

**INSTITUTO POLITÉCNICO NACIONAL**



**ESCUELA NACIONAL DE CIENCIAS BIOLÓGICAS**

**POSGRADO EN BIOMEDICINA Y BIOTECNOLOGÍA MOLECULAR**

**ESTUDIO DE MUTAGÉNESIS INSERCIONAL DEL  
RETROVIRUS MMTV/HMTV EN MUESTRAS DE  
ADENOCARCINOMAS MAMARIOS DE MUJERES  
MEXICANAS**

**T E S I S**

Que como uno de los requisitos  
para obtener el grado de:

**MAESTRO EN CIENCIAS EN BIOMEDICINA  
Y BIOTECNOLOGÍA MOLECULAR**

PRESENTA:

**IBQ. ABRAHAM PEDROZA TORRES**

DIRECTORES:

DR. NORMAND GARCÍA HERNÁNDEZ

DR. ALFONSO MÉNDEZ TENORIO



# INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 13:00 horas del día 22 del mes de junio del 2011 se reunieron los miembros de la Comisión Revisora de Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de la ENCB-IPN para examinar la tesis titulada:

**Estudio de mutagénesis insercional del retrovirus MMTV/HMTV en muestras de adenocarcinomas mamarios de mujeres mexicanas**

Presentada por el alumno:

**Pedroza**

Apellido paterno

**Torres**

Apellido materno

**Abraham**

Nombre(s)

Con registro: 

B	0	9	1	0	4	9
---	---	---	---	---	---	---

aspirante de:

**Maestría en Ciencias en Biomedicina y Biotecnología Molecular**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA

Director de tesis

Dr. Alfonso Méndez Tenorio

Dr. Normand García Hernández

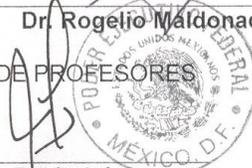
Dr. José Luis Muñoz Sánchez

Dr. Sergio Enrique Meza Toledo

Dr. Diego Julio Arenas Aranda

Dr. Rogelio Maldonado Rodríguez

PRESIDENTE DEL COLEGIO DE PROFESORES



Dr. Manuel Jesús Piñón López

S.E.P.  
Escuela Nacional de Ciencias Biológicas  
Sección de Estudios de Posgrado e Investigación



**INSTITUTO POLITECNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

*CARTA CESIÓN DE DERECHOS*

En la Ciudad de México el día **20** del mes **Junio** del año **2011** , el (la) que suscribe **Abraham Pedroza Torres** alumno del Programa de **Maestría en Biomedicina y Biotecnología Molecular** con número de registro **B0910489** , adscrito al **Laboratorio de Genómica Funcional y Proteómica, UIMGH del Hospital de pediatría, CMN S-XXI** , manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de **Dr. Normand García Hernández** y **Dr. Alfonso Méndez Tenorio** y cede los derechos del trabajo escrito titulado **“Estudio de mutagenesis insercional del retrovirus MMTV/HMTV en muestras de adenocarcinomas mamarios de mujeres mexicanas”**, al Instituto Politécnico Nacional para su difusión y fines académicos.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **normandgarcia@yahoo.com**, Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Atentamente,

Abraham Pedroza Torres  
Nombre y firma del alumno(a)

Dr. Alfonso Méndez Tenorio  
Nombre y firma del Director de tesis

Dr. Normand García Hernández  
Nombre y firma del Director de tesis  
**LABORATORIO DE GENÓMICA**  
**FUNCIONAL Y PROTEÓMICA**  
**IMSS U.I.M. EN GENÉTICA HUMANA**

Este trabajo se realizó en la Laboratorio de Genética Molecular, UIMGH del hospital de Pediatría, CMN S-XXI, IMSS y en el Laboratorio de Biotecnología y Bioinformática Genómica, Departamento de Bioquímica, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional. El proyecto se desarrolló mediante el financiamiento 80620 del CONACyT, FIS/IMSS/PROT/G09/741 y FIS/IMSS/PROT/G09/602.

*A Dios (porque yo sé que existe):*  
Por su gran misericordia para conmigo

*A mis Padres:*

María Felix Torres Cornelio y Arturo Pedroza Ramírez. No tengo palabras suficientes para agradecerles todo lo que han hecho por mí. Que les debo todo, que me dieron la vida y han estado conmigo en todo momento, que a pesar de los momentos difíciles han logrado salir adelante. Que me han dado las mejores armas para ser alguien en la vida -escuela y amor- sin dudarlo los volvería a escoger como padres. A ellos les dedico esto.

*A Rosa, Maribel y Sonia:*

Mis hermanas, que han luchado conmigo todos y cada uno de estos años, que me han apoyado en todo cuanto han podido, que han vivido de cerca mis logros y mis fracasos, que me sirven de inspiración para salir adelante por ellas y para ellas.

Quiero agradecer sinceramente al Dr. Normand García Hernández mi tutor de estudios y asesor en este trabajo, por sus valiosas aportaciones y su intensa labor a la hora de revisar este trabajo, si su ayuda este trabajo no sería lo que es, mis más sinceros reconocimientos.

También quiero agradecer al Dr. Alfonso Méndez Tenorio por darme la oportunidad de pertenecer a ese maravilloso grupo de trabajo del cual él forma una parte muy importante -el laboratorio de Bioinformática Genómica en la ENCB-. Así como sus valiosas aportaciones y su interés en el trabajo. Gracias.

A mis sinodales el Dr. Diego J. Arenas Aranda, el Dr. José Luis Muñoz, el Dr. Sergio Meza, el Dr. Rogelio Maldonado gracias por su participación y su empeño en que este trabajo quedara lo más presentable posible así como todos y cada uno de sus concejos metodológicos en la realización de este trabajo, a todos ellos mi reconocimiento y agradecimiento.

A mis compañeros de laboratorio en genética humana Xchel Rivera, Gaby bravo, Ignacio Virgilio, Susana, Nitzia, Erick, Alex, Alberto, Mariana gracias por su apoyo y por compartir conmigo todas esas experiencias que nos da hacer ciencia día con día.

A mis compañeros de la maestría Iveth, Paco, Memo, Joaquín, Alberto, Sofía, gracias por todos los momentos que pasamos juntos por todas las palabras de aliento, a ustedes chicos no tengo más que decirles si no GRACIAS.

## **ABREVIATURAS**

<b>DNA</b>	Deoxyribonucleic acid (Ácido desoxirribonucleico)
<b>MMTV</b>	Mouse Mammary Tumor Virus (Virus del tumor mamario murino)
<b>HERVs</b>	Human endogenous retroviruses (Retrovirus endógenos humanos)
<b>RVs</b>	Retrovirus exógenos
<b>PCR</b>	Polymerase Chain Reaction (Reacción en cadena de la polimerasa)
<b>LTR</b>	Long Terminal Repeats (Repeticiones terminales largas)
<b>RNA</b>	Ribonucleic acid (Ácido ribonucleico)
<b>MHC</b>	Major Histocompatibility Complex (Complejo mayor de Histocompatibilidad)
<b>IMMS</b>	Instituto Mexicano del Seguro Social
<b>NCBI</b>	National Center for Biotechnology Information (Centro Nacional de Información sobre Biotecnología)
<b>EMBL-EBI</b>	European Molecular Biology Laboratory- European Bioinformatics Institute (Laboratorio Europeo de Biología Molecular-Instituto Europeo de Bioinformática)
<b>BLAST</b>	Basic Local Alignment Search Tool (Herramienta básica de búsqueda de alineamientos locales)
<b>Dx</b>	Diagnóstico
<b>CDI</b>	Carcinoma Ductal Infiltrante
<b>CLI</b>	Carcinoma Lobulillar Infiltrante
<b>CSC</b>	Carcinoma Sin Clasificar
<b>TNA</b>	Tejido No Afectado
<b>FBA</b>	Fibroadenoma

## ÍNDICE

1. Introducción .....	14
1.1 Cáncer .....	14
1.2 El cáncer de mama en México .....	18
1.3 Factores de Riesgo .....	19
1.4 El cáncer de mama y su posible origen viral.....	19
1.5 Los retrovirus endógenos humanos o HERVs.....	23
1.6 Estructura general de los retrovirus.....	25
2. Antecedentes.....	26
3. Planteamiento del problema .....	30
4. Justificación .....	31
5. Objetivos .....	32
6. Materiales y métodos .....	33
6.1 Obtención de las muestras de adenocarcinomas .....	33
6.2 Obtención de DNA a partir de las muestras obtenidas .....	33
6.3 PCR anidada para la amplificación del gen <i>env</i> .....	33
6.4 PCR splinkerette para amplificar las secuencias flanqueadas por las LTRs virales	35
6.5 Búsqueda Bioinformática.....	37
7. Resultados.....	39
7.1 Obtención de las muestras de adenocarcinomas .....	39
7.2 Obtención del DNA a partir de las muestras obtenidas.....	39
7.3 PCR anidada para la amplificación del gen <i>env</i> .....	41
7.4 Purificación y análisis de las secuencias obtenidas a partir de los amplificados para el gen <i>env</i> .....	51

7.5 PCR splinkerette para amplificar las secuencias flanqueadas por las LTRs virales	61
7.6 Búsqueda Bioinformática	68
8. Discusión	106
9. Conclusiones	113
10. Referencias	114

## ÍNDICE DE FIGURAS

Figura 1. Procesos que interfieren en cada una de las capacidades adquiridas y necesarias para el crecimiento y la progresión tumoral	10
Figura 2. Divisiones mitóticas de un óvulo fecundado hasta una sola célula cancerosa	12
Figura 3. Organización genómica un retrovirus en su forma libre como RNA y la integración en el genoma como provirus	19
Figura 4. Organización genómica de los retrovirus exógenos	20
Figura 5. Estrategia de amplificación del gen <i>env</i> del MMTV	29
Figura 6. Estructura generada para el adaptador (zee linker)	30
Figura 7. Esquema del método de PCR splinkerette	32
Figura 8. Muestras de tejidos mamarios obtenidos (adenocarcinomas y tejidos mamarios no afectados)	33
Figura 9. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 001CM- 006CM del gen <i>env</i> del MMTV	36
Figura 10. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 007CM-012CM del gen <i>env</i> del MMTV	37
Figura 11. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 013CM-018CM del gen <i>env</i> del MMTV	37
Figura 12. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 019CM-024CM del gen <i>env</i> del MMTV	38
Figura 13. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 025CM-028CM del gen <i>env</i> del MMTV	38
Figura 14. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 001CM-011CM del gen <i>env</i> del MMTV	39
Figura 15. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 012CM- 023CM del gen <i>env</i> del MMTV	40
Figura 16. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 024CM-030CM del gen <i>env</i> del MMTV	40
Figura 17. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 031CM-037M del gen <i>env</i> del MMTV	41

Figura 18. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 001CM-012CM del gen GAPDH	43
Figura 19. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 013CM-019CM del gen GAPDH	43
Figura 20. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 020CM-023CM del gen GAPDH.	44
Figura 21. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 024CM-030CM del gen GAPDH	44
Figura 22. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 031CM-049CM del gen GAPDH	45
Figura 23. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 049CM-056CM del gen GAPDH	45
Figura 24. Secuencias obtenidas de los amplificados de las muestras de adenocarcinomas 002CM, 004CM, 007CM y 027CM así como para el control positivo	47
Figura 25. Secuencias obtenidas a partir de los amplificados de las muestras de adenocarcinomas 031CM, 032CM para el gen env y de las muestras 001CM y 002CM un fragmento del gen GAPDH.	48
Figura 26. Alineamiento obtenido del amplificado de la muestra 001CM del gen GAPDH	49
Figura 27. Alineamientos obtenidos de la secuencia amplificada del gen env del MMTV de una muestra de adenocarcinoma mamario (Muestra 002CM)	50
Figura 28. Alineamiento de la secuencia obtenida del gen env del MMTV de una muestra de adenocarcinoma mamario (muestra 002CM)	52
Figura 29. Alineamiento de la secuencia obtenida del gen env del MMTV de una muestra de adenocarcinoma mamario (muestra 002CM)	52

## Resumen

En México a partir del 2006 el cáncer de mama es la primera causa de muerte por neoplasias en mujeres, los datos de incidencia se reportan en 38 por 100,000 habitantes. Se han identificado la mayoría de los factores de riesgo, sin embargo, la etiología aún es desconocida. Los factores identificados incluyen: a) Factores genéticos: historia familiar con cáncer, principalmente mutaciones en los genes BRCA-1 y/o BRCA-2, b) Factores endocrinos: larga exposición a estrógenos, nuliparidad, menarca temprana y menopausia tardía y c) Factores exógenos: Dieta, consumo de alcohol, cigarrillos, agentes químicos, radiaciones ionizantes, etc. Recientemente diversos grupos de trabajo han sugerido la posibilidad de que un retrovirus (el Virus del Tumor Mamario Murino o MMTV) pudiera ser el agente etiológico de algunos cánceres de mama. El objetivo del presente trabajo es identificar un fragmento de 250 pb. del gen *env* del MMTV en adenocarcinomas mamarios en mujeres mexicanas y evaluar los sitios de inserción del retrovirus mediante PCR splinkerette. Se analizaron 56 muestras de adenocarcinoma mamarios y tejidos no afectados. Las muestras fueron procesadas y el DNA extraído. La extracción del DNA se realizó mediante el equipo QIAcube siguiendo las indicaciones del fabricante. Se verificó la integridad y se cuantificó para una posterior PCR anidada, para la cual se utilizaron los oligonucleótidos propuestos por Wang y col. en 1995. Se utilizó una construcción hecha con el gen *env* del MMTV cepa C3H sobre un plásmido pBR322 y propagado en E.coli HB101 como control positivo de reacción, así mismo se amplificó un fragmento de 500pb del gen GAPDH para verificar la utilidad de los DNAs. Una vez identificados los productos de PCR positivos para *env* estos fueron secuenciados y las secuencias alineadas con BLAST 2.2.24. Una vez detectadas las muestras positivas para MMTV éstas fueron evaluadas mediante PCR splinkerette para detectar los posibles sitios de inserción del retrovirus. Se detectó la presencia de secuencias parecidas al gen *env* del MMTV en 16 (40%) de las muestras analizadas y en ninguno de los tejidos no afectados. Los alineamientos hechos a las secuencias amplificadas arrojaron una similitud del 96- 98% entre las secuencias amplificadas y el gen *env* del MMTV. Los análisis hechos mediante PCR splinkerette arrojaron diversos sitios de amplificación, se pudieron detectar 12 secuencias (genes y regiones intergénicas) entre ellas: proteína nucleolar 4, Factor de iniciación 5, proteína de ensamble a nucleosoma 1, Factor 6 de crecimiento, proteína. Es posible detectar secuencias del MMTV en adenocarcinomas mamarios de mujeres mexicanas. Las secuencias obtenidas de los amplificados detectados en los adenocarcinomas mamarios corresponden hasta en un 98% de homología al gen *env* del MMTV depositado en el NCBI. Algunos de los genes detectados están involucrados en procesos diferenciación y crecimiento celular.

## **Abstract**

In Mexico from 2006 breast cancer is the leading cause of death from malignancy in women, incidence data are reported at 38 per 100,000 habitants. Have identified most of the risk factors, however, the etiology is unknown. The factors identified include: a) Genetic factors: family history of cancer, especially mutations in the BRCA-1 and / or BRCA-2 genes, b) Endocrine factors: long exposure to estrogen, nulliparity, early menarche and late menopause c) Factors Exogenous: diet, alcohol, cigarettes, chemicals, radiation, etc.. Recently, several working groups have suggested the possibility that a retrovirus (the murine mammary tumor virus or MMTV) may be the causative agent of some breast cancers. The aim of this study is to identify a fragment of 250 bp. MMTV *env* gene in mammary adenocarcinomas in Mexican women and evaluate the retrovirus insertion sites by PCR splinkerette. We analyzed 56 samples of breast adenocarcinoma and uninvolved tissue. Samples were processed and DNA extracted. DNA extraction was performed using the computer QIAcube following manufacturer's instructions. Integrity was verified and quantified for subsequent nested PCR, for which we used oligonucleotide proposed by Wang et al. in 1995. We used a construction made by the *env* gene of MMTV C3H strain on a plasmid pBR322 and propagated in E.coli HB101 as a positive control reaction, so it was amplified a 500pb fragment of the gene GAPDH to verify the usefulness of the DNAs. Having identified the positive PCR products were sequenced and *env* these sequences aligned with BLAST 2.2.24. Having detected these MMTV-positive samples were tested by PCR to detect possible splinkerette insertion sites of retrovirus. Detected the presence of sequences similar to MMTV *env* gene in 16 (40%) of the samples and the tissues unaffected. The alignments made to the amplified sequences showed a similarity of 96 - 98% between the amplified sequences and the MMTV *env* gene. The analysis made by PCR yielded different sites splinkerette amplification could be detected 12 sequences (genes and intergenic regions) including: protein nucleolar 4, initiation factor 5, a nucleosome assembly protein 1, Factor 6, growth, protein. Sequences can be detected MMTV in mammary adenocarcinomas of Mexican women. The sequences obtained from amplicons detected in breast adenocarcinomas are up to 98% homology to MMTV *env* gene deposited in the NCBI. Some of the genes identified are involved in cell growth and differentiation processes.

# 1. Introducción

## 1.1 Cáncer

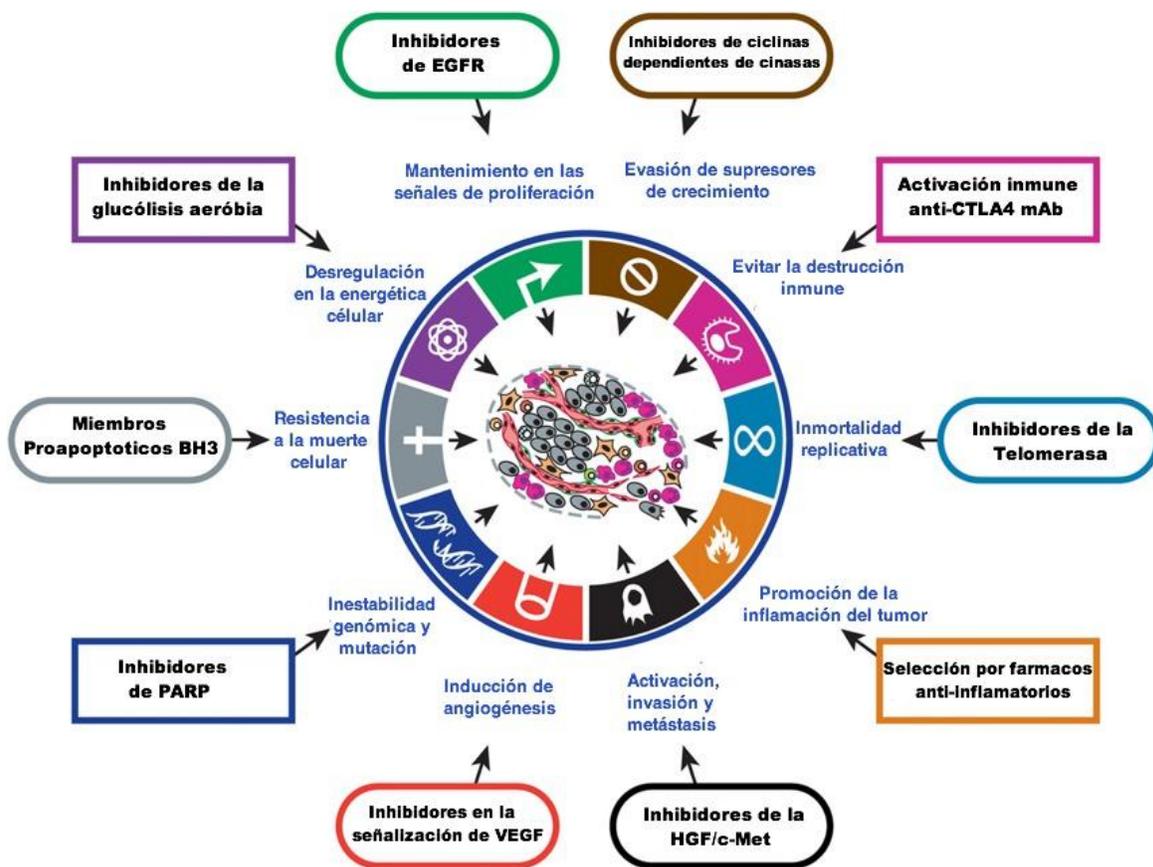
La teoría del “daño múltiple” para explicar el origen del cáncer sugiere que, las células acumulan daño aleatorio en sus genes (mutaciones). Por accidente, algo del daño ocurre en genes críticos que normalmente controlan de forma rigurosa los procesos de proliferación y apoptosis en las células (oncogenes, genes supresores de tumor o genes de reparación del DNA). El cáncer ha sido llamada “la enfermedad de los genes”, debido a que se ven afectados un sin número de genes en el desarrollo de esta enfermedad, los cuales están implicados en procesos distintos tales como la multiplicación, especialización y/o muerte celular. Se conocen algunos genes involucrados en diversos procesos del desarrollo neoplásico, por ejemplo: PCNA en proliferación, los oncogenes FOS, RAS, MYC, supresores de tumor como P53, RB, BRCA1, BRCA2, WT1, WT2, FOXP3, ATM, EGF en crecimiento celular, y en apoptosis BAX, BAD, BID, BCL-2, MCL-1 y FAS (Adams J y Cory S., 2007, Amit I y col., 2007, Osin y Lakhani, 1999) (ver figura 1).

Si alguna célula acumula suficientes daños, múltiples mutaciones (de cinco a diez), la célula se transforma en un fenotipo maligno con pérdida en el control de la proliferación y la muerte celular, es decir una neoplasia (Kesteloot H y Zhang J, 2006).

Los tumores pueden originarse en cualquier célula del organismo, pero es raro observar tumores a partir de células cuyo índice mitótico es bajo o nulo. Algunos proceden de una única estirpe celular, pero, el pleiomorfismo y la variabilidad son las propiedades más comunes a las células cancerosas. El cáncer ha sido definido como un crecimiento neoplásico que tiene la capacidad de invadir los tejidos circundantes y diseminarse por la corriente sanguínea y/o linfática.

Existen cuatro características que describen la forma en que las células cancerosas actúan de modo distinto a las células normales, dichas características son las siguientes:

- Proliferación: el cáncer se origina de una única célula progenitora que se divide y da lugar a células malignas.
- Autonomía: el crecimiento no es regulado de forma adecuada por las influencias bioquímicas y físicas normales del ambiente.
- Anaplasia: existe una ausencia de diferenciación celular normal y coordinada.
- Metástasis: las células cancerosas tienen la capacidad de crecer y diseminarse a otras partes del cuerpo (Isselbacher y col. 1994).

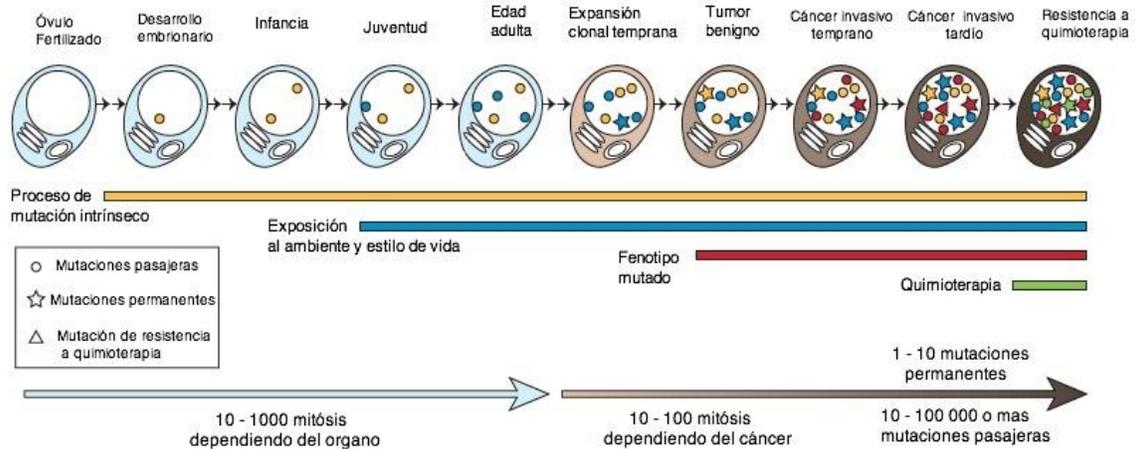


**Figura 1.** Procesos que interfieren en cada una de las capacidades adquiridas y necesarias para el crecimiento y la progresión tumoral. Las capacidades adquiridas por la célula cancerosa incluyen procesos muy diversos y complejos, tales como; reprogramación del metabolismo energético y activación de mecanismos inmunológicos, así como inhibición de la Telomerasa, entre otros (Modificado de *Hallmarks of Cancer: The next generation*, Weinberg R y Hanahan D, 2011).

Estas propiedades pueden ser expresadas por las células normales no malignas durante épocas adecuadas, por ejemplo durante la embriogénesis. No obstante en las células cancerosas, estas características tienen un grado inapropiado o excesivo. Los tumores benignos son clónales y tienen un cierto grado de autonomía, pero permanecen diferenciados y no realizan metástasis. El proceso mediante el que una célula normal se convierte en una célula maligna que presenta estas características se denomina transformación maligna (Isselbacher y col. 1994).

Al igual que todas las células del cuerpo, una célula cancerosa es descendiente directa a través de un linaje de divisiones celulares mitóticas, de un óvulo fertilizado y por lo tanto lleva una copia de su genoma diploide. Sin embargo, la secuencia de DNA del genoma de una célula cancerosa, y de hecho la mayoría de los genomas de células normales, han adquirido un conjunto de diferencias respecto a su progenitor óvulo fertilizado (ver figura 2). Estas son denominadas colectivamente mutaciones somáticas para distinguirlas de las mutaciones en la línea germinal que se heredan de los padres y se transmiten a la descendencia.

Las mutaciones somáticas en el genoma de una célula de cáncer pueden abarcar diversos tipos de alteraciones para cambiar la secuencia de DNA. Estas incluyen sustituciones de una base por otra; inserciones o eliminaciones de pequeños o grandes segmentos de DNA; reordenamientos, en la cual un fragmento de DNA se ha roto y luego se reincorpora a un segmento de DNA en otras partes del genoma, aumento en el número de copias (de las dos copias en la genoma diploide normal) hasta algunas veces varios cientos de copias (proceso conocido como amplificación de un gen) y reducciones en el número de copias que pueden dar lugar a la ausencia total de una secuencia de DNA en el genoma de una célula de cáncer (Stratton M, Campbell P, Futreal A., 2009)



**Figura 2. Divisiones mitóticas de un óvulo fecundado hasta una sola célula cancerosa.** Las mutaciones pueden ser adquiridas, mientras el linaje de la célula es fenotípicamente normal, solo refleja tanto las mutaciones intrínsecas adquiridas durante la división celular normal y los efectos de los mutágenos exógenos. Durante el desarrollo de los procesos de todo tipo de cáncer, por ejemplo, defectos de reparación del DNA, contribuyen a la carga mutacional. Las mutaciones pasajeras no tienen ningún efecto sobre las células cancerosas, pero las mutaciones permanentes (en genes implicados en procesos de desarrollo celular) causará una expansión clonal. La recaída después de la quimioterapia puede estar asociada a mutaciones de resistencia que a menudo son anteriores a la iniciación del tratamiento (Stratton M, Campbell P, Futreal A., 2009)

En adición, la célula cancerosa puede haber adquirido, desde fuentes exógenas, secuencias completas de DNA nuevo, particularmente de virus tales como; el virus del papiloma humano, virus de Epstein Barr, virus de la hepatitis B, Virus linfotrópico de células T y el Virus del herpes humano ocho, cada uno de ellos se sabe que puede contribuir a la génesis de uno o más tipos de cancer (Talbot S y col., 2004.). En comparación con el óvulo fertilizado, el genoma de un cáncer también ha adquirido cambios epigenéticos que alteran la estructura de la cromatina y la expresión de genes y que se manifiestan a nivel de secuencias de DNA por cambios en el estado de metilación de algunos residuos de citosina. Los cambios epigenéticos suelen estar sujetos a la misma presión de selección.

No hay que olvidar que existe otro genoma que se alberga dentro de la célula cancerosa. Las miles de mitocondrias presentes suelen llevar un genoma circular de aproximadamente 17 kilobases. Diversas mutaciones somáticas en genomas

mitocondriales han sido reportados en muchos cánceres humanos, aunque su papel en el desarrollo de la enfermedad no está aún claro (Chatterjee A y col., 2006).

La mayoría de los cánceres de mama son considerados neoplasias epiteliales, es decir carcinomas, y se suelen originar en la unidad ducto-lobulillar. Al proceder de una glándula se denominan más adecuadamente adenocarcinomas. Las neoplasias que se presenta con mayor frecuencia son los tumores de estirpe epitelial (en el epitelio ductal en un 70-80% de los casos y en el lobulillar en un 10-20%), aunque cualquiera de los tejidos que componen la mama puede originar un cáncer (Ruiz E, 1993).

## 1.2 El cáncer de mama en México

En México un país con una población un poco mayor de 105 millones de habitantes, el cáncer de mama es hoy en día uno de los desafíos más importantes para el sector salud (Lozano R y col. 2008).

Las tasas de mortalidad por cáncer de mama en México muestran un aumento notorio en las últimas cinco décadas. Entre 1955 y 1960, a partir de la disposición de los primeros datos confiables, la tasa era alrededor de dos a cuatro muertes por cada 100,000 mujeres, luego se elevó de manera sostenida en las mujeres adultas de todas las edades hasta alcanzar una cifra cercana a 9 por cada 100,000 mujeres para la mitad de la década de 1990. Para el año 2005, el cáncer de mama se había convertido en la segunda causa de muerte más común en México entre las mujeres de entre 30 a 54 años y la tercera más frecuente entre el grupo de 30 a 59 años (después de la diabetes y las cardiopatías). Tan solo en el 2006 murieron 4,451 mujeres mexicanas, lo cual implica un fallecimiento cada 2 horas. La tasa de mortalidad por cáncer mamario se incrementó 2.5 veces de 1992 a 2006, de tal forma que a partir de ese año la tasa de mortalidad de cáncer de mama es superior a la de cáncer cérvico uterino (Lozano R y col. 2008).

A principios de 2006, el cáncer de mama se ubicó como la primera causa de mortalidad por tumores malignos entre las mujeres en algunos estados de la

república, tales como Jalisco; donde desde 1997 ocupa el primer lugar, con una tasa de mortalidad de 15.82% por 100,000 mujeres y la frecuencia de cáncer de mama es ya casi similar (17.9% vs. 18.8%) a la del cáncer cérvicouterino. Esta tendencia parece mantenerse en la mayoría de los estados, lo que permite estimar que en muy poco tiempo será ya la primera causa de muerte por cáncer en todo el país (Lozano R y col. 2008).

### 1.3 Factores de Riesgo

El 80% de los cánceres de mama no tienen una causa conocida, aunque aparentemente el cáncer de mama depende de una conjunción de varios factores. Por un lado están los factores hormonales (los estrógenos, la menarca temprana y nuliparidad) y por otro, se encuentran los factores ambientales (la dieta, la edad, el sedentarismo, etc.), que van a influir en el riesgo de padecer cáncer de mama a través de las modificaciones que ejercen en el organismo de cada persona. La edad y el sexo son los factores de riesgo más importantes, ya que el riesgo de cáncer de mama aumenta conforme va aumentando la edad, y el 99% de los cánceres de mama aparecen en mujeres (Zamora P y col. 2001).

Otro factor importante es el factor genético y familiar, ya que representa alrededor del 10-20 % de estos tipos de cáncer, sin embargo, pudieran ayudar a comprender algunos mecanismos del desarrollo tumoral. Las alteraciones genéticas, a su vez, pueden ser heredadas como mutaciones de la línea germinal o como mutaciones somáticas adquiridas. Estas últimas podrían ocurrir como resultado de la exposición a carcinógenos ambientales, ya sean físicos (ej. Radiaciones ionizantes excesivas), químicos (ej. hidrocarburos policíclicos, nitrosoureas) y/o biológicos (ej. virus) (Martínez J, 1995; Russo J y Russo I, 2000).

### 1.4 El cáncer de mama y su posible origen viral

Se estima que un poco más del 15 por ciento de todos los tipos de cáncer en humanos pueden atribuirse a una infección viral, lo que representaría una porción significativa del número total de casos que se presentan anualmente (1.5 millones

de casos al año). Se ha demostrado que tanto los virus de DNA como los de RNA son capaces de causar cáncer en humanos. El virus de Epstein-Barr, el Virus del papiloma humano, el Virus de la hepatitis B y el Herpes virus humano 8 son los cuatro virus de DNA que se sabe son capaces de causar cáncer en humanos. El Virus linfotrópico de células T humanas y el Virus de la hepatitis C son los dos virus de RNA que están asociados a algunos cánceres en humanos (ver tabla 1) (Zur Hausen H, 1991).

A pesar de que los virus oncogénicos humanos pertenecen a diferentes familias de virus y suelen utilizar diversas estrategias para contribuir al desarrollo del cáncer, comparten muchas características comunes. Una característica fundamental es su capacidad para infectar, pero no matar, su célula huésped. A diferencia de muchos otros virus que causan enfermedades, los virus oncogénicos tienen la tendencia a establecer infecciones persistentes a largo plazo. En consecuencia, han desarrollado estrategias para evadir la respuesta inmune del hospedero. A pesar de la etiología viral de varios tipos de cáncer, parece que los virus a menudo pueden contribuir, pero solos no son suficientes para iniciar el proceso de carcinogénesis, de hecho, la mayoría de las personas infectadas por algún virus del tumor no desarrollan cáncer, es decir, al parecer es necesario un conjunto de eventos para iniciar el proceso de carcinogénesis (Cohen S y Parsonnet J., 1999).

Los virus pueden contribuir al desarrollo de tumores por diferentes mecanismos: indirectamente, mediante la inducción de inmunosupresión o mediante la modificación del genoma del hospedero sin la persistencia del DNA viral y directamente induciendo oncoproteínas o mediante la alteración de la expresión de las proteínas del huésped en el lugar de la integración del DNA viral (mutagénesis insercional) (Minami M y col., 2005).

**Tabla 1. Virus asociados a distintos tipos de cáncer en humanos** (Modificado de *Human tumor-associated viruses and new insights into the molecular mechanisms of cancer*, Martin D y Gutkind J., 2009).

Virus	Familia	Tipo de cáncer	Casos/año	Mecanismo	Oncogenes	Función del oncogén	Referencia
Virus de la Hepatitis B (HBV)	Hepadnaviridae	Carcinoma Hepatocelular	3400 000	Inflamación crónica	Proteína X CREB, ATF	Desregulación del señalamiento molecular Inhibición de p53	Ganem y Prince, 2004; Guidotti y Chisari, 2006 Kao y Chen, 2002
Virus de la Hepatitis C (HCV)	Flaviviridae	Carcinoma hepatocelular	195 000	Inflamación crónica	-	-	Colombo y col., 1989; Thomas y col., 2000
Virus de Epstein-Barr (EBV)	Herpesviridae	Linfoma de Burkitt Linfoma de Hodgkin Carcinoma nasofaríngeo	113 400	Oncogénico	LMP-1	Desregulación del señalamiento molecular Activación de NF- $\kappa$ B Linfoproliferación	Arvanitakis y col., 1995; Brown y col., 2001; Mosialos y col., 1995)
Virus del Papiloma humano (HPV)	Papilomaviridae	Cáncer cérvico Cáncer anal Cáncer de pene Carcinoma de cabeza y cuello	561 180	Oncogénico	E6/E7	Inhibición de p53 y Rb Desregulación en los mecanismos de adhesión celular	Beaudenon y col., 1986; Dyson y col., 1989b; Scheffner y col., 1990
Virus linfotrópico de células T humanas (HTLV-1)	Retroviridae	Leucemia de células T en adultos	3 330	Oncogénico	Tax	Desregulación del señalamiento molecular Activación de NF- $\kappa$ B	Matsuoka and Jeang, 2007; Poesz y col., 1980
Herpes virus asociado al sarcoma de Kaposi (Herpes virus humano 8) (KSHV) (HHV8)	Herpesviridae	Sarcoma de Kaposi Linfoma de efusión Pleural Enfermedad Multicéntrica de Castleman	102 300	Oncogénico	vGPCR, vIL6, vBcl2, vMIPs, vFlip, vCyclin, LANA, Kaposin B	Múltiples eventos de desregulación del ciclo celular Inhibición de la apoptosis Evasión inmune	Arvanitakis y col., 1997; Bais y col., 1998; Cesarman y col., 1995; Chang y col., 1994; Ganem, 2006; Montaner y col., 2003; Yang y col., 2000

El interés sobre el MMTV (Mouse Mammary Tumor Virus) como agente cancerígeno fue generado varias décadas atrás, cuando Bittner demostró en 1939 que un factor extra cromosómico en los ratones contribuía a la incidencia del cáncer de mama. En sus experimentos, utilizó una línea de ratones con alta tasa de incidencia tumoral (>80%), éstos ratones fueron amamantados por ratones hembras con una incidencia baja a tumores mamarios. Los ratones amamantados y su progenie vivían mucho más tiempo y tenían una incidencia de tumores de sólo el 7,4%. Con este y otros experimentos similares, llegó a la conclusión de que un agente de la leche influía en la desarrollo de los tumores de mama (Bittner J., 1939).

En 1945, Andorvent describe este agente de la leche como un virus (Gross L., 1970), y poco a poco, se hizo ampliamente conocido como el "Virus de Bittner" actualmente conocido como MMTV. Se ha demostrado que éste virus se transmite por vía endógena y exógena. En el primer caso se transmite a través de la línea germinal y todas las células heredan la secuencia viral. Para el caso de la vía exógena el virus es transmitido a través de la leche materna. El virus responde a un estímulo estrogénico aumentando su expresión durante la lactancia de tal manera que las partículas virales en la leche son capaces de infectar a la progenie, transmitiéndose en forma exógena (Choi Y y col., 2004, Brandat-Carlson C y col., 1993).

Actualmente se ha sugerido que el ratón doméstico pudiera ser una fuente de transmisión para contraer el cáncer mamario en humanos (Stewart T., 2000). En el año 2004, un grupo de investigadores informó que había hallado DNA similar al del MMTV en los tumores de mama y en linfomas no-Hodgkin en pacientes que sufrían ambas enfermedades (Etkind P., 2004). Mediante análisis de similitud entre las secuencias identificadas en los tumores mamarios se encontró alta similitud entre las secuencias del retrovirus humano con el MMTV (90-95%) y una menor con los HERV's (57%). Adicionalmente, diversos grupos de trabajo han demostrado que secuencias parecidas al gen env del virus del tumor mamario del

ratón están presentes en tumores mamarios humanos (Indik S y col., 2005, Etkinnd P y col., 2000, Ford C y col., 2004, Pogo B y col., 1997, Wang Y y col. 1995).

Se reconocen tres clases de evidencia que sustentan esta fuerte asociación entre el MMTV y el cáncer de mama:

1. La evidencia física de que ciertos tumores de mama humanos, líneas celulares y leche materna humana contienen DNA idéntico al DNA del MMTV
2. Las respuestas inmunológicas al MMTV
3. Estudios moleculares que demuestran que las secuencias virales de DNA o RNA similar al MMTV puede hallarse en más de un tercio de los cánceres de mama humanos pero no se hallan en el tejido de mama normal (Melana S., 2007).

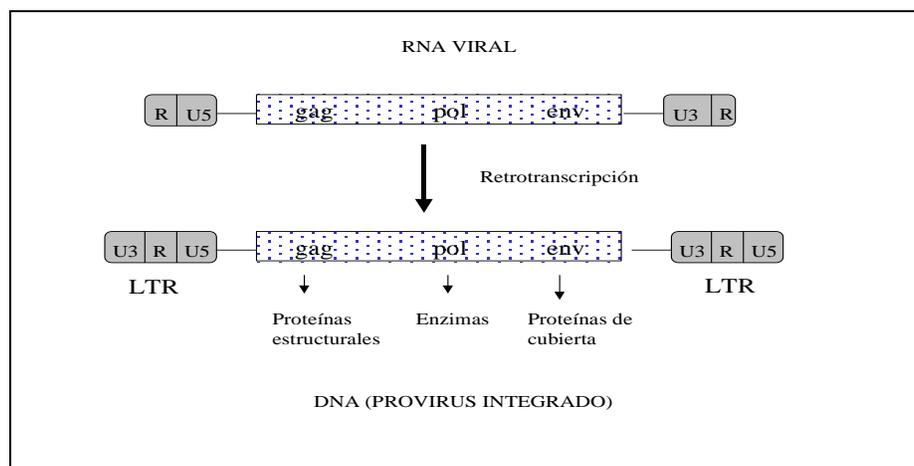
#### 1.5 Los retrovirus endógenos humanos o HERVs

Cerca del 10% del genoma humano está compuesto por secuencias de origen retroviral muy diferentes entre sí, pero englobadas bajo la denominación genérica de retrovirus endógenos humanos o HERVs (de sus siglas en inglés). La existencia de miles de secuencias de las distintas familias de HERVs en el genoma humano que presentan similitud parcial y capacidad de retrotransposición es una fuente enorme de variabilidad y plasticidad genómica (Bock M y Stoye J., 2000).

La estructura de los HERVs es la característica de los retrovirus exógenos (RVs) en su forma de provirus, contienen dos repeticiones terminales largas en los extremos (LTRs) y los genes *gag*, *pol* y *env*, que codifican las proteínas necesarias para la formación de nuevas partículas virales incluyendo la Transcriptasa reversa (RT) la cual permite la transcripción de un molde de RNA en DNA (retrotranscripción) y es parte esencial del ciclo vital de los retrovirus (ver figura 3). Las LTRs se generan durante el proceso de retrotranscripción y son necesarias para la integración viral en el DNA genómico. Además, presentan elementos

necesarios para regular la expresión de sus genes, incluyendo las secuencias a las que se unen factores de transcripción y hormonas, y las señales de poliadenilación para el correcto procesamiento de los RNA virales. Gracias a sus regiones con similitud de secuencias promueven la recombinación no homologa en distintas regiones del genoma, promoviendo en general reorganizaciones genómicas y en algunos casos concretos, duplicaciones génicas, tanto de sí mismas, como de cualquier gen que esté flanqueado por las regiones de similitud (Löwer R y col., 1996)

Los retrovirus endógenos se encuentran integrados en el genoma del hospedero, los HERVs se transmiten de modo vertical (mendeliano) a la descendencia, y no tienen la capacidad de producir partículas virales, mientras que los RVs se transmiten de modo horizontal (infectivo) entre los individuos de una población. La integración de HERVs produce además diversidad alélica en las poblaciones, como se ha demostrado en la región del Complejo Mayor de Histocompatibilidad (MHC) (Tassabehji y col., 1994; Dawkins y col., 1999; Kulski y col., 1999) y puede producir respuestas individuales diferenciales ante determinadas enfermedades, sobre todo contra las infecciones virales exógenas (Dangel et al., 1994; Schneider et al., 2001)



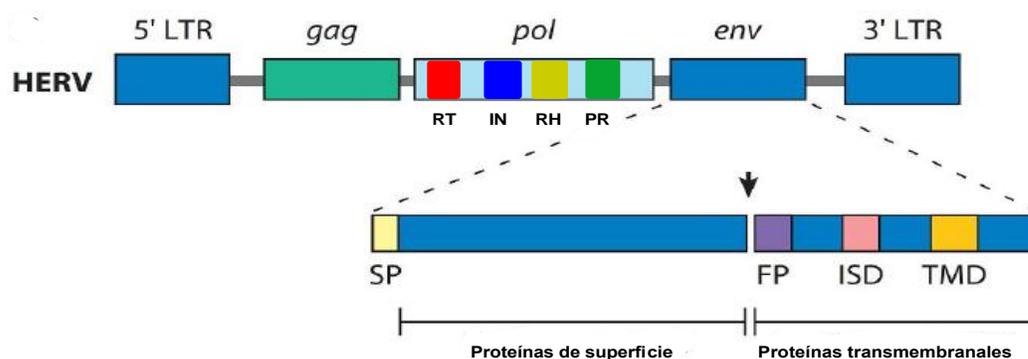
**Figura 3. Organización genómica un retrovirus en su forma libre como RNA y la integración en el genoma como provirus.** Los retrovirus incluyen en su genoma los genes *env*, *pol* y *gag* necesarios para la integración posterior a las células hospederas. Esta estructura y sus variantes defectivas son las que corresponden a los HERVs.

## 1.6 Estructura general de los retrovirus

La envoltura externa de los retrovirus proviene de la membrana plasmática de la célula huésped. Las proteínas de envoltura (antígenos de superficie) son codificadas por el gen *env*. Dentro de la envoltura hay una cápside icosaédrica que contiene proteínas codificadas por el gen *gag* (Antígeno grupo-específico), dichas proteínas también cubren el RNA genómico. El gen *pol* codifica para varias proteínas (igual como sucede con los genes *gag* y *env*, una poliproteína única es sintetizada y luego dividida (ver figura 4)

Los productos del gen *pol* son:

- Transcriptasa inversa (RT) (una polimerasa que copia RNA en DNA)
- Integrasa (IN) (integra el genoma viral al genoma de la célula huésped)
- RNasa H (divide el RNA a medida que el DNA es transcrito de modo que la transcriptasa inversa pueda sintetizar la segunda cadena complementaria de DNA)
- Proteasas (PR) (dividen las poliproteínas traducidas de los mRNA de los genes *gag* y *pol*). Nota: esta es una proteína codificada también en los virus exógenos y es blanco de una nueva generación de drogas antivirales.



**Figura 4. Organización genómica de los retrovirus exógenos.** Los retrovirus incluyen en su genoma dos regiones LTR en sus extremos 5' y 3' y los genes *env*, *pol* y *gag* necesarios para la integración posterior a las células hospederas. El gen *env* codifica para las proteínas de membrana y las proteínas superficiales. El gen *pol* codifica para las enzimas retrotranscriptasa, integrasa, proteasa y RNasa H.

## 2. Antecedentes

En la actualidad, se ha demostrado que el MMTV causa tumores de mama en ratones de laboratorio. Diversos trabajos se han realizado para identificar un posible análogo humano del MMTV que pudiera causar cáncer en humanos. La mayor parte de los datos disponibles y su conclusión sugiere indicar la presencia de este tipo de analogía entre el cáncer de mama humano y los tumores murinos.

Uno de los estudios más ampliamente utilizado consiste en la identificación de las proteínas de la envoltura del MMTV en cáncer de mama humano mediante el uso de anticuerpos. En uno de estos trabajos, las reacciones positivas se observaron en 51 de 131 (39%) de los adenocarcinomas mamarios con resultados negativos en todas los tejidos normales de mama (Mesa-Tejada et al, 1978). Sin embargo, el uso de éstos anticuerpos contra la proteína de envoltura del MMTV ha sido motivo de disputa y dio lugar a la incertidumbre sobre la validez de estos resultados (Hareuveni y Torno, 1990).

Una de las principales dificultades surgidas en ese momento fue la identificación de los retrovirus endógenos humanos (HERVs) en el genoma humano. La gran similitud entre los HERVs y el MMTV ha hecho extremadamente difícil diferenciar entre los dos. Como consecuencia de ello, a principio los resultados fueron cuestionados con demandas que la identificación del MMTV en realidad puede ser un HERV.

Este problema, fue superado cuando Wang Y y colaboradores en 1995 aislaron una secuencia de 660 pb en adenocarcinomas mamarios, donde reportaban la detección de secuencias parecidas al gen *env* del MMTV en 121 (38.5%) de 314 muestras de cáncer de mama, en cultivos de líneas celulares MCF-7 y ED (2 de 10 líneas celulares mamarias), en 2(6.9%) de 29 fibroadenomas mamarios y en 2 (1.8%) de 107 muestras provenientes de reducción de mamoplastias (Wang y col., 1995). Estas secuencias resultaron con una similitud del 98% a la del gen *env* del MMTV y muy baja con secuencias HERVs. Diversos trabajos han sido publicados

en los cuales se ha empleado la estrategia de detectar secuencias de DNA o RNA del MMTV en tumores mamarios y tejido mamario no afectado de diversas poblaciones.

Etkind y colaboradores utilizaron en el año 2000 un enfoque similar y analizaron la presencia del MMTV mediante la amplificación de secuencias del gen *env* en los tumores de mama. La estrategia consistía en la amplificación de un fragmento de DNA de 250bp utilizando una PCR anidada. La secuencia se encontró en el 37% de las 73 muestras de los adenocarcinomas mamarios y en ninguno de los tejidos normales de mama analizados. La secuenciación de los amplificados confirmó que estas secuencias eran entre 99-100% similares a las secuencias del gen *env* de MMTV. Curiosamente, también identificaron estas secuencias en 3 de 19 linfomas no-Hodgkin (Etkind y col., 2000).

En un estudio posterior Wang Y y colaboradores en 1998 determinaron la expresión del gen *env* del MMTV en muestras con secuencias provirales positivas (muestras *env* positivas) y detectaron la amplificación de una secuencia de 660pb y la de una secuencia interna (250 pb) en 66% y 100% de las muestras analizadas, respectivamente. Estos resultados indican que el virus detectado en las muestras de adenocarcinomas tiene la capacidad de integrarse y expresarse en el tejido mamario humano (Wang y col, 1998).

Posteriormente en en 2001, Liu B y colaboradores detectaron la estructura proviral completa en dos tumores de mama humanos *env* positivos utilizando una estrategia de amplificación por PCR sobrelapada de segmentos LTR-gag, gag-pol, and pol-env. El provirus LTR-gag-pol-env-LTR de 9.9 kb manifestó una homología del 95% al MMTV y del 57% a los retrovirus endógenos de humanos (HERVs) en los genes gag y pol. Se analizó la secuencia amplificada de 660 pb del gen *env* y se encontró una homología del 90 al 98% con el gen *env* de MMTV. Estos resultados demuestran una vez más que el retrovirus MMTV es capaz de infectar células humanas y lograr la integración de su genoma completo en el genoma humano. La técnica empleada (caminata cromosómica) demostró ser eficaz al

lograr la amplificación completa del fragmento de 9.9 kb. El análisis realizado a las secuencias provirales demostró que contaban con las regiones reguladoras para lograr expresarse y replicarse. La región LTR amplificada contenía todos los elementos reguladores tales como la región enhancer y la región del promotor, características de una replicación competente del MMTV así como los elementos de respuesta a glucocorticoides y las secuencias codificantes para el superantígeno (*gag*).

Al mismo tiempo Melana y colaboradores en 2001, trataron de establecer como una fuente exógena la infección por MMTV mediante el análisis de una secuencia de 250 pb en tumores mamarios y los respectivos tejidos normales del mismo paciente. En las muestras pareadas de mama (normal y cáncer) de 106 pacientes encontraron que el 30% de los 106 cánceres de mama humanos y sólo 1 de 106 tejidos mamarios normales resultaron *env* positivos. Estos resultados indicaban que el tejido mamario no afectado de los pacientes con cáncer *env* positivos no contiene dicha secuencia y llevó a los autores a concluir que estas secuencias son de origen exógeno. Sin embargo, ellos no podían explicar el hallazgo de esta secuencia en el tejido mamario normal de un individuo. El correspondiente tejido mamario afectado de este paciente resultó ser *env* negativo. Después de varias pruebas, concluyeron que esto puede haberse debido a la posible contaminación con tejido mamario o con DNA tumoral (Melana y col., 2001).

En otro estudio, Ford y colaboradores en 2003, determinaron la presencia de secuencias *env* en sólo una de 120 muestras y en cero de 40 tejidos de adenocarcinomas mamarios a partir de la biopsias de mujeres vietnamitas y australianas, respectivamente. El mismo estudio mostró la secuencia en 19 de 45 cánceres de mama de las mujeres caucásicas (Ford y col., 2003). Anteriormente, esta secuencia se detectó en 37,7% de 70 bloques embebidos en parafina de adenocarcinomas mamarios de mujeres italianas (Pogo y col., 1999).

Ford y colaboradores en 2004, encontraron una relación entre el grado de malignidad del cáncer de mama y el porcentaje de secuencias relacionadas con

MMTV. Ellos detectaron 43 de 136 (32%) tumores con secuencias de MMTV de éstos tumores estudiados el 23, 34 y 38% se observaron en carcinoma ductal infiltrante de primer, segundo y tercer grado, respectivamente. Estos resultados suponen una relación de las secuencias de MMTV con la manifestación de tumores de mama y una asociación con el incremento de gravedad del cáncer.

Un estudio posterior, realizado por Zapata P y colaboradores determinó la presencia de secuencias *env* en cinco de 119 (4,2%) de biopsias de cáncer de mama de mujeres mexicanas. Ellos también reportaron que en la línea celular MCF-7 (que estaba reportada como control positivo) no se obtuvo amplificación de las secuencias *env* del MMTV (Zapata P y col., 2007).

Los rangos de detección son variables aunque es claro que hay una diferencia significativa en los porcentajes de detección entre el tejido neoplásico y el tejido mamario no afectado (30-60% en neoplasia mamaria, 1-4% en tejido mamario no afectado). Con estos estudios se ha demostrado que secuencias del retrovirus MMTV están presentes en la población mexicana, sin embargo no han determinaron los posibles sitios de inserción de estas.

### 3. Planteamiento del problema

Desde hace algunos años diversos estudios han demostrado que el MMTV puede infectar a las células mamarias humanas y se ha propuesto como una de las causas del cáncer de mama en humanos, pudiendo estar asociado a este retrovirus o algún homólogo (HMTV).

La existencia de secuencias HERVs en el genoma humano que contienen similitud parcial y capacidad de retrotransposición es una fuente enorme de variabilidad y plasticidad genómica, ya que afectan la expresión de genes cercanos debido a sus secuencias reguladoras (las LTRs); también se ha encontrado respuesta inmunológica contra el virus del MMTV, así como secuencias virales en neoplasias mamarias. Algunos investigadores han reportado asociación entre la etapa tumoral y el número de secuencias relacionadas con el MMTV. Por lo anterior consideramos necesario determinar la presencia de secuencias tipo *env* en muestras de adenocarcinomas mamarios, determinar cuáles son los posibles sitios de inserción de éstas secuencias retrovirales y evaluar la participación de los genes localizados en estas.

## 4. Justificación

El cáncer mamario es un problema de salud muy importante en nuestro país ya que ocupa el primer lugar como causa de muerte en mujeres mayores de 35 años. El estudio de las neoplasias se ha dificultado debido a la gran heterogeneidad histológica y clínica que presentan. Aunque se conocen algunos genes y proteínas asociados al origen del cáncer, la identificación y el entendimiento de los genes involucrados en estas patologías, sus interacciones génicas, las alteraciones bioquímicas y genéticas resultan de trascendental importancia. Por lo que el conocimiento en el ámbito molecular de los cambios necesarios para el avance del cáncer es una de las áreas de mayor desarrollo en la investigación biomédica actual. La amplia diversidad de tumores, aun los que derivan del mismo tejido, presenta un reto para un diagnóstico temprano y acertado. Debido a que los retrovirus pueden ocasionar re-arreglos genómicos que pudieran modificar la expresión de diversos genes, es importante averiguar si el MMTV/HMMT o algún homólogo están relacionados con el desarrollo de cáncer de mama en humanos debido a la inserción y posible mutagénesis insercional.

## 5. Objetivos

Objetivo general: Determinar si el retrovirus MMTV o algún homólogo humano están relacionados con el desarrollo de cáncer de mama en humanos debido al fenómeno de mutagénesis insercional.

Objetivos particulares:

- Identificar y determinar secuencias tipo *env* del retrovirus HMTV/MMTV en muestras tumorales mamarias.
- Identificar secuencias LTRs del retrovirus HMTV/MMTV en muestras tumorales mamarias mediante el método de la PCR *splinkerette*.
- Identificar posibles efectos biológicos de la mutagénesis insercional de las LTRs y la participación de los genes del retrovirus HMTV / MMTV en el desarrollo y avance de la neoplasia.
- Identificar mediante una búsqueda informática secuencias *env* en el genoma humano y correlacionar con la presencia de secuencias en las muestras tumorales MMTV/HMTV.

## 6. Materiales y métodos

### 6.1 Obtención de las muestras de adenocarcinomas

Se obtuvieron muestras de adenocarcinomas mamarios de mujeres mexicanas referidas del hospital de Oncología, Centro Médico Nacional S-XXI, IMSS. Los tejidos fueron transportados al laboratorio en un contenedor con nitrógeno líquido. Las muestras se almacenaron a  $-70^{\circ}\text{C}$  hasta antes de usarlas.

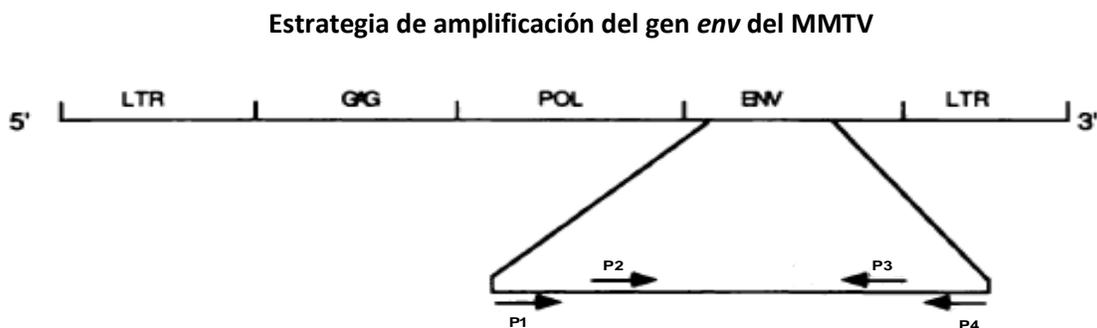
### 6.2 Obtención de DNA a partir de las muestras obtenidas

Para la extracción de DNA a partir de tumor, se pesaron aproximadamente 100mg de cada muestra, se seccionó en pequeñas piezas, con ayuda de un bisturí, el tejido congelado en presencia de nitrógeno líquido fue pulverizado sobre un mortero, posteriormente el tejido pulverizado fue recuperado, utilizando proteínasa K y el paquete comercial QIAamp DNA Mini Kit (Quiagen) y siguiendo las especificaciones del fabricante, se llevó a cabo la extracción del material genético en un sistema automatizado de purificación de ácidos nucleicos QIAcube (Quiagen), para asegurar concentraciones y pureza similar entre las muestras. Se verificó la concentración e integridad del DNA de las muestras mediante espectrofotometría (nanodrop technologies) y electroforesis en geles de agarosa. Las muestras de DNA se almacenaron a  $-70^{\circ}\text{C}$ , hasta su uso.

### 6.3 PCR anidada para la amplificación del gen *env*

La detección de las secuencias MMTV/HMTV se realizaron de acuerdo a los métodos reportados por Zapata-Benavides y col, 2007, la amplificación con oligonucleótidos específicos se realizó flanqueando las secuencias del gen *env* del MMTV. La amplificación se realizó en un termociclador PalmCycler (tiempo final). Todas las reacciones se realizaron por duplicado y se incluyeron controles negativos sin molde y positivos de reacción.

La estrategia de amplificación fue la siguiente: se utilizaron los primers 1 y 4 (CCTCACTGCCAGATC, GAATCGCTTGGCTCG) para amplificar un fragmento del gen *env* de 685 pb. (56 °C para la alineación, 35 ciclos), posteriormente usando el producto de la reacción como molde se utilizaron los primers 2 y 3 (TACATCTGCCTGTGTGTTAC, ATCTGTGGCATACT) para amplificar un producto de 250 pb. (ver figura 5).



**Figura 5. Estrategia de amplificación del gen *env* del MMTV.** Se amplificó un producto de 686 pb. Utilizando los oligonucleótidos P1 y P4 (sobre las bases 685 y 1371 del gen *env* respectivamente), posteriormente se utilizaron los oligonucleótidos P2 y P3 (sobre las bases 1098 y 1350 del gen *env* respectivamente) para obtener un producto de 252 pb.

Para ambas amplificaciones se utilizó la siguiente mezcla de reacción:

Mezcla de reacción y programa de ciclado		Componente	Concentración Final
<b>1er Paso</b>	94 °C 10 min	Buffer de reacción 10X	1X
		MgCl <sub>2</sub> 50mM	1.5 mM
<b>2do Paso</b>	35 ciclos {	dNTP's 10mM	2 mM
		Iniciador F 20 pmol	20 pmol
		Iniciador R 20 pmol	20 pmol
<b>3er Paso</b>	72 °C 8 min	Taq pol (5u/μL)	1 U
		DNA 200 ng	200ng
<b>Almacenar a 4 °C</b>		Agua	X
		Volumen de reacción	30 μL

Como control positivo de reacción se utilizó la construcción hecha con el gen env del MMTV cepa C3H sobre el plásmido pBR322 y propagado en E.coli HB101 (ATCC 45006), así mismo se amplificó un fragmento de 500pb del gen GAPDH para verificar la utilidad de los DNAs.

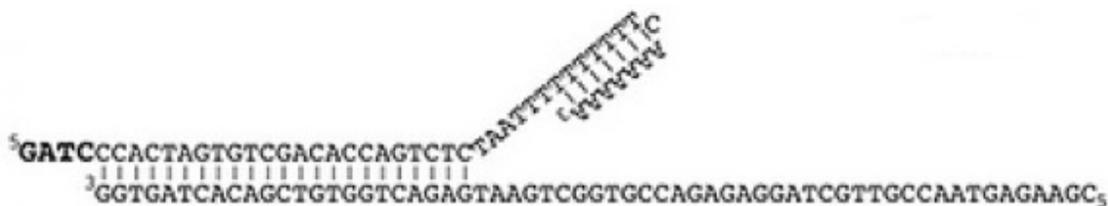
#### 6.4 PCR splinkerette para amplificar las secuencias flanqueadas por las LTRs virales

Para detectar e identificar los sitios de inserción del MMTV/HMTV flanqueados por las secuencias LTR se utilizó el método de PCR splinkerette reportado por Theodorou V y col. en el 2007, el cual se describe a continuación:

A partir de 1 mg de DNA tumoral se realizó una digestión con la enzima *XhoI* (6 hrs a 37 °C, 5 U) cuya secuencia de reconocimiento es 5'...C<sup>^</sup>TCGAG...3'. Posteriormente se ligaron 300 ng del DNA digerido con *XhoI* con el "splinkerette linker" en una proporción molar de 1:15 con con 2 unidades de T4 DNA ligasa durante 5 hrs a 37 °C . El adaptador (splinkerette linker) fue generado utilizando dos oligonucleótidos (ver figura 6), HMSpA y HMSpBD cuyas secuencias son las siguientes:

HMSpA           CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTGCGACATCGATGG  
HMSpBD         GATCCCATCGATGTCGACACCAGTCTCTAATTTTTTTTTTTTCAAAAAA

Estructura del splinkerette generado



**Figura 6. Estructura generada para el adaptador (splinkerette linker).** El adaptador fue generado apartir de los oligonucleotidos HMSpA y HMSpBD. La estructura generada brinda una mayor especificidad en la reacción de amplificación debido a la estructura formada en el extremo 3' del oligonucleotido HMSpBD.

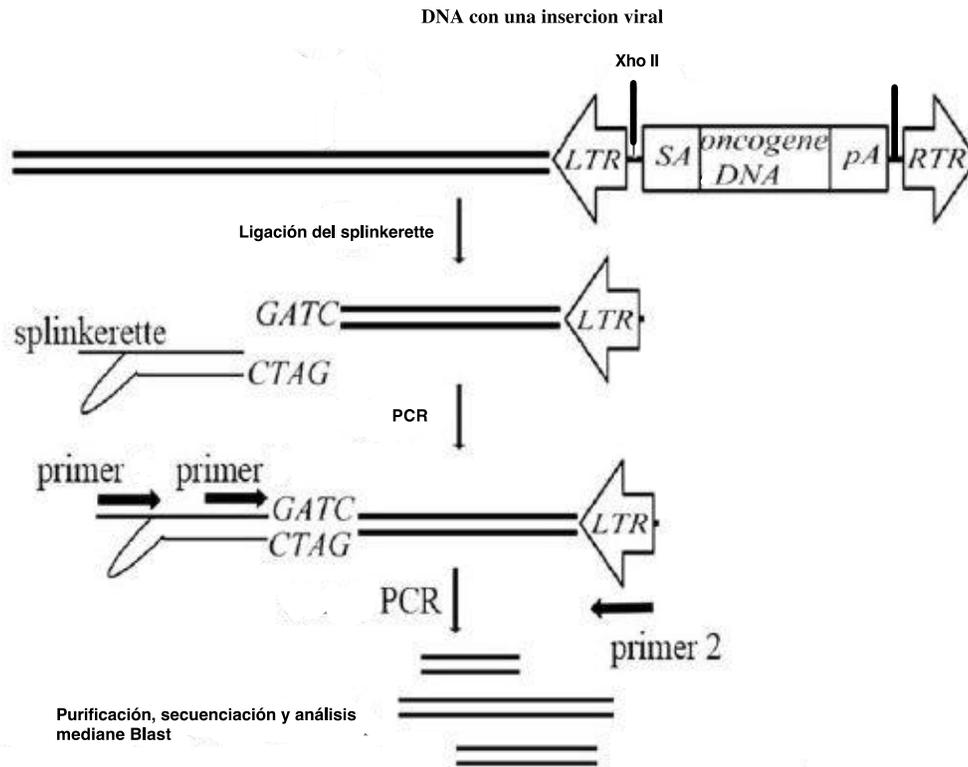
El adaptador fue generado utilizando 150 pmol de cada oligonucleótido e incubando la mezcla a 94 °C durante 3 minutos. Posteriormente se llevó a temperatura ambiente a una tasa de 1 C por 15 segundos. Una vez realizada la reacción de ligación las muestras fueron digeridas con la enzima *DraI* (2 hr. De incubación 37 °C, 5 U). Después se realizó la purificación del producto de digestión en una columna Microcon YM-30 (Amicon BioSeparations) de acuerdo al manual de usuario. Se utilizó el DNA ligado para la amplificación utilizando los oligonucleótidos LTR-beg2 (que se alinea a la LTR viral) y 4HMSP1a (que se alinea al enlazador splinkerette) (68 °C de alineación, 35 ciclos) cuya secuencias se muestran en seguida:

LTR-beg2+6: GTCTTTGTCTGATGGGCTCATCCGTTTGTG  
 LTR-beg4-C3H+1: CCCTAGGTGTAGGACACTCTCG  
 4HMSP-1a: AGAGTAACCGTTGCTAGGAGAG

Posteriormente se reamplificaron los productos de PCR con el oligonucleótido LTR-beg4-C3H+1 y 4HMSP1a (11 ciclos de alineación a 68 °C seguido de 11 ciclos de 67 °C y 13 ciclos en 66 °C). Después de la electroforesis, los productos fueron purificados y reamplificados para ser enviados al servicio de secuenciación de la FES-UNAM Iztacala a cargo del M en C Alejandro Monzalvo (ver figura 6).

Para ambas amplificaciones se utilizó la siguiente mezcla de reacción:

Mezcla de reacción y programa de ciclado		Componente	Concentración Final
<b>1er Paso</b>	94 °C 10 min	Buffer de reacción 10X	1X
		MgCl <sub>2</sub> 50mM	1.5 mM
<b>2do Paso</b>	30 ciclos { 94 °C 30 seg 68 °C 30 seg 72 °C 30 seg	dNTP's 10mM	2 mM
		Iniciador F 20 pmol	20 pmol
		Iniciador R 20 pmol	20 pmol
<b>3er Paso</b>	72 °C 8 min	Taq pol (5u/μL)	1 U
		DNA 200 ng	200ng
<b>Almacenar a 4 °C</b>		Agua	X
		Volumen de reacción	30 μL



**Figura 7. Esquema del método de PCR splinkerette.** El método consisten en digerir con la enzima *Xho II* el DNA asociado a la LTR viral del MMTV, posteriormente se genera un adaptador llamado “Splinkerette” que es ligado al DNA digerido previamente con *Xho II*, después de ser purificado y desalado el DNA ligado al adaptador es amplificado con primers específicos tanto para el adaptador como para la LTR viral.

## 6.5 Búsqueda Bioinformática

Para identificar posibles sitios de inserción del retrovirus MMTV/HMTV en el genoma humano se realizó una búsqueda bioinformática en la base de datos del NCBI y del EMBL-EBI de acuerdo en la siguiente estrategia:

1. Se descargó la secuencia completa del gen *env* del MMTV y las secuencias LTR 5' y 3' del MMTV de las páginas del NCBI y del EMBL-EBI.
2. El número de acceso para la secuencia completa del MMTV fue NC\_001503 (NCBI)

3. Posteriormente, cuando las secuencias fueron descargadas se procedió a realizar una búsqueda de similitud de secuencias mediante la herramienta Blast (programa nucleotide Blast, algoritmo Blastn, base de datos Human genomic plus transcript), para cada una de las secuencias descargadas.

4. Los parámetros de búsqueda fueron los siguientes:

Número máximo de secuencias: 10 000

Valor de corte (Expect threshold): 10

Tamaño de palabra (Word size): 11

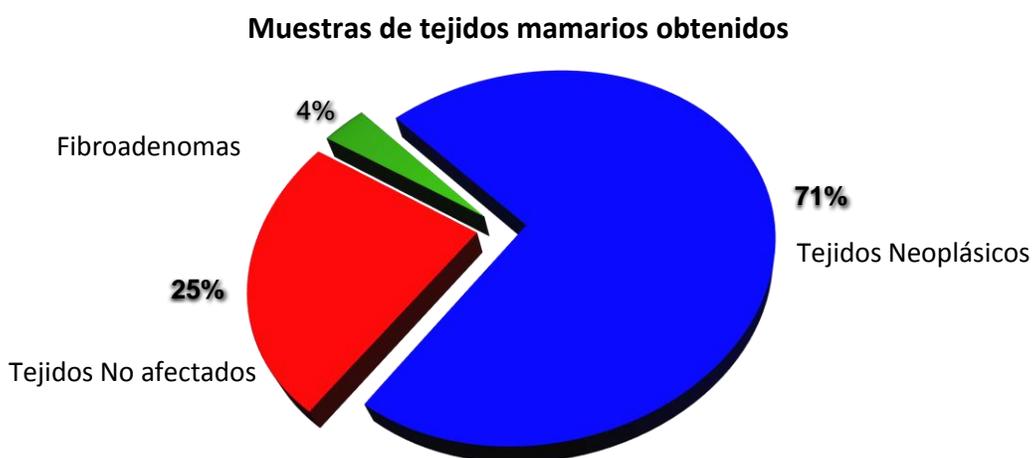
Puntaje Match/Mismatch: 2,-3

Puntaje gaps (gaps costs): inicio 5, extensión 2

## 7. Resultados

### 7.1 Obtención de las muestras de adencarcionamas

Se obtuvieron 56 muestras de adenocarcinomas mamarios y tejidos no afectados de mujeres mexicanas referidas al hospital de Oncología, Centro Médico Nacional S-XXI del IMSS. Los porcentajes de los tejidos obtenidos fueron de la siguiente manera: 40 tejidos neoplásicos (71%), 2 fibroadenomas (4%) y 16 tejidos mamarios no afectados (25%) (ver figura 8).



**Figura 8. Muestras de tejidos mamarios obtenidos (adenocarcinomas y tejidos mamarios no afectados).** Se obtuvieron 56 muestras de tejido mamario. El 71% de las muestras obtenidas corresponde a 40 tejidos neoplásicos, el 4% a 2 fibroadenomas y el 25% a tejidos mamarios no afectados.

### 7.2 Obtención del DNA a partir de las muestras obtenidas

Para la extracción de DNA a partir de cada muestra, se pesaron aproximadamente 100mg de cada muestra y se procesaron de acuerdo a los descrito en los métodos, se asignó una clave de identificación a cada muestra. En la tabla 2 y 2A, se pueden observar los datos pertenecientes a cada muestra así como las claves asignadas.

**Tabla 2. Relación de las muestras obtenidas.** En la tabla se muestran los datos relacionados con las muestras obtenidas, concentración y cociente 260/280. Dx= diagnostico, CDI= Carcinoma Ductal Infiltrante, CLI= Carcinoma Lobulillar infiltrante, CSC= Carcinoma sin clasificar, TNA = Tejido no afectado, FBA= fibroadenoma

Muestra	Dx	Concentración (ng/μl)	Cociente 260/280	Muestra	Dx	Concentración (ng/μl)	Cociente 260/280
001CM	CDI	345.54	1.67	016CM	CDI	432.13	1.88
002CM	CDI	453.54	1.84	017CM	CDI	984.23	1.89
003CM	CDI	314.41	1.78	018CM	CDI	431.31	1.80
004CM	CDI	434.31	1.98	019CM	TNA	654.14	1.93
005CM	TNA	414.63	1.85	020CM	FBA	354.75	1.98
006CM	TNA	435.13	1.67	021CM	CDI	564.13	1.85
007CM	CDI	764.89	1.90	022CM	CDI	765.34	1.67
008CM	TNA	967.65	1.84	023CM	CDI	345.13	1.90
009CM	CLI	673.65	1.82	024CM	CDI	876.34	1.84
010CM	CDI	865.33	1.88	025CM	TNA	344.33	1.67
011CM	TNA	978.56	1.89	026CM	CDI	431.34	1.90
012CM	CDI	537.46	1.80	027CM	CDI	765.34	1.84
013CM	CDI	387.36	1.93	028CM	CDI	982.34	1.82
014CM	CDI	845.64	1.92	029CM	TNA	650.78	1.83
015CM	CSC	532.76	1.90	030CM	TNA	433.14	1.88

**Tabla 2A. Continuación**

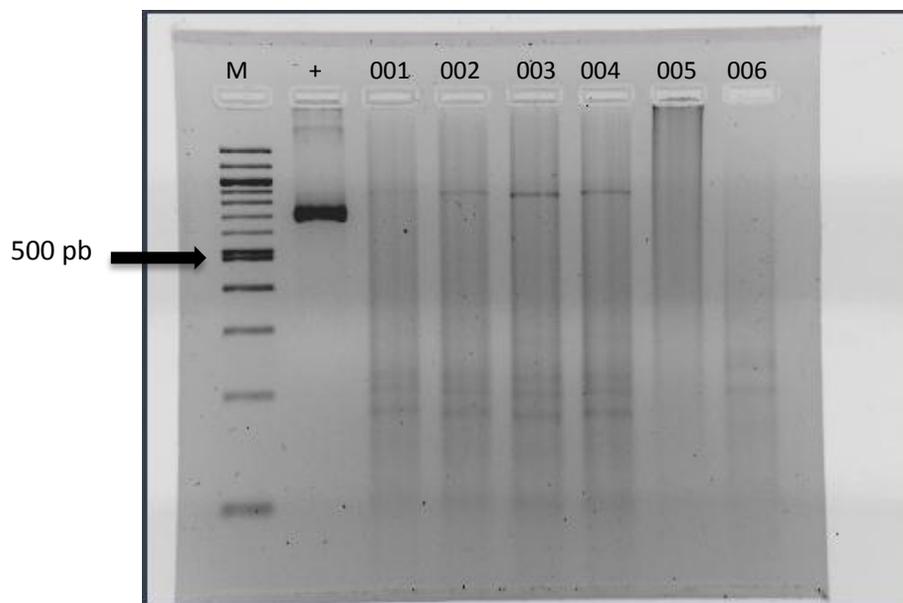
Muestra	Dx	Concentración (ng/μl)	Cociente 260/280	Muestra	Dx	Concentración(ng/μl)	Cociente 260/280
031CM	CDI	865.34	1.68	046CM	TNA	563.78	1.90
032CM	CDI	525.80	1.84	047CM	CDI	764.65	1.84
033CM	TNA	452.00	1.78	048CM	CDI	674.52	1.78
034CM	CLI	553.00	1.98	049CM	TNA	845.31	1.90
035CM	CLI	230.00	1.85	050CM	CDI	665.23	1.85
036CM	CDI	748.23	1.67	051CM	CDI	765.23	1.67
037CM	CDI	423.64	1.72	052CM	CLI	643.34	1.84
038CM	CDI	242.95	1.82	053CM	CDI	454.62	1.88
039CM	CDI	562.33	1.67	054CM	TNA	122.63	1.90
040CM	CDI	456.63	1.90	055CM	CDI	754.24	1.81
041CM	FBA	765.34	1.84	056CM	CDI	143.75	1.77
042CM	CDI	876.24	1.82				
043CM	CDI	145.65	1.88				
044CM	TNA	546.87	1.96				
045CM	TNA	254.86	1.78				

Una vez extraídos los DNA de cada muestra fueron etiquetados y guardados a -30 °C hasta la su utilización.

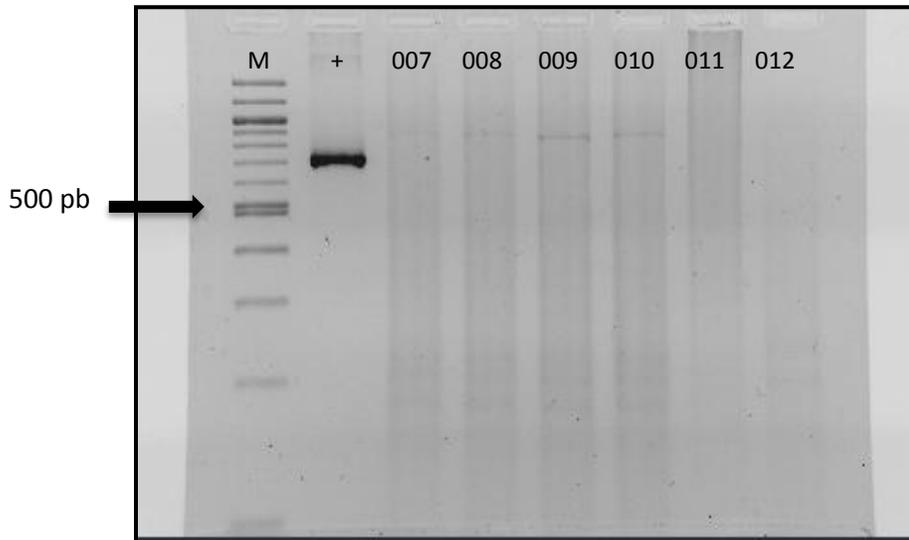
### 7.3 PCR anidada para la amplificación del gen *env*

La PCR anidada para la detección de secuencias *env* del MMTV en las muestras obtenidas fue realizada de acuerdo a los métodos reportados por Zapata-Benavides y col, en 2007. A partir de 200 ng de DNA se realizó la primer PCR utilizando los oligonucleotidos CCTCACTGCCAGATC y GAATCGCTTGGCTCG para amplificar un fragmento del gen *env* de 685 pb. de acuerdo a los métodos descritos anteriormente. Una vez realizada la PCR los productos de amplificación

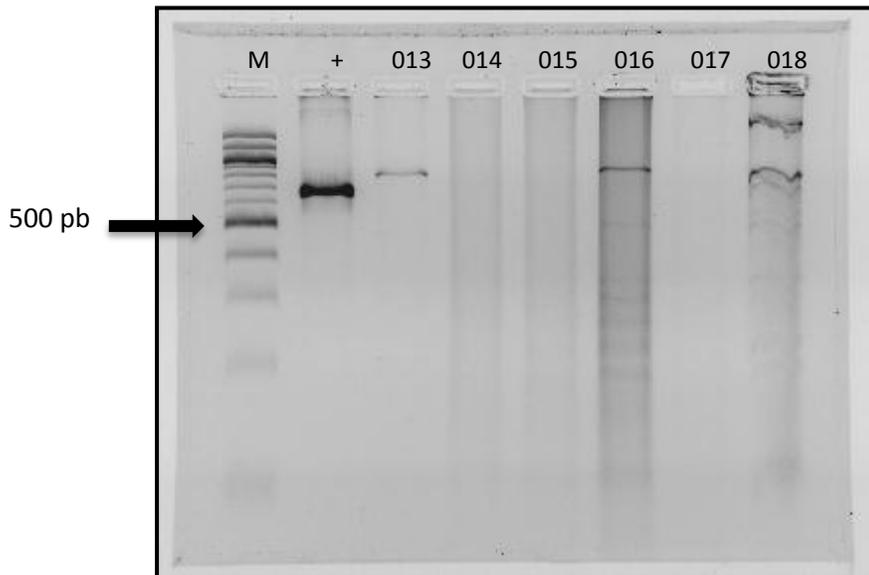
fueron separados en geles de agarosa al 1.2% teñidos con SYBER green. En las figuras 9-13 se pueden observar los electroferogramas de las muestras 001CM-029CM. En la figura 9 es posible observar el marcador de escalera de 100 pb (carril M), el control positivo de reacción que corresponde al producto de PCR con el DNA del plásmido hecho con la construcción del gen *env* del MMTV C3H (carril +) y los amplificados de las muestras 001CM-006CM. En la figura es posible observar un amplificado en el carril (+) correspondiente al fragmento de aproximadamente 685 pb del gen *env* del MMTV y amplificados en las muestras 001CM-004CM



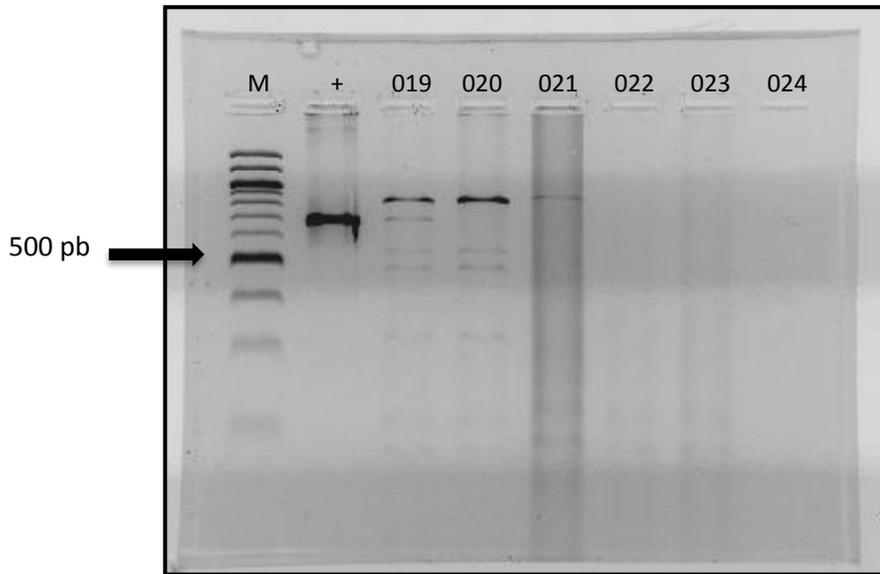
**Figura 9. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 001CM al 006CM del gen *env* del MMTV.** En la figura se muestran los amplificados de las muestras 001CM al 006CM, M= escalera de peso molecular de 100pb, el control positivo de reacción (+) corresponde al amplificado de la construcción hecha con el gen *env* del MMTV C3H y los amplificados para las muestras 001CM-006CM.



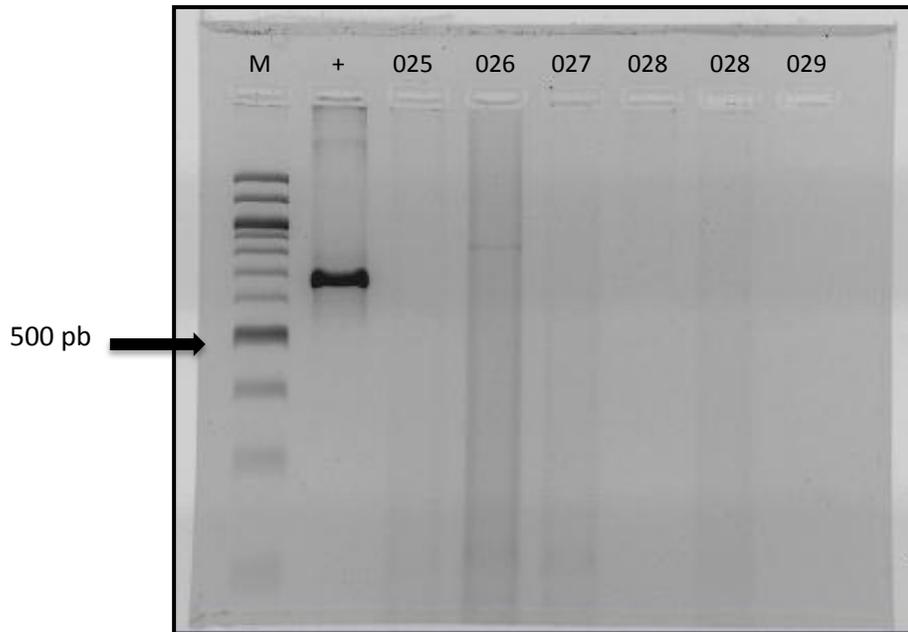
**Figura 10. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 007CM-012CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 007CM al 012CM, M= escalera de peso molecular de 100pb, el control positivo de reacción (+) corresponde al amplificado de la construcción hecha con el gen *env* del MMTV C3H y los amplificados para las muestras 007CM-012CM.



**Figura 11. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 013CM-018CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 013CM al 018CM, M= escalera de peso molecular de 100pb, el control positivo de reacción (+) corresponde al amplificado de la construcción hecha con el gen *env* del MMTV C3H y los amplificados para las muestras 013CM-018CM.



**Figura 12. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 019CM-024CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 019CM al 024CM, M= escalera de peso molecular de 100pb, el control positivo de reacción (+) corresponde al amplificado de la construcción hecha con el gen *env* del MMTV C3H y los amplificados para las muestras 019CM-024CM.

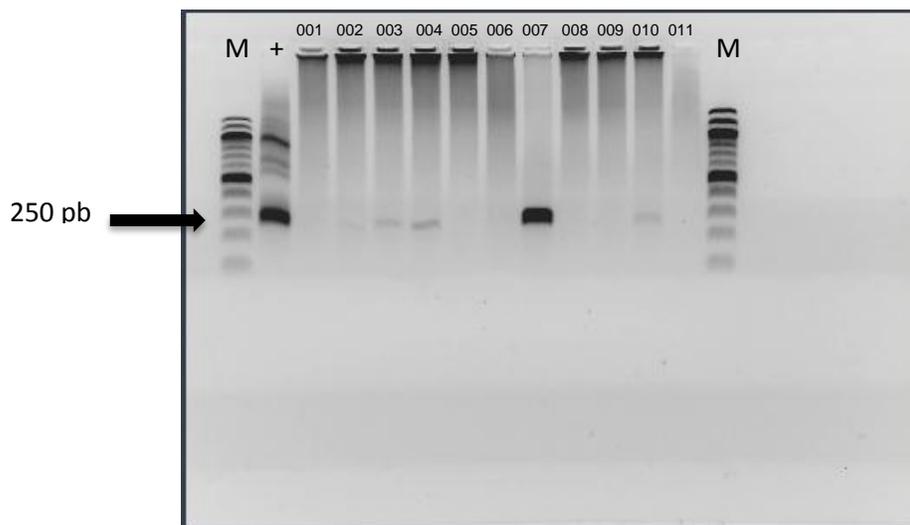


**Figura 13. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 025CM-028CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 025CM al 029CM, M= escalera de peso molecular de 100pb, el control positivo de reacción (+) corresponde al amplificado de la construcción hecha con el gen *env* del MMTV C3H y los amplificados para las muestras 025CM-029CM.

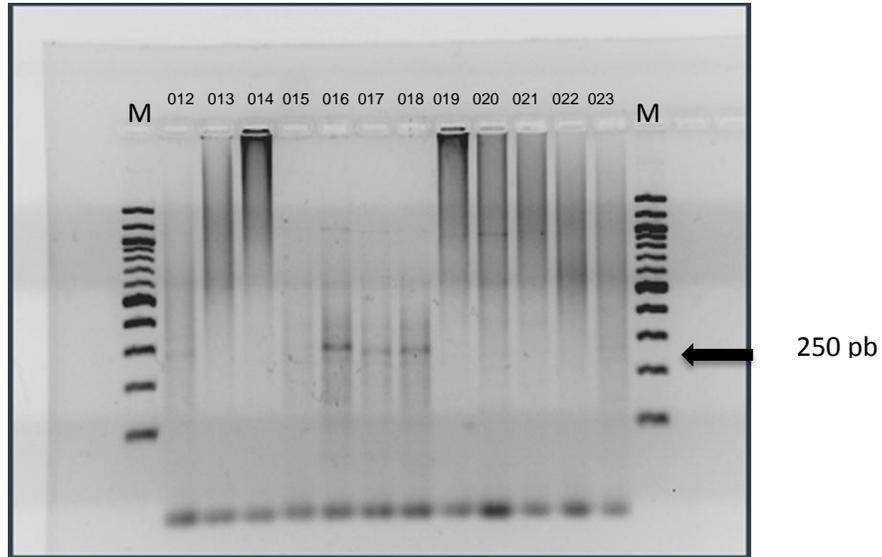
Una vez que las muestras fueron amplificadas se procedió a realizar la segunda PCR con los oligonucleotidos (TACATCTGCCTGTGTGTTAC y ATCTGTGGCATAACCT) para amplificar un producto de 250 pb la reacción se realizó de acuerdo a los métodos descritos anteriormente.

En las figuras 14-17 se pueden observar los electroferogramas de las muestras 001CM-029CM para el amplificado de 250pb. Los correspondientes geles de agarosa fueron realizados al 1.2% y teñidos con Syber green.

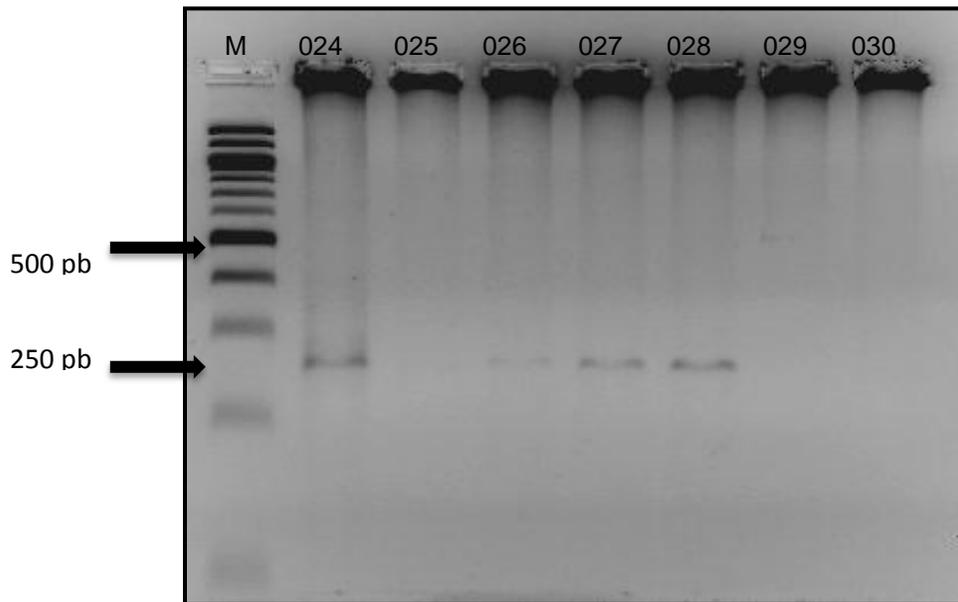
En la figura 14 es posible observar un marcador de escalera de 100 pb (carril M), el control positivo de reacción que corresponde al producto de la primer PCR con el DNA del plásmido hecho con la construcción del gen *env* del MMTV C3H (carril +) y los amplificados de las muestras 001CM-011CM. Se observan amplificados en el carril (+) correspondiente al fragmento de aproximadamente 250 pb del gen *env* del MMTV C3H y amplificados en las muestras 002CM, 003CM, 004CM, 007CM y 010CM de aproximadamente 250 pb.



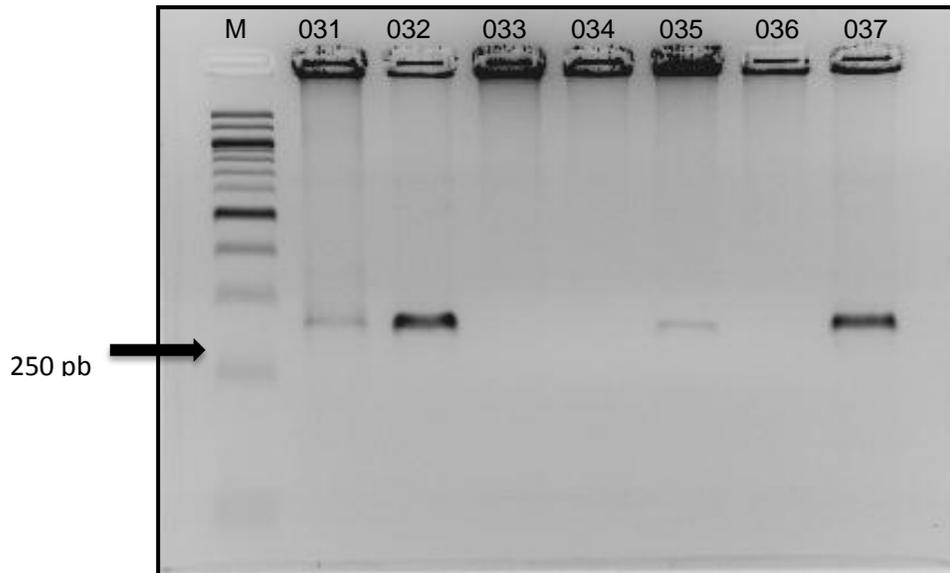
**Figura 14. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 001CM-011CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 001CM al 011CM, M= a marcador de Peso molecular de 100pb, el control positivo de reacción(+) corresponde al amplificado de la construcción hecha con el gen *env* del MMTV C3H. Se observan amplificados de las muestras 002CM, 003CM, 004CM, 007CM y 010CM.



**Figura 15. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 012CM-023CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 012CM al 023CM, M= a marcador de Peso molecular de 100pb. Es posible observar amplificados en las muestras 012CM, 016CM, 017CM, 018CM.



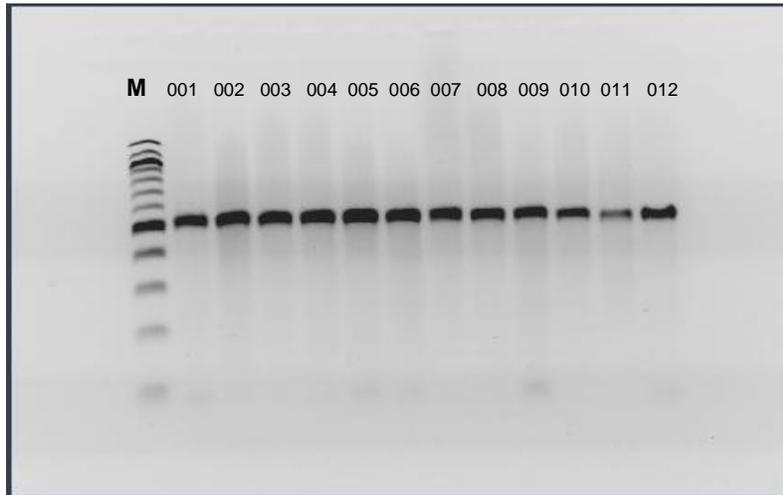
**Figura 16. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 024CM-030CM del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 024CM al 030CM, M= a marcador de Peso molecular de 100pb. Es posible observar amplificados en las muestras 024CM, 026CM, 027CM, 028CM.



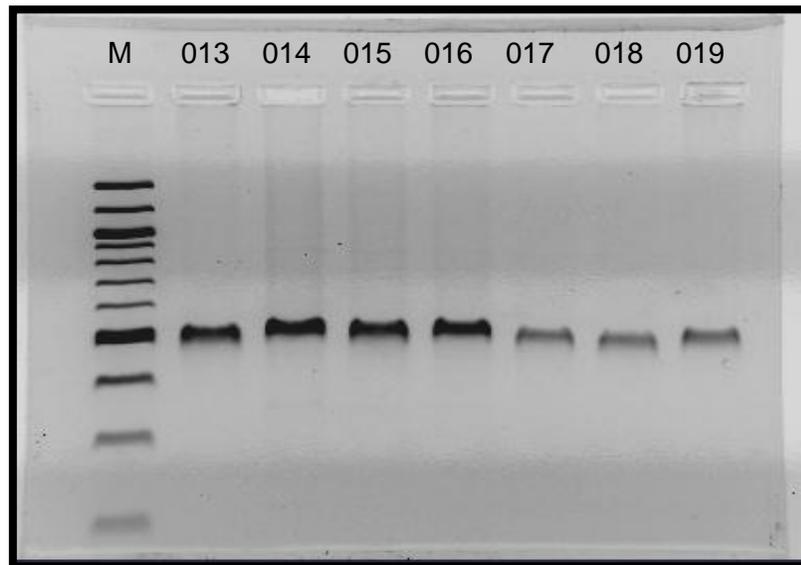
**Figura 17. Electroforesis en gel de agarosa al 1.2 % para los amplificados de las muestras 031CM-037M del gen *env* del MMTV.** En la figura se muestran los amplificados las muestras 031CM al 037CM, M= a marcador de Peso molecular de 100pb. Es posible observar amplificados en las muestras 031CM, 032CM, 035CM, 037CM.

Al mismo tiempo fueron realizadas las reacciones de amplificación para un fragmento de 500 pb del gen GAPDH mediante la metodología descrita anteriormente. Posteriormente los amplificados fueron separados en gels de agarosa al 1.2% teñidos con Syber green.

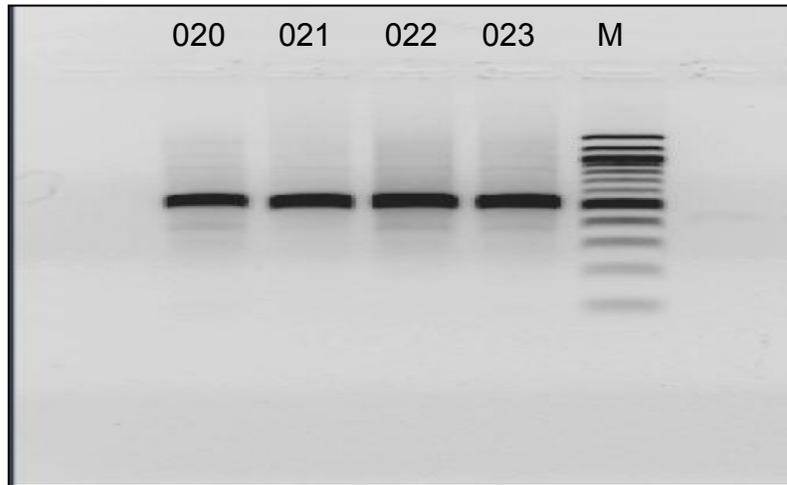
Como se puede observar en las figuras 18-22 se observan amplificados de aproximadamente 500pb en las muestras empleadas en el estudio. En la figura 18 es posible observar los amplificados de las muestras 001CM-012CM de aproximadamente 500 pb. demostrando la utilidad de los DNAs extraídos tanto para los tejidos tumorales (muestras 001CM-004CM, 007CM, 010CM, 012CM) y los tejidos no afectados (006CM, 008CM, 011CM)



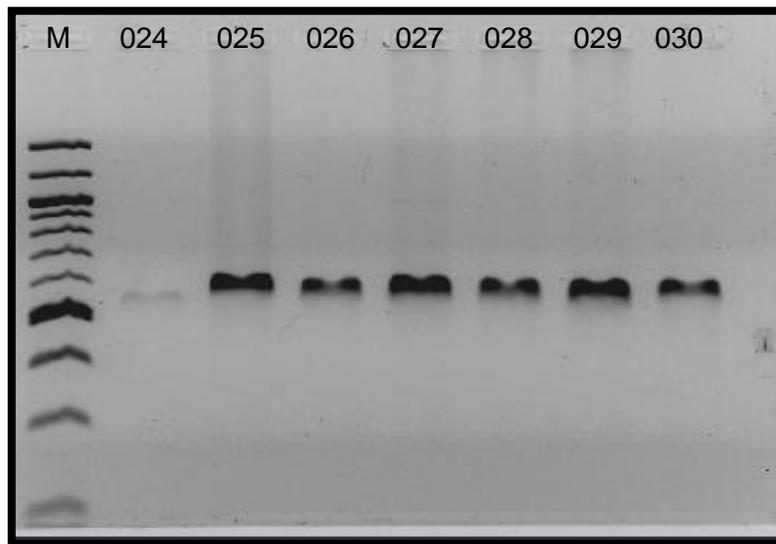
**Figura 18. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 001CM-012CM del gen GAPDH.** En la figura se muestran los amplificados las muestras 001CM al 012CM de aproximadamente 500pb , M= escalera de peso molecular de 100pb



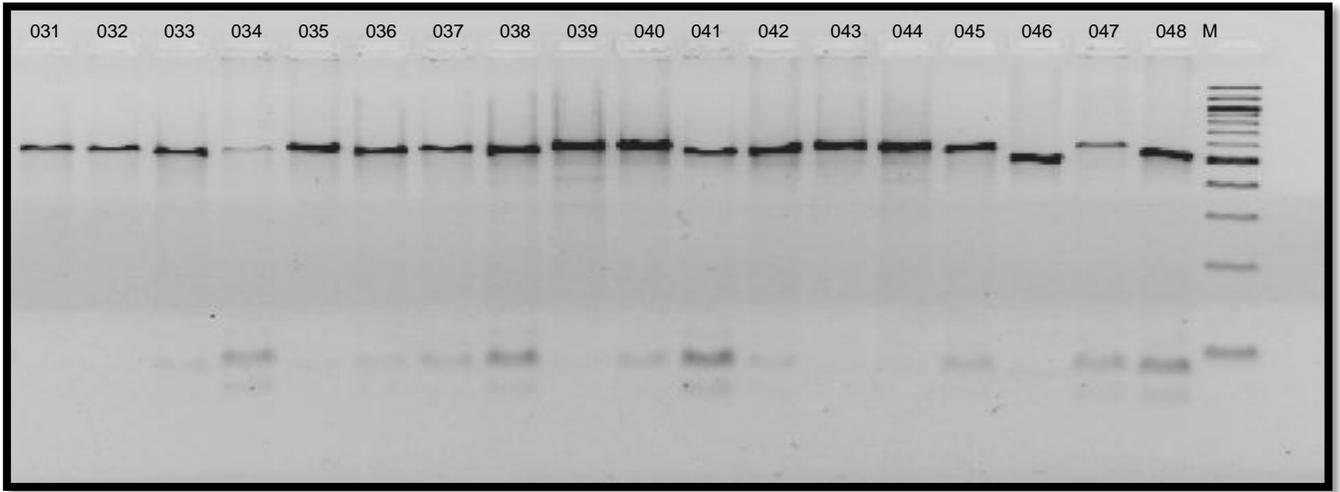
**Figura 19. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 013CM-019CM del gen GAPDH.** En la figura se muestran los amplificados las muestras 013CM al 019CM de aproximadamente 500pb , M= escalera de peso molecular de 100pb



**Figura 20. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 020CM-023CM del gen GAPDH.** En la figura se muestran los amplificados las muestras 020CM al 023CM de aproximadamente 500pb , M= escalera de peso molecular de 100pb

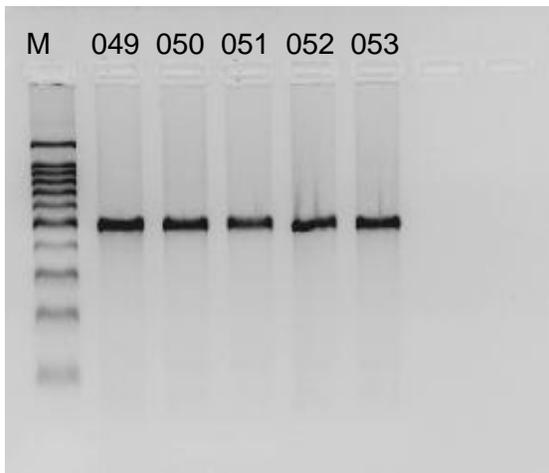


**Figura 21. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 024CM-030CM del gen GAPDH.** En la figura se muestran los amplificados las muestras 024CM al 030CM de aproximadamente 500pb , M= escalera de peso molecular de 100pb

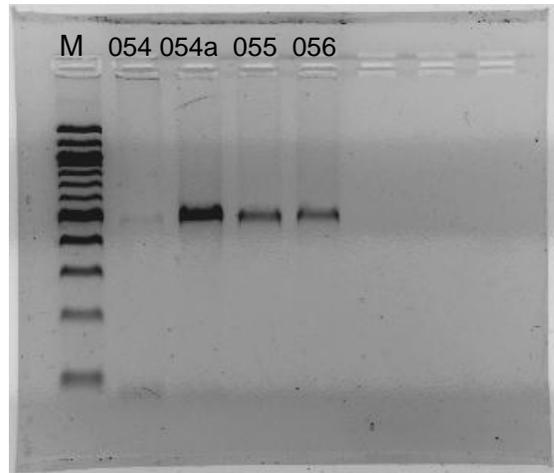


**Figura 22. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 031CM-049CM del gen GAPDH.** En la figura se muestran los amplificados las muestras 031CM al 049CM de aproximadamente 500pb , M= escalera de peso molecular de 100pb

**A)**



**B)**



**Figura 23. Electroforesis en gel de agarosa al 1.2% para los amplificados de las muestras 049CM-056CM del gen GAPDH.** En las figuras A) y B) se muestran los amplificados las muestras 049CM-056CM de aproximadamente 500pb , M= escalera de peso molecular de 100pb. En la figura B) se pueden observar los carriles 054 y 054a que corresponden a la misma muestra (muestra 54CM)

Como se puede observar en las figuras 13-16 se detectó un amplificado de 250 pb del gen *env* del MMTV en 16 muestras de adenocarcinomas mamarios (muestras 002CM, 003CM, 004CM, 007CM, 010CM, 012CM, 016CM, 017CM, 018CM, 024CM, 026CM, 027CM, 028CM, 031CM, 032CM, 035CM) lo que representa el 40% de los 40 adenocarcinomas mamarios estudiados, así mismo no fue posible detectar el amplificado en los 2 fibroadenomas muestreados ni en los 14 tejidos no afectados. Las 56 muestras analizadas en el presente estudio amplificaron un fragmento de 500 pb del gen GAPDH.

#### 7.4 Purificación y análisis de las secuencias obtenidas a partir de los amplificados para el gen *env*

Posteriormente, los amplificados de las muestras analizadas fueron purificados, tanto para el gen *env* del MMTV (muestras 002CM, 004CM, 007CM, 027CM, 031CM 032CM) como el control positivo (amplificado del plásmido del MMTV C3H) y el control de reacción (fragmento de 500pb de gen GAPDH de las muestras 001CM y 002CM), los amplificados fueron enviados al servicio de secuenciación de la FES-UNAM Iztacala a cargo del M en C Alejandro Monzalvo. Una vez realizadas las reacciones de secuenciación las secuencias obtenidas se pueden observar en las figuras 24 y 25.

En la figura 24 se pueden observar las secuencias obtenidas de los amplificados de las muestras de adenocarcinomas 002CM, 004CM, 007CM y 027CM así como del control positivo (plásmido con el gen *env* de MMTV C3H). La secuencia que se obtuvo para el control positivo mostro un tamaño de 291 pb mientras que las muestras 002CM y 004CM resultaron de 286 y 254pb respectivamente. La secuencia obtenida para la muestras 007CM fue de 242pb mientras que la secuencia de la muestra 027CM fue de 318 pb.

```

>Plasmido_MMTV (291 pb)
CTCCTTGGGGCAAATTTATTAGGATTACCTCAGCTATAGATATAGAGAAAG
AGGATCTACTTTTCATATTTCTGTCTTCTTGTAGATTGACTAATTGTT
TAGATTCTTCTGCCTACGACTATGCAGCGATCATAGTCAAGAGGCCGCCA
TACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGATTC
TGCCATTCAAACCTTTAGGTATGCCACAGATATAAATTCGAGCCAAGCGAT
TCAGGCCTCACTGCCAGATCAATTAAGGAAAGGAAGAGGGAA

>MMTV-002CM (286 pb)
TCCTTTAAACCAAATTTATTAAGGATTACCTCAGCTATAGATCTAGAGAA
AAGAGGATCTACTTTTCATATTTCTGTCTTCTTGTAGATTGACTAATT
GTTTAGATTCTTCTGCCTACGACTATGCAGCGATCATAGTCAAGAGGCCG
CCATACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA
TTCTGCCATTCAAACCTTTAGGTATGCCACAGATATAAATTCGAGCCAAGCG
ATTCAGCCTCCTGCAGACCCAAAAAACAGGGGAT

>MMTV-004CM (235 pb)
GGCCCTTTTGGCCAAATTTTAAGGATTACCTCAGCTATAGATATAAGAGA
AAAGAGGATCTACTTTTCATATTTCTGTCTTCTTGTAGATTGACTAAT
TGTTTAGATTCTTCTGCCTACGACTATGCAGCGATCATAGTCAAGAGGCC
GCCATACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATG
ATTCTGCCATTCAAACCTTTAGGTATGCCACAGAT

>MMTV-007CM (242)
GGAGCAGACTCGTCTCCCTCCCAGAAGGCGTCTTCTTAAAGGCGATCTG
GCAGTGAGGACCTGAAATCGCTTGACTCGAATTAATGTGTGGCAGTGAA
AAGCTTGGGTCTGGTGTAATGATCAGCCCTGCTTCCCTCTCGATATGTCT
AAAAAGATGAGCCGTCTTGCTCTCCTTCTTAATTCATACTGCAGAATAAT
CTGAAAGGTGAAATCTTTATCCCTTCAGAAAGGCACAAAAA

>MMTV-027 (318)
GATGATGGCCGAATCATCAGCCATGGTTTCATCACCAATATCTACAGGTAG
CAGCACGTATGGCGGCCCTTTGACTATGATCGCTGCATAGTCGTAGGCAG
AAGAATCTAAACAATTAGTCAATCTACAAGAAGAACAGGAAATATGAAAA
GTAGATCCTCTTTCTCTATATCTATTAGCTGAGGTAATCCTAATAATAT
GGCATAAGGGTAAGTAACACAGGCCGATGTATAAAATTCGAGCCAAGCGA
TTCAGGTCTCCCTGCCAGATCCACCTAGCCAGGGCGTCACCGCCAAACC
AGAGGAGTGGAGGCGGGG

```

**Figura 24** Secuencias obtenidas de los amplificadores de las muestras de adenocarcinomas 002CM, 004CM, 007CM y 027CM así como para el control positivo. En la figura se pueden observar las secuencias obtenidas a partir del amplificador detectado para el gen *env* del MMTV.

```

>MMTV-31 (292 pb)
GCGCTTGAGGGAAAATTATTAGGATTACCTCAGCTAATAGATATAAGAGA
AAAGAGGATCTACTTTTCATATTTCTGTCTTCTTGTAGATTGACTAAT
TGTTTAGATTCTTCTGCCTACGACTATGCAGCGATCATAGTCAAGAGGCC
GCCATACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATG
ATTCTGCCATTCAAACCTTTAGGTATGCCACAGAT

>MMTV-32 (292 pb)
TTTTTCCCCAAATTATTAAGAATTACCTCAGCTAATAGATATAGAGAAAA
GAGGATCTACTTTTCATATTTCTGTCTTCTTGTAGATTGACTAATTGT
TTAGATTCTTCTGCCTACGACTATGCAGCGATCATAGTCAAGAGGCCGCC
ATACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGATT
CTGCCATTCAAACCTTTAGGTATGCCACAGATA

>GAPDH001CM
AGAATGGGGAAGAAAAACCAAGGGCCTTCATCGTGATGGTAGTGCAACTG
ACGGCATGATGTTAAAAATCGGTAAAAATGCCACCTCCATATTTTGAG
GAAAAATGAACTGAATGTGTTCGGTGACTTATTTCCATCATCGTCCTTAGGG
AACTTGGGTAGGGGCAGGCGTGTACTGGGACCTAGTCCACCCCTGGCTCT
GCCCTGAACGGCTCAGTGCTTTGGGCAGTTACTCCCCGGCCTCACTTTGC
AGTGTGCTTACCTATGAACAAAATACATACCTCAACGGCACCCCTGAACCC
CACATTTGGGAGGCCCGGTGGTGTATCCACC

>GAPDH002CM
AAAGTTGTTGTGGGGTAACCAGGGTCCTTCCATCGTGCCGAAAGGGGTA
AGGACTATATGTGTTAAAAATCGGTAAAAATGCCACCTCGCATAGTTT
TGAGGAAGATGAACTGAGATGTGTCAGGGTGACTTATTTCCATCATCGTC
CTTAGGGGAACCTGGGTAGGGGCAAGGCGTGTAGCTGGGACCTAGTCCAG
ACCCCTGGCTCTGCCACTGAACGGCTCATTGCTTTGGGCAGTTACTCCCG
GGCCTCACTTTGCACGTGTGCTTACCTATGGAGACAAAAGTACATACCTC
GTAGAGCGCGCACCCCTGTAACCCACCCCTTTGGGAGGCAAGGTGGGTGAT
CACCTG

```

**Figura 25. Secuencias obtenidas a partir de los amplificadores de las muestras de adenocarcinomas 031CM, 032CM para el gen *env* y de las muestras 001CM y 002CM un fragmento del gen GAPDH.** En la figura se pueden observar las secuencias obtenidas a partir de los amplificadores detectados para en el gen *env* del MMTV en las muestras 031CM y 032CM, así como para los amplificadores del gen GAPDH en las muestras 001CM y 002CM

Posteriormente, las secuencias fueron analizadas mediante la herramienta bioinformática BLAST 2.2., se realizó un Blast en el recurso electrónico de la página del NCBI para las secuencias obtenidas de los amplificadores del gen *env* del MMTV y para el amplificador del gen GAPDH. Se obtuvieron las siguientes homologías:

Para el amplificador del gen GAPDH (muestra GAPDH 001CM) se obtuvo un hit a 733 pb del extremo 5' del gen gliceraldehido-3-fosfato deshidrogenasa con una identidad de 97% (312/324) y un valor *e* de  $2 \times 10^{-146}$  (ver figura 26).

```

Features flanking this part of subject sequence:
 743 bp at 5' side: glyceraldehyde-3-phosphate dehydrogenase

Score = 523 bits (283), Expect = 2e-146
Identities = 312/324 (97%), Gaps = 10/324 (3%)
Strand=Plus/Plus

Query  43      AAGG-GGTAAGGACTATAT-GTGTTTTAAAAATCGGTAAAAATGCCACCTCGCATAGTTT  100
      ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct  6588079    AAGGCGGTAAGGACTATATAATGTTTTAAAAATCGGTAAAAATGCCACCTCGCATAGTTT  6588138

Query  101     TGAGGAAGATGAACTGAGATGTGTCAGGGTGACTTATTTCCATCATCGTCCTTAGGGGAA  160
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct  6588139    TGAGGAAGATGAACTGAGATGTGTCAGGGTGACTTATTTCCATCATCGTCCTTAGGGGAA  6588198

Query  161     CTTGGGTAGGGGCAAGGCCTGTAGCTGGGACCTAG-ICCAGACCCCTGGCTCTGCCACTG  219
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct  6588199    CTTGGGTAGGGGCAAGGCCTGTAGCTGGGACCTAGGTCCAGACCCCTGGCTCTGCCACTG  6588258

Query  220     AACGGCTCA-ITGCTTTGGGCAGTTACTCCCGGGCCTCACTTTGCACGTGTGCTTACCTA  278
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct  6588259    AACGGCTCAGTTGCTTTGGGCAGTTACTCCCGGGCCTCACTTTGCACGTGTGCTTACCTA  6588318

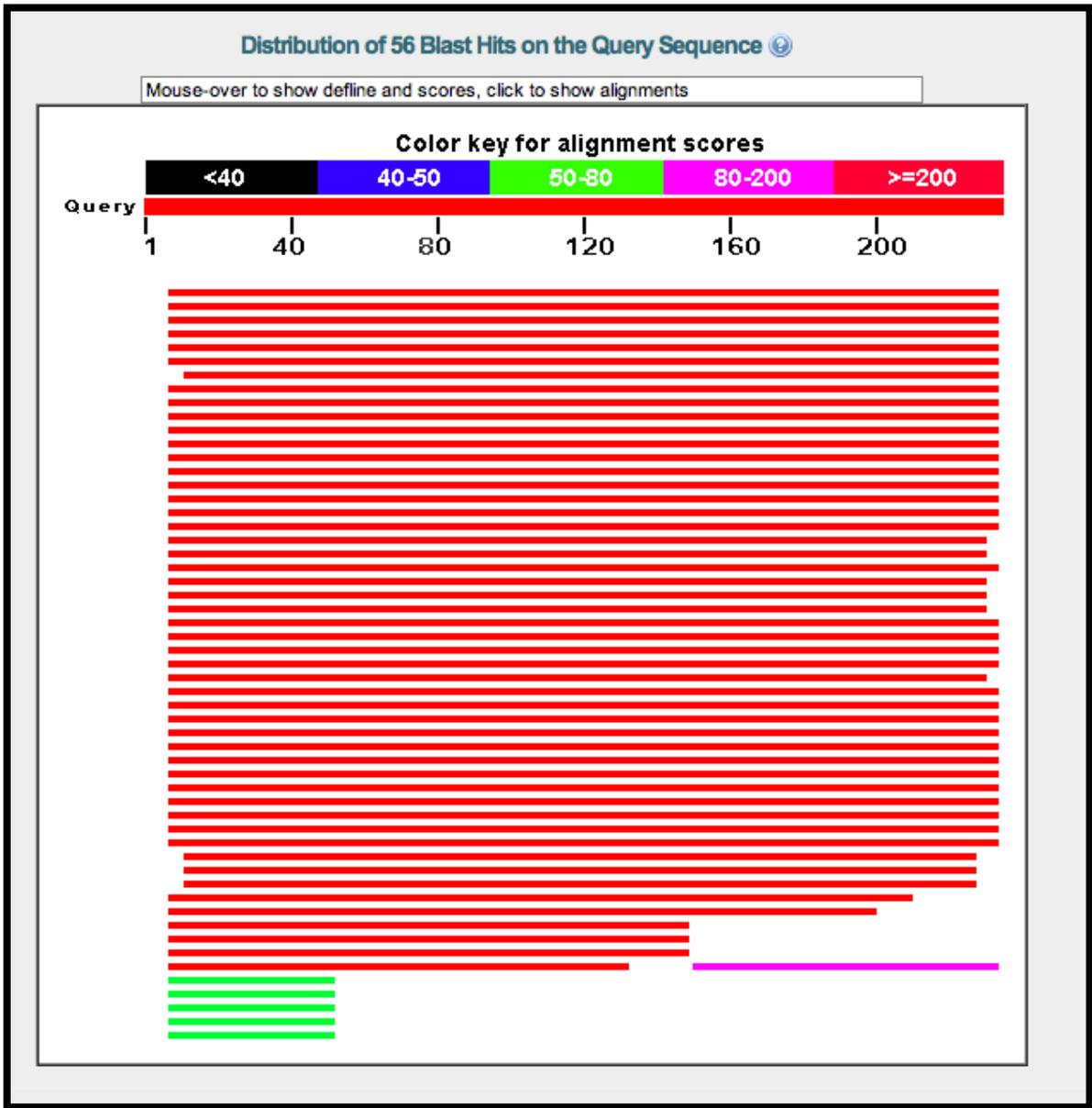
Query  279     -TGGAGACAAAAGTACATACCTCG-TAGAGCGCGCAC-CCTGTAACCCACC-CTTTGGG  334
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct  6588319    GTGGAGACAAAAGTACATACCTCGGTAGAGCGCGCACGCCTGTAACCCACCAGCACTTTGGG  6588378

Query  335     AGGC-AAGGTGGGTG-ATCACCTG  356
      ||||| ||||||| |||||||
Sbjct  6588379    AGGCCAAGGTGGGTGTATCACCTG  6588402

```

**Figura 26. Alineamiento obtenido del amplificado de la muestra 001CM del gen GAPDH.** En la figura se muestra en alineamiento obtenido de la muestra 001CM GAPDH con un hit en el gen de la gliceraldehido 3-fosfato deshidrogenasa de humano.

Para los amplificadores de gen *env* del MMTV (muestra MMTV 001CM y 002CM) se obtuvieron 56 hits en secuencias depositadas de aislados del gen *env* del MMTV con identidad de hasta el 99% de las secuencias y valores de *e* de  $1 \times 10^{-107}$  (ver figuras 27, 28 y 29).



**Figura 27. Alineamiento obtenido de la secuencia amplificada del gen *env* del MMTV de una muestra de adenocarcinoma mamario (Muestra 002CM).** En la figura se muestran los 56 hits obtenidos con scores mayores de 50 -80 puntos, 80-200 puntos y mayores de 200 (50 hits obtenidos).

Las secuencias obtenidas para cada alineamiento en contrado con la secuencia de la muestra 002CM, así como los valores de score y los valores de *e* se muestran en la tabla 3.

**Tabla 3. Alineamientos obtenidos para la búsqueda realizada con la secuencia de la muestra 002CM.**  
 Las secuencias encontradas arrojaron valores de  $e$  de  $1 \times 10^{-108}$  para las secuencias depositadas del gen env del MMTV

Sequences producing significant alignments:	Score (Bits)	E Value
gb AY702717.1  Mouse mammary tumor virus isolate 5 envelope p...	401	1e-108
gb AY152721.1  Mouse mammary tumor virus strain C3H/C envelop...	401	1e-108
gb AF228552.1 AF228552 Exogenous mouse mammary tumor virus C3...	401	1e-108
gb K00556.1 MMTENV mouse mammary tumor virus env mrna	401	1e-108
gb AY702718.1  Mouse mammary tumor virus isolate 6 envelope p...	398	1e-107
gb AY702716.1  Mouse mammary tumor virus isolate 4 envelope p...	398	1e-107
emb X01811.1  Mouse mammary tumor virus env gene for envelope...	398	1e-107
gb AF513921.1  Human betaretrovirus isolate Es177-3 env and s...	401	1e-108
gb AY326252.1  Human betaretrovirus isolate Es238-86 envelope...	392	6e-106
gb AC122322.4  Mus musculus BAC clone RP23-321G20 from chromo...	392	6e-106
gb AC140374.3  Mus musculus BAC clone RP24-76K4 from 6, compl...	392	6e-106
gb M22028.1 MMTENV Endogenous mouse mammary tumor virus (loc...	392	6e-106
gb M11024.1 MUSERMMTR Mouse endogenous mammary tumor virus (M...	387	2e-104
gb AF033807.1 AF033807 Mouse mammary tumor virus complete pro...	383	3e-103
gb M15122.1 MMTPROCG Mouse mammary tumor virus complete provi...	383	3e-103
emb AL683884.26  Mouse DNA sequence from clone RP23-117B19 on...	381	1e-102
gb AY152722.1  Mouse mammary tumor virus strain 40412 envelop...	379	4e-102
gb DQ767968.1  Mouse mammary tumor virus envelope glycoprotei...	378	1e-101
gb BC003207.1  Mus musculus cDNA clone IMAGE:3587523, partial...	378	1e-101
dbj AK085738.1  Mus musculus 10 days lactation, adult female ...	378	1e-101
gb BC018102.1  Mus musculus, clone IMAGE:3495376, mRNA	378	1e-101
gb AF228550.1 AF228550 Endogenous mouse mammary tumor virus M...	378	1e-101
gb AF071010.1 AF071010 Mouse mammary tumor virus putative int...	378	1e-101
gb AF263910.1 AF263910 Mus musculus endogenous mouse mammary ...	372	5e-100
gb AF228551.1 AF228551 Exogenous mouse mammary tumor virus He...	372	5e-100
gb AF346816.1  Human mammary tumor virus envelope protein gen...	370	2e-99
gb AY702714.1  Mouse mammary tumor virus isolate 2 envelope p...	369	7e-99
gb AY702713.1  Mouse mammary tumor virus isolate 1 envelope p...	369	7e-99
gb AF243039.1  Human mammary tumor virus SAG pseudogene, comp...	365	8e-98
gb EF495356.1  Human mammary tumor virus isolate 11 envelope ...	363	3e-97
gb AY702715.1  Mouse mammary tumor virus isolate 3 envelope p...	363	3e-97
gb AC140344.3  Mus musculus BAC clone RP23-237M10 from chromo...	361	1e-96
dbj AK033592.1  Mus musculus adult male cecum cDNA, RIKEN ful...	361	1e-96
gb AF043688.1 AF043688 Mus musculus Mouse mammary tumor virus...	361	1e-96
gb AY659980.1  Mouse mammary tumor virus from Homo sapiens cl...	358	1e-95
gb EF495355.1  Human mammary tumor virus isolate 12 envelope ...	356	4e-95
gb AY659983.1  Mouse mammary tumor virus from Homo sapiens cl...	354	1e-94
gb AY659981.1  Mouse mammary tumor virus from Homo sapiens cl...	354	1e-94
dbj D16249.1 MMTPROVR Mouse mammary tumor virus proviral DNA ...	352	5e-94
gb AY659984.1  Mouse mammary tumor virus from Homo sapiens cl...	349	6e-93
gb AY659982.1  Mouse mammary tumor virus from Homo sapiens cl...	349	6e-93
dbj AK145024.1  Mus musculus mammary gland RCB-0526 Jyg-MC(A)...	347	2e-92
gb AY659985.1  Mouse mammary tumor virus from Homo sapiens cl...	345	8e-92
dbj AK145002.1  Mus musculus mammary gland RCB-0526 Jyg-MC(A)...	343	3e-91
gb AF239172.1 AF239172 Homo sapiens Env-like protein gene, pa...	313	4e-82
gb AF043689.1 AF043689 Mus musculus Mouse mammary tumor virus...	288	2e-74
gb HM636471.1  Mouse mammary tumor virus isolate mmtv101 enve...	224	2e-55
gb GU252129.1  Human mammary tumor virus nonfunctional envelo...	221	2e-54
gb HM636470.1  Mouse mammary tumor virus isolate MMTV14hz non...	215	9e-53
gb DQ925473.1  Human mammary tumor virus envelope protein (en...	206	5e-50
gb DQ910867.1  Human mammary tumor virus env/LTR fragment gen...	159	6e-36
gb AC242833.4  Rattus norvegicus Y Chr BAC RNAEX-235F02 (Ampl...	80.6	5e-12

```
gb|AC242860.2| Rattus norvegicus Y Chr BAC RNAEX-132E23 (Ampl... 80.6 5e-12
gb|AC241951.4| Rattus norvegicus Y Chr BAC RNECO-297N16 (Ampl... 80.6 5e-12
```

```
>gb|K00556.1|MMTENV mouse mammary tumor virus env mrna
Length=3513

Score = 396 bits (214), Expect = 4e-107
Identities = 219/221 (99%), Gaps = 1/221 (0%)
Strand=Plus/Plus

Query 12 ATTATTAAGAATTACCTCAGCTAATAGATATAGAGAAAAGAGGATCTACTTTTCATATTT 71
      |||
Sbjct 1421 ATTATT-AGGATTACCTCAGCTAATAGATATAGAGAAAAGAGGATCTACTTTTCATATTT 1479

Query 72 CCTGTTCTTCTTGTAGATTGACTAATTGTTTAGATTCTTCTGCCTACGACTATGCAGCGA 131
      |||
Sbjct 1480 CCTGTTCTTCTTGTAGATTGACTAATTGTTTAGATTCTTCTGCCTACGACTATGCAGCGA 1539

Query 132 TCATAGTCAAGAGGCCGCCATACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGT 191
      |||
Sbjct 1540 TCATAGTCAAGAGGCCGCCATACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGT 1599

Query 192 TTGATGATTCTGCCATTCAAACCTTTAGGTATGCCACAGAT 232
      |||
Sbjct 1600 TTGATGATTCTGCCATTCAAACCTTTAGGTATGCCACAGAT 1640
```

Figura 28. Alineamiento de la secuencia obtenida del gen env del MMTV de una muestra de adenocarcinoma mamario (muestra 002CM). En la figura se muestra un alineamiento con un hit en el gen env del MMTV en la posición 1421 hasta la posición 1640, el porcentaje de identidad en dichas secuencias fue de 99% (219/221), el valor e obtenido fue de  $4 \times 10^{-107}$ .

```
>gb|EF495356.1| Human mammary tumor virus isolate 11 envelope protein (env) gene,
partial sequence
Length=255

Score = 392 bits (212), Expect = 5e-106
Identities = 221/225 (98%), Gaps = 1/225 (0%)
Strand=Plus/Plus

Query 8 CCAAATTATTAAGAATTACCTCAGCTAATAGATATAGAGAAAAGAGGATCTACTTTTCAT 67
      |||
Sbjct 32 CCATATTATT-AGGATTACCTCAGCTAATAGATATAGAGAAAAGAGGATCTACTTTTCAT 90

Query 68 ATTTCTGTCTTCTTGTAGATTGACTAATTGTTTAGATTCTTCTGCCTACGACTATGCA 127
      |||
Sbjct 91 ATTTCTGTCTTCTTGTAGATTGACTAATTGTTTAGATTCTTCTGCCTACGACTATGCA 150

Query 128 GCGATCATAGTCAAGAGGCCGCCATACGTGCTGCTACCTGTAGATATTGGTGATGAACCA 187
      |||
Sbjct 151 GCGATCATAGTCAAGAGGCCGCCATATGTGCTGCTACCTGTAGATATTGGTGATGAACCA 210

Query 188 TGGTTTGATGATTCTGCCATTCAAACCTTTAGGTATGCCACAGAT 232
      |||
Sbjct 211 TGGTTTGATGATTCTGCCATTCAAACCTTTAGGTATGCCACAGAT 255
```

Figura 29. Alineamiento de la secuencia obtenida del gen env del MMTV de una muestra de adenocarcinoma mamario (muestra 001CM). En la figura se muestra un alineamiento con un hit en la secuencia EF495356.1 que pertenece a un aislado del gen env del Human mammary tumor virus (HMTV) en la posición 32 hasta la posición 255, el porcentaje de identidad en dichas secuencias fue de 98% (221/225), el valor e obtenido fue de  $5 \times 10^{-106}$ .

Posteriormente con la ayuda de la herramienta bioinformática T-coffe (EMBL-EBI) se alinearon las secuencias obtenidas para los amplificadores del gen *env* del MMTV de las muestras de adenocarcinomas (002CM, 004CM, 007CM, 027CM, 031CM y 032CM), el amplificador del control positivo (gen *env* del MMTV C3H) y la secuencia depositada en la base de datos del NCBI para el MMTV (MMTV\_NC001503.1) se obtuvo el siguiente alineamiento:

```

MMTV002CM      -----ATAGATCTA---G-AGAA---AAGAGGATCTA-CTTT
MMTV004CM      -----ATAGATATAA---G-AGAA---AAGAGGATCTA-CTTT
MMTV007CM      -----AAGGCGATC---T-GGCA---GTGAGGACCTGAAATC
MMTV027CM      -----ATATCTACAGGTAGCAGCACGATGGCGGC---CTCT
MMTV031CM      -----ATAGATATAA---G-AGAA---AAGAGGATCTA-CTTT
MMTV032CM      -----ATAGATATA---G-AGAA---AAGAGGATCTA-CTTT
MMTVNC_001503  TTACCTCAGTTAATAGATAT-A---G-AGAA---AAGAGGATCTA-CTTT
PlasmidoMMTV   -----ATAGATATA---G-AGA---AAGAGGATCTA-CTTT
                *                *                *  *

```

```

MMTV002CM      TCA---TATTTCTGTTCTTCTTGTAG-ATTGACTAAT-T--GTTTAG--
MMTV004CM      TCA---TATTTCTGTTCTTCTTGTAG-ATTGACTAAT-T--GTTTAG--
MMTV007CM      GCT---TGACTCGAATTAAATGTGTGGCAGTGAAGGCTTGGGTCTGGTG
MMTV027CM      TGAATGATGATCGCTGCA-TAGTCGTAG-GCAGAAGAAT-C--TAAACA--
MMTV031CM      TCA---TATTTCTGTTCTTCTTGTAG-ATTGACTAAT-T--GTTTAG--
MMTV032CM      TCA---TATTTCTGTTCTTCTTGTAG-ATTGACTAAT-T--GTTTAG--
MMTVNC_001503  TCA---TATTTCTGTTCTTCTTGTAG-ATTGACTAAT-T--GTTTAG--
PlasmidoMMTV   TCA---TATTTCTGTTCTTCTTGTAG-ATTGACTAAT-T--GTTTAG--
                *                ** *                ** *

```

```

MMTV002CM      -ATTCTTCTGCC-TACGACTAT-GCAGCGATCATAGTCAAGAGGCCGCCA
MMTV004CM      -ATTCTTCTGCC-TACGACTAT-GCAGCGATCATAGTCAAGAGGCCGCCA
MMTV007CM      TAATGATCAGCCCTGCTTCCCT---CTCGAT-ATGTCTAAAAAGAT--GA
MMTV027CM      -ATTAGTCAATC-TACAAGAAGAACAGGAAATATG---AAAAGT-----A
MMTV031CM      -ATTCTTCTGCC-TACGACTAT-GCAGCGATCATAGTCAAGAGGCCGCCA
MMTV032CM      -ATTCTTCTGCC-TACGACTAT-GCAGCGATCATAGTCAAGAGGCCGCCA
MMTVNC_001503  -ACTTCTTCTGCC-TACGACTAT-GCAGCGATCATAGTCAAGAGGCCGCCA
PlasmidoMMTV   -ATTCTTCTGCC-TACGACTAT-GCAGCGATCATAGTCAAGAGGCCGCCA
                * * ** * * *                * ** * * *

```

```

MMTV002CM      TACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA-TT
MMTV004CM      TACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA-TT
MMTV007CM      GCCGTCTTGCTCTCCTTCTTAAATTCATACTGCAGAATAATCTGAAAGGTG
MMTV027CM      GATCCTCTTTTCTCTATATCTATTAG--CTGAGGTAA-TCCTAATAA-T-
MMTV031CM      TACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA-TT
MMTV032CM      TACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA-TT
MMTVNC_001503  TATGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA-TT
PlasmidoMMTV   TACGTGCTGCTACCTGTAGATATTGGTGATGAACCATGGTTTGATGA-TT
                * * * * * * * * * *

```

```

MMTV002CM      CTGCCATTCA-ACC-TTTAGGTATGCC-ACAGATAAATTCGAGCCAAGC
MMTV004CM      CTGCCATTCAAAC-TTTAGGTATGCC-AC-----
MMTV007CM      AAAATCTTTATCCC-TTCAGAA-----
MMTV027CM      ATGGCATAAGGGTAAGTAACACAGGCCGATGTATAAAATTCGAGCCAAGC
MMTV031CM      CTGCCATTCAAAC-TTTAGGTATGCC-ACA-----
MMTV032CM      CTGCCATTCAAAC-TTTAGGTATGCC-ACA-----
MMTVNC_001503  CTGCCATTCAAAC-TTTAGGTATGCC-ACAGATTTAATTCGAGCTAAGC
PlasmidoMMTV   CTGCCATTCAAAC-TTTAGGTATGCC-ACAGATAAATTCGAGCCAAGC
                *                * *

```

```

MMTV002CM      GATTC-----AG-----
MMTV004CM      -----
MMTV007CM      -----
MMTV027CM      GATTC-----AGGT-----
MMTV031CM      -----
MMTV032CM      -----
MMTVNC_001503  GATTCGTCGCAGCCATTATCCTGGGCATATCTGCTTTGATTGCTATTATT
PlasmidoMMTV   GATTC-----AGG-----

```

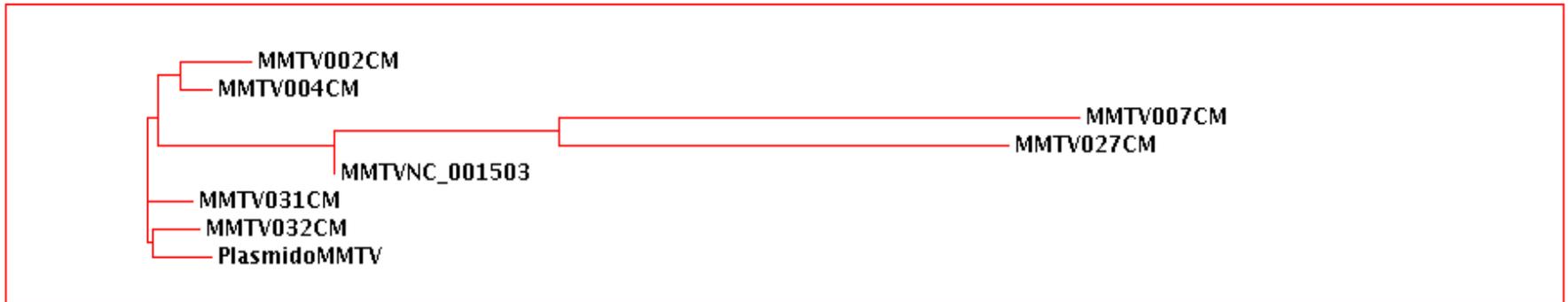
Con las secuencias alineadas se realizó un árbol filogenético con la finalidad de evaluar las distancias entre cada secuencia así como verificar homología entre las secuencias obtenidas y la secuencia depositada en el NCBI para el MMTV. Las distancias obtenidas mediante la opción filogenia con Clustal W2 fueron las siguientes:

```

(
(
(
MMTV002CM:0.02063,
MMTV004CM:0.00937)
:0.00625,
(
(
MMTV007CM:0.15000,
MMTV027CM:0.13000)
:0.06500,
MMTVNC_001503:-0.03500)
:0.05125)
:0.00313,
MMTV031CM:0.01312,
(
MMTV032CM:0.01312,
PlasmidoMMTV:0.01688)
:0.00187);

```

Con dichas distancias obtenidas se construyó un árbol filogenético en cual se puede observar en la figura 30. En dicha figura se observan las secuencias ordenadas en tres principales ramas incluyendo la secuencia depositada en el NCBI para el MMTV



**Figura 30. Árbol filogenético construido con las distancias calculadas a partir del alineamiento realizado mediante T-coffee.** En la figura se puede observar la distribución de las secuencias en base a las distancias calculadas mediante la herramienta bioinformática T-coffee del EMBL-EBI (Modelo de distancias Tamura-Nei, parametros de construcción Neighbor-Joining, 1000 réplicas).

### 7.5 PCR splinkerette para amplificar las secuencias flanqueadas por las LTRs virales

Una vez obtenidas las muestras MMTV positivas y las secuencias verificadas mediante secuenciación, se realizó el procedimiento de PCR splinkerette de acuerdo al método reportado por Theodorou V y col., 2007, se seleccionaron 14 muestras *env* positivas para llevar a cabo el procedimiento. Las muestras seleccionadas fueron las siguientes: 002CM, 003CM, 004CM, 007CM, 010CM, 012CM, 016CM, 017CM, 018CM, 024CM, 026CM, 027CM, 028CM, 031CM.

Se realizó la digestión con *Xho*II para cada una de las muestras seleccionadas y se verificó la reacción separando los productos en un gel de agarosa al 2%. En la figura 31 se puede observar un marcador de peso molecular lambda/hindII (carril M) un control positivo reacción para verificar que la digestión se llevaba a cabo (DNA del fago lambda) (carril +) la muestra 002CM sin digerir y la muestra 002CM digerida (carril 002D)



**Figura 31. Electroforesis en gel de agarosa al 2% para los productos de restricción con la enzima *Xho* II.** En la figura se observa en el carril 1 un marcador de peso Fago lambda, carril número 2 el fago lambda digerido con *Xho* II utilizado como control de la reacción de restricción, carril 3 la muestra de DNA 002 sin digerir, carril 4 muestra 002 digerido con *Xho* II, carril 5 muestra 003 sin digerir y carril 6 muestra 003 digerida con *Xho* II

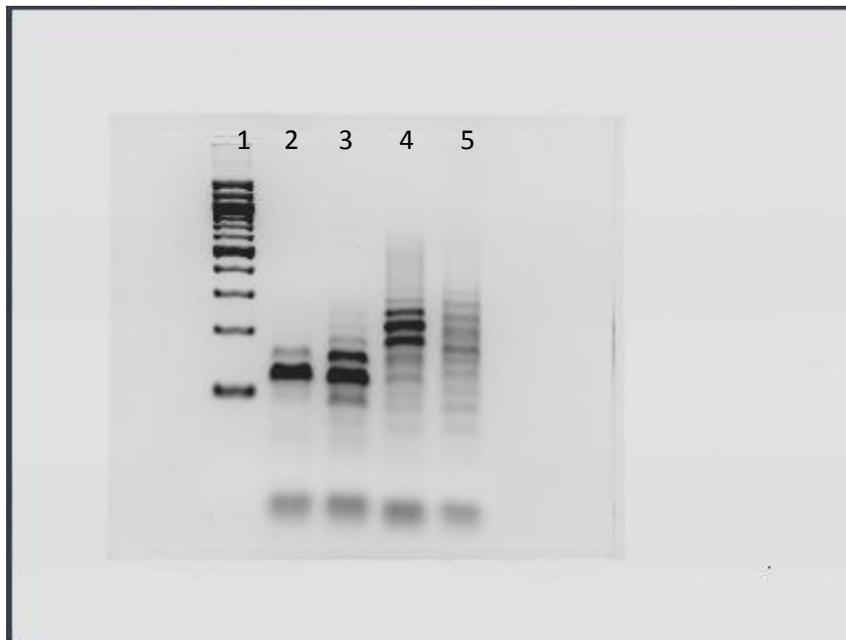
Posteriormente se llevaron a cabo las ligaciones del adaptador a las muestras digeridas con *XhoI* y se llevó a cabo el procedimiento completo de PCR splinkerette. Se amplificaron las muestras con los oligonucleótidos LTR-beg2 (que se alinea a la LTR viral) y 4HMSP1a (que se alinea al enlazador splinkerette) (68 °C de alineación, 35 ciclos). Se reamplificaron los productos de PCR con un cebador LTR y un primer 4HMSP1a (11 ciclos de alineación a 68 °C seguido de 11 ciclos de 67 °C y 13 ciclos en 66 °C). Y se obtuvieron los amplificadores correspondientes (ver figuras 31, 32 y 33).

En la figura 31 se pueden observar los amplificadores obtenidos mediante PCR splinkerette de las muestras MMTV+ 002CM, 003CM, 004CM, 007CM, 010CM, 012CM. Así mismo se pueden observar los diversos tamaños en los amplificadores obtenidos en la muestra 010CM (carril 6) los cuales van desde 80 pb hasta los 350 pb aproximadamente.



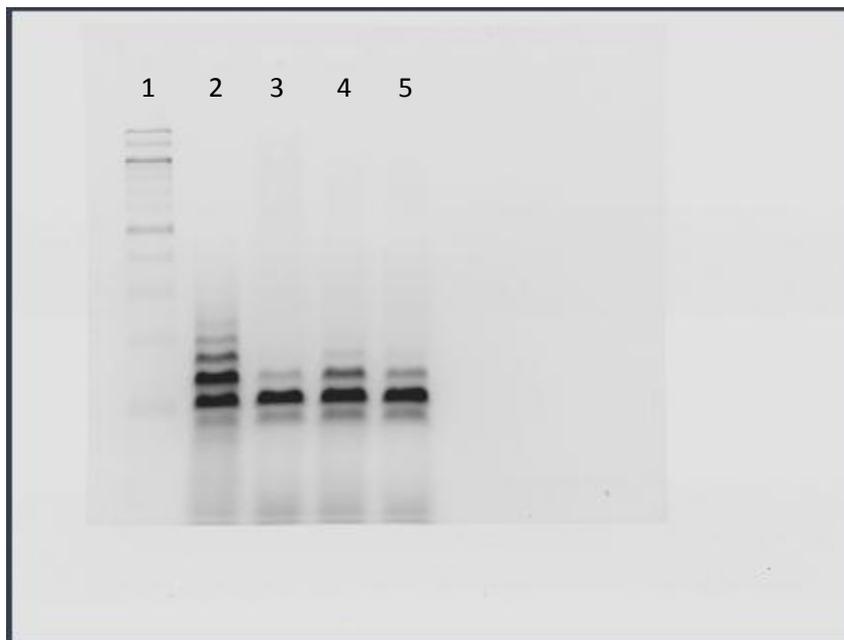
**Figura 31. Electroforesis en gel de agarosa al 4% para los productos de PCR splinkerette.** En la figura se observa en el carril 1 un marcador de peso molecular en escalera de 100pb., carril 2 el producto de PCR splinkerette de la muestra 002 CM (1 amplificado), carril 3 producto de PCR splinkerette de la muestra 003CM (2 amplificadores aprox 100 pb), carril 4 producto de PCR splinkerette de la muestra 004CM (3 amplificadores), carril 5 producto de PCR splinkerette de la muestra 007CM (4 amplificadores), carril 6 producto de PCR splinkerette de la muestra 010CM (6 amplificadores), carril 7 producto de PCR splinkerette de la muestra 012CM.

En la figura 32 es posible observar los amplificadores obtenidos mediante PCR splinkerette de las muestras MMTV+ 016CM, 017CM, 018CM, 024CM. Así mismo se pueden observar los diversos tamaños en los amplificadores obtenidos en las muestra 018CM y 024CM (carril 4 y 5). En el carril uno se observa una escalera de peso molecular de 100 pb de bases, en el carril dos se observan dos amplificadores productos de la muestra 0016CM (aprox. 120 y 160pb), en el carril 3 los productos de la PCR splinkerette de la muestra 017CM (4 amplificadores aprox 100,110, 130, 170 pb), en el carril 4 los productos de la PCR splinkerette de la muestra 0018CM (8 amplificadores) y en el carril 5 los productos de la PCR splinkerette de la muestra 024CM (8 amplificadores).



**Figura 32. Electroforesis en gel de agarosa al 4% para los productos de PCR splinkerette.** En la figura se observa en el carril 1 un marcador de peso molecular en escalera de 100pb., carril 2 el producto de PCR splinkerette de la muestra 0016 CM (2 amplificado aprox. 120 y 160pb), carril 3 producto de PCR splinkerette de la muestra 0017CM (4 amplificadores aprox 100,110, 130, 170 pb), carril 4 producto de PCR splinkerette de la muestra 0018CM (8 amplificadores), carril 5 producto de PCR splinkerette de la muestra 024CM (8 amplificadores).

En la figura 33 es posible observar los amplificados obtenidos mediante PCR splinkerette de las muestras MMTV+ 026CM, 027CM, 028CM, 031CM. Así mismo es posible observar los diversos tamaños en los amplificados obtenidos en la muestra 026CM (carril 2). En el carril uno se observa una escalera de peso molecular de 100 pb de bases, en el carril dos se observan seis amplificados productos de la muestra 0026CM (aprox. 120 y 160pb carril 3 producto de PCR splinkerette de la muestra 0027CM (3 amplificados aprox 100,110, 130 pb), carril 4 producto de PCR splinkerette de la muestra 028CM (4 amplificados), carril 5 producto de PCR splinkerette de la muestra 031CM (3 amplificados). Como se observa en los carriles 3 y 4 (muestras 027CM y 031CM) se observan tres amplificados en ambos carriles de aproximadamente 90, 110 y 130 pb en ambos carriles.



**Figura 33. Electroforesis en gel de agarosa al 4% para los productos de PCR splinkerette.** En la figura se observa en el carril 1 un marcador de peso molecular en escalera de 100pb., carril 2 el producto de PCR splinkerette de la muestra 026 CM (6 amplificado aprox. 90 y 220pb), carril 3 producto de PCR splinkerette de la muestra 0027CM (3 amplificados aprox 90,110, 130 pb), carril 4 producto de PCR splinkerette de la muestra 028CM (4 amplificados), carril 5 producto de PCR splinkerette de la muestra 031CM (3 amplificados).

Los amplificadores obtenidos mediante la PCR splinkerette fueron seleccionados y reamplificados, posteriormente fueron purificados y enviados al servicio de secuenciación de la FES-UNAM Iztacala a cargo del M en C Alejandro Monzalvo con la finalidad de conocer cuales son los genes asociados a LTRs virales del MMTV.

En la tabla 4 se pueden observar los amplificadores que fueron seleccionados y reamplificados para ser enviados al servicio de secuenciación. Los amplificadores fueron seleccionados de acuerdo la disposición del volumen de cada producto así como de su concentración en los geles de agarosa. En la tabla se muestra el tamaño de los amplificadores obtenidos así como la muestra de la cual fueron amplificadores. Adicionalmente fueron asignadas claves de localización a cada amplificador con la finalidad de ubicarlos con mayor facilidad.

**Tabla 4. Relación de las muestras de PCR splinkerette que fueron seleccionadas para ser enviadas al servicio de secuenciación.** La tabla muestra el tamaño de los amplificadores obtenidos mediante PCR Splinkerette así como las claves asignadas a cada amplificador

<b>Muestra MMTV +</b>	<b>Tamaño aprox.</b>	<b>Clave</b>	<b>Figura</b>
002CM	100 pb	02a1	30
003CM	100 pb	03a1	30
007CM	150 pb	07a1	30
012CM	120 pb	12a1	30
017CM	130 pb	17a1	31
018CM	200 pb	18a1	31
026CM	100 pb	26a1	32
026CM	150 pb	26a2	32
027CM	100 pb	27a1	32
028CM	100 pb	28a1	32

Los amplificadores fueron enviados al servicio de secuenciación de la FES-UNAM Iztacala a cargo del M en C Alejandro Monzalvo. Las secuencias obtenidas se pueden observar en las figuras 34 y 35.

En la figura 34 se pueden observar las secuencias obtenidas a los amplificadores 02a1, 03a1, 07a1 y 12a1 así como sus respectivos tamaños.

```
>02a1
GATGATGGCCGAATCATCAGCCATGGTTCATCACCAATATCTAC
AGGTAGCAGCACGTATGGCGGCTCTTACTATGATCGTGCAT
AGGTCTGGTAGATCAGAAATTCCTGGGCAGCCCACCTCACACC
CTGCAATTCGTTCCACATGCTGTAATTATAAAAAGCCACTTA
ATAGACTCTCGCTCTCCGTGTGAAAAGCACAGGCTCTTCTATAC
ACAGAGAGATCTAAAATTCGAGCCAAGCGATTCAGGTCTCCCT
GCCAGATCCACCTAGCCAGGGCGTCACCGCAAACCAGAGGAGT
GGAGGCGGGG

>03a1
GATCCATGTGACGACAGTGGAGGTATTTAGGGACAAATGACAT
TTAGGGGACCACTCTGTCTTGAGTCCAGGTCCAGACATTTCC
CTGAGGTCTTCCATAATATGTGCCTTCGGGCAGAGGGACAGAT
GTAATTTGATGGTCACTGACATGTTCTCCACACAGTTGTCTTG
GGTCTCAGGACCAAGGTGTCCAGGCTGCTGGCAGATGTCATGG
GTCTCCAGGATGTAA

>07a1
GGATGATGTGCGTGCTCAGCAGCCTCTCCGTTCAAAGGGCAG
AGTCCAGACTCTGACATCGTGGCTGGTCTCTCTTGGGGGATT
CAGGGCTTCATGTTTTCAGTAGGTCTGCTGCTTTGGGGCGGTGG
TTGGGGTTCTCTCCAGGGCAGCTTCTATCAGCTCCCTCATGC
CTGGACTGCAGTCACCAGCGATGTCTTCCAGGGGAGGTGCCTG
CTTGTGCAGATGTATAAATTCGAGCCAAGCGATTCAGGTCTC
ACTGCCAAATCCAGCCAAGGAGCCCACTCCCGCGTTTCAGAGG
GGGCGGTCAACGG

>12a1
GATGATGTGAAGTTAGGATCTGGGGTACTATTGAGGCTTCTCT
TGATGAGGGTGCTGGAGCCCTAAGCTAGACCACAGCGGGAGGG
CTCTTTCTGAGTCCCCACCTCTGGAGTAGGGAAAAGGCAGCC
CAAGAGCTCTTCCATTTAGGGGAAGAAGGCCAAACCATGTA
TAACTTCTGAGATGGGCCAGCACTGCTCAGTAACTTCTACCCT
ACTCCAGAACCAGGATCCGAGCCAAGCGATTCAGGCCTCCCT
GCCAGATCGGGAAGGGAGGTTCGGCTTACGGCGATTTGCGAGG
GGCGA
```

**Figura 34. Secuencias obtenidas de los amplificadores de PCR splinkerette.** En la figura se pueden observar las secuencias obtenidas a partir de adenocarcinomas mamarios MMTV + 02CM, 03CM, 07CM y 012CM. Las secuencias obtenidas corresponden a los amplificadores 02a1, 03a1, 07a1 y 12a1.

```
>18a1
GATGATGGCAGATCATCAACCATGGTTCATCACCAATATCTACA
GGTAGCAGCACGTATGGCGGCCTCTTGACTATGATCGCTGCATA
GTCGTAGGCAGAAGAATCTAAACAATTAGTCAATCTACAAGAAG
AACAGGAAATATGAAAAGTAGATCCTCTTTTCTCTATATCTATT
AGCTGAGGTAATCCTAATAATATGGCATAAGGGTAAGTAACACA
GGCAGATGTAAAAATCATGTCATGTTCTTGAAATCCTGGCCTTT
TGAGAATAAGATATCTTGAGGCTGCAACTAAGCGAAACAATTCT
GTATGGGGAAGTACAGCAGTAAAATCTCTATCATTGGGATCCTT
AGGAGAATTTTCCCAGAACCATGTAGGTCTACCCATCTGCTT
CATACCATCGATGAACCTCTTTATTGACAGGGGAGTATAATTT
CCAAATAGATCCTTTTTGCTTTTAATCTGATCTGACTGATCTAC
ACTAGGCGAGGGAAGGGAGAAATCCCAAAGTAACCCAAGGCC
CTTTTGGAGAAAAACTCACCCCTGGTCAGGGAAGGCGCAAGGC
AACCACCGTGGAGGAGCAGACTCGTCTCCCTCCCAGAAGGCGTC
CTTCTTAAAGGCGATCTGGCAGTGAAGACCTGAAATCGCTTGAC
TCGAATTAATGTGTGGCAGTAAAAGCTTGGGTCTGGTGTAAAT
GATCAGCCCTGCTTCCCTCTCGATATGTCTAAAAAGATGAGCCG
TCTTGCTCTCCTTCTTAATTCATACTGCAGAATAATCTGAAAGG
TGAAAATCTTTATCCCTTCAGAAAGGCACAAAAA
```

```
>26a2
GATCGATGGCAGAATCATCAACCATGGTTCATCACCAATATCT
ACAGGTAGCAGCACGTATGGCGGCCTCTTGACTATGATCGCTG
CATAGTCGTAGGCAGAAGAATCTAAACAATTAGTCAATCTACA
AGAAGAACAGGAAATATGAAAAGTAGATCCTCTTTTCTCTATA
TCTATTAGCTGAGGTAATCCTAATAATATGGCATAAGGGTAAG
TAACACAGGCAGATGTAAAAATCATGTCATGTTCTTGAAATCC
TGGCCTTTTGAGAATAAGATATCTTGAGGCTGCAACTAAGCGA
AACAAATCTGTATGGGGAAGTACAGCAGTAAAATCTCTATCAT
TGGGATCCTTAGGAGAATTTCCCAGAACCATGTAGGTCTAC
CCATCTGCTTCATACCATCGATGAACCTCTTTATTGACAGGG
GGAGTATAAATTTCAAATAGATCCTTTTTGCTTTTAATCTGAT
CTGACTGATCTACACTAGGCGAGGGAAGGGAGAAATCCCAAAG
TAACCCAAGGCCCTTTTGGAGAAAAACTCACCCCTGGTCA
GGGAAGGCGCAAGGCAACCACCGTGGAGGAGCAGACTCGTCTC
CCTCCCAGAAGGCGTCTTCTTAAAGGCGATCTGGGCAGTGAG
GACCTGAATCGCTTGGGCTCGATTTGAATGTGTGGCAGTGAAG
AACTTGTCTTTGGTAGAATTATCAGGACTACGTTTCATTTATA
AAAATATAAAAAAGAAAAGGACTACCTTGCATATTTCTTGTT
AAGAGATGATGATTAGCT
```

```
>28a1
GAAGATGTCCGAATCTTCAACATGGTTCATCACCTATATCTAC
AGGTAGCAGCACGTATGGCGGCCTCTTGAGTATGATCGCTGCA
TAGTCGTAGGCTGAAGAATCTAAACAATTAGTCAATCTACAAG
AAGAACAGGAAATATGAAAAGTAGATCCTCTTTTCTCTATATC
TATTAGCTGAGGTAATCCTAATAATATGGCATAAGGGTAAGTA
ACACAGGCCGATGTATAAAATTCGAGCCAAGCGATTACAGGTCC
TCACTGCCAGATCCACCCCGCAGTTCAGATTACGGCCGGAG
```

**Figura 35. Secuencias obtenidas de los amplificadores de PCR splinkerette.** En la figura se pueden observar las secuencias obtenidas a partir de adenocarcinomas mamarios MMTV + 02CM, 03CM, 07CM y 012CM. Las secuencias obtenidas corresponden a los amplificadores 02a1, 03a1, 07a1 y 12a1.

Las secuencias obtenidas fueron analizadas y posteriormente se realizó una búsqueda de secuencias mediante la herramienta en línea Blastn del NCBI con los parámetros descritos anteriormente, las regiones encontradas y los genes asociados a dichas secuencias amplificadas se muestran en la siguiente tabla 5.

**Tabla 5. Genes encontrados a partir de las secuencias obtenidas de los amplificadores de PCR splinkerette.** La tabla muestra los fragmentos amplificados, el largo de las secuencias obtenidas y las regiones encontradas mediante Blast

Secuencia	Tamaño (pb)	Región encontrada	Num. Registro
02a1	318	Stromal interaction molecule 2, transcript variant 1	NT_006316.16
03a1	232	Fibroblast growth factor 8 isoform E precursor	NT_030059.13
07a1	314	MAP3K8 mitogen-activated protein kinase	NM_005204.2
12a1	306	Proto-oncogen WNT-1	NT_029419.12
18a1	826	Ubiquitin-like modifier-activating enzyme 6	NT_022778.16
26a2	792	Gonadotropin-releasing hormone receptor	NM_001012763.1
26a1	309	Peroxisomal NADH pyrophosphatase NUDT12	NT_034772.6

### 7.6 Búsqueda Bioinformática

Para identificar posibles sitios de inserción del retrovirus MMTV/HMTV en el genoma humano se realizó una búsqueda bioinformática en la base de datos del NCBI y del EMBL-EBI de acuerdo en la siguiente estrategia:

1. Se descargó la secuencia completa del gen *env* del MMTV y las secuencias LTR 5' y 3' del MMTV de las páginas del NCBI y del EMBL-EBI.

2. El número de acceso para la secuencia completa del MMTV fue NC\_001503 (NCBI)
3. Las secuencias descargadas para las regiones LTRs fueron las siguientes:

**5´ LTR (176 pb) Número de acceso M10761.1 (EMBL-EBI)**

AGAGACGCTCAACCTCAATTGAAGAGCAGGTGCAAGGAAGTTAAGTTTGTGGTTACAAACTGTTCTTAAA  
 ACAAGGATGTGAGACTAGAGAGACGCTCAACCTCAATTGAAGAGCAGGTGCAAGGAAGTTAAGTTTATGG  
 TTACAAACTGTTCTTAAAACGAGGATGTGAGACAAG

**3´ LTR (245 pb) Número de acceso S75270.1 (NCBI)**

TAACCTTTATGAGCCCAACCTTGCGGTTCCCAGGGTTCAAATAAGTTCTTGGTCACAAACTGTTCTTAAA  
 ACAAAGATGTGAGACAAGTGGTTTTCTGACTTGGTTTTGGTATCAAATGTTTTGATCTAAGCTCTGAGTGT  
 TCTATTCTCCTATGTTCTTTTAGAACTTATCCAAGTCTTATGTAAATGCTTATGTAAACCATGATATAAAA  
 AGAGTGCTGATTTTTTGGAGTAAACTTGCAACAGTC

4. La secuencia descargada para el gen env fue la siguiente:

**Gen env del MMTV (2066 pb) NC\_001503 (NCBI)**

ATGCCGAAACACCAATCTGGGTCCCCGACCGATTTCATCCGACCTTTTACTGAGCGGAAAGAAGCAACGCC  
 CACACCTGGCACTGCGGAGAAAACGCCGCCGAGATGAGAAAGATCAACAGGAAAGTCCCAAGAATGAA  
 TCTAGTCCCCATCAAAGAGAAGACGGCTTGGAACATCTGCAGGCGTTGATCTCCGAAGCGGAGGAGGTC  
 CTTAAAACCTCACAAACTCCCCAAACCTCTTTGACCTTATTTCTTGCTTTGTTGTCTGTCTCGGCCCCC  
 CGCCTGTGACAGGGGAGAGTTATTGGGCCTACCTACCTAAACCACCTATTCTCCATCCCGTGGGATGGGG  
 AAGTACAGACCCCATTAGAGTTCTGACAAATCAAACCATGTATTTGGGTGGTTTCGCCTGACTTTTCATGGG  
 TTCAGAAATATGTCTGGTAATGTACATTTTGAGGGGAAGTCTGATACGCTCCCCATTTGCCTTTCTTCT  
 CCTTTTCTACCCCCACGGGCTGCTTTCAAGTAGACAAGCAAGTATTTCTTTCTGATACACCCACGGTTGA  
 TAATAATAAACCTGGGGGAAAGGGTGATAAAAGGCGTATGTGGGAACTTTGGTTGACTACCTTGGGGAAC  
 TCAGGGGCCAATACAAAACCTGGTCCCTATAAAAAAGAAGTTGCCCCCAAATATCCTCACTGCCAGATCG  
 CCTTTAAGAAGGACGCCTTCTGGGAGGGGAGACGAGTCTGCTCCTCCACGGTGGTTGCCTTGCCTTCCC  
 TGACCAGGGGGTGAAGTTTTCTCCAAAAGGGGCCCTTGGGTTACTTTGGGATTTCTCCCTTCCCTCGCCT  
 AGTGTAGATCAGTCAGATCAGATTTAAAAGCAAAAAGAATCTATTTGGAAATTATACTCCCCCTGTCAATA  
 AAGAGGTTTCATCGATGGTATGAAGCAGGATGGGTAGAACCTACTTGGTTCTGGGAAAATTCTCCTAAGGA  
 TCCCAATGATAGAGATTTCACTGCACTAGTCCCCATACAGAATTGTTTCGCTTAGTTCGCAGCCTCAAGA  
 CATCTTATTCTCAAAAAGGCCAGGATTTCAAGAACATGAGATGATTCCCTACATCTGCCTGTGTTACTTACC  
 CTTATGCCATATTATTAGGATTACCTCAGTTAATAGATATAGAGAAAAGAGGATCTACTTTTCATATTTCT  
 CTGTTCTTCTTGTAGATTGACTAATTGTTTAGACTCTTCTGCCTACGACTATGCAGCGATCATAGTCAAG  
 AGGCCGCCATATGTGCTGCTACCTGTAGATATTGGTGTGTAACCATGGTTTGTGATGATTCTGCCATTCAA  
 CCTTTAGGTATGCCACAGATTTAATTCGAGCTAAGCGATTTCGTTCGCAGCCATTATCCTGGGCATATCTGC  
 TTTGATTGCTATTACTTCTTTGCTGTAGCTACTACTGCCTTAGTTAAGGAGATGCAAACCTGCTACG  
 TTTGTTAATAATCTTCATAGGAATGTTACATTAGCCTTATCTGAACAACGGATAATAGATTTAAAATTAG  
 AAGCTAGACTTAATGCTTTAGAAGAAGTAGTTTTAGAGTTGGGACAAGATGTGGCCAATTTAAAGACCAG  
 AATGTCCACTAGGTGTGCAAAATTATGACTTTATCTGCCTTACACCTTTACCCTATAATGCTACTGAG  
 AACTGGGAAAGAACCAGGGCTCATTTATTGGGCATTTGGAATGATAATGAGATTTTCATATAACATACAGG  
 AGTTAACCAACCTGATTAGTGATATGAGCAACAACATATTGATGCAGTGGACCTTAGTGGCTTGGCTCA

```

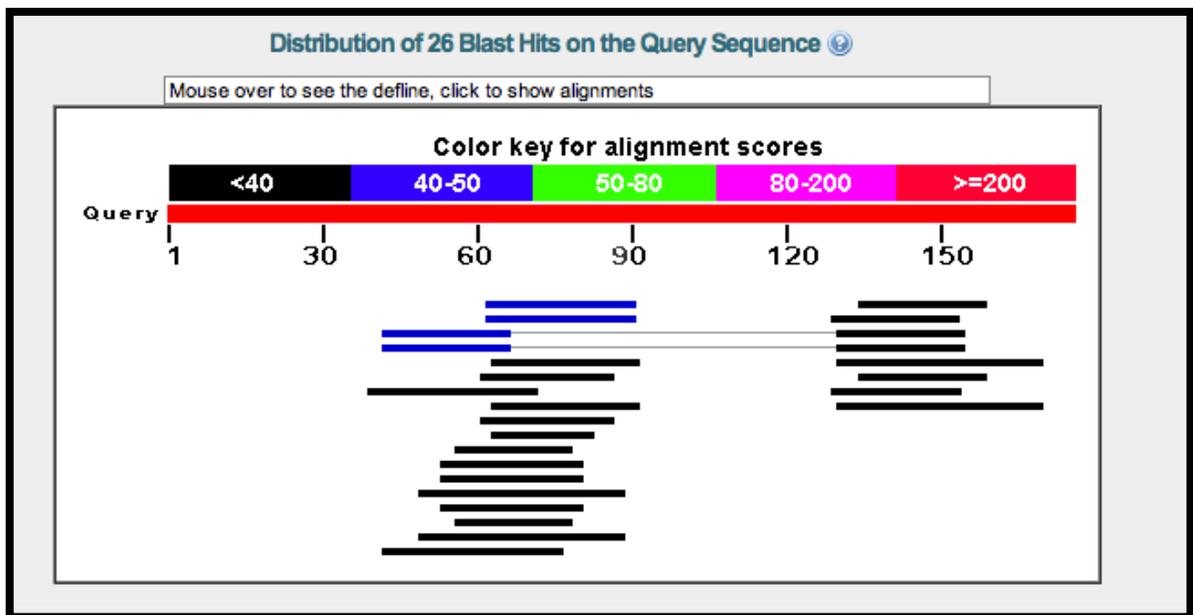
GTCTTTTGCCAATGGAGTGAAGGCTTTAAATCCATTAGATTGGACACAATATTTTCATTTTTATAGGTGTT
GGAGCCCTGCTTTTAGTCATTGTACTTATGATTTTCCCCATTGTTTTCCAGTGCCTTGCGAAGAGCCTTG
ACCAAGTGCAGTCAGATCTTAACGTGCTTCTTTTAAAAAAGAAAAAAGGGGAAATGCCGCGCCTGCAGC
AGAAATGGTTGAACTCCCAGAGTGTCTTACACTTAG

```

- Posteriormente, cuando las secuencias fueron descargadas se procedió a realizar una búsqueda de similitud de secuencias mediante la herramienta Blast (programa nucleotide Blast, algoritmo Blastn, base de datos Human genomic plus transcript), para cada una de las secuencias descargadas. Los parámetros de búsqueda fueron los siguientes:

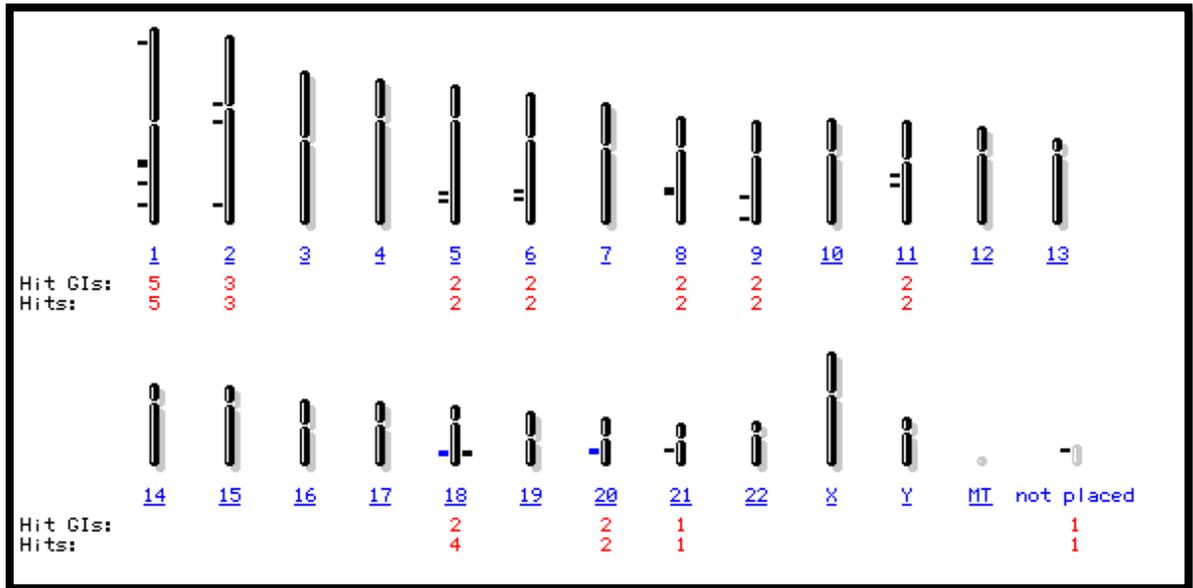
Número máximo de secuencias: 10 000  
 Valor de corte (Expect threshold): 10  
 Tamaño de palabra (Word size): 11  
 Puntaje Match/Mismatch: 2,-3  
 Puntaje gaps (gaps costs): inicio 5, extensión 2

- Para la región 5' LTR del MMTV se encontraron 26 hits como se puede observar en la siguiente figura:



**Figura 36. Distribución de los 26 hits encontrados mediante Blastn para la secuencia 5' LTR del MMTV.** En la figura se observan 26 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 42.8 y 37.4.

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con la región 5´LTR se muestra en la siguiente figura:



**Figura 37. Distribución genómica de los 26 hits encontrados mediante Blastn para la secuencia 5´ LTR del MMTV.** En la figura se observan 25 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas y un alineamiento no encontrado (NW\_001842097.1). Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados se muestran en la siguiente figura donde es posible observar los valores del score asignado a cada alineamiento, porcentaje de similitud en la búsqueda, valor e y porcentaje de identidad entre las secuencias.

Sequences producing significant alignments:							
Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
<b>Genomic sequences</b>							
<a href="#">NT_011362.10</a>	Homo sapiens chromosome 20 genomic contig, GRCh37.p2 r	<a href="#">42.8</a>	42.8	16%	0.11	93%	
<a href="#">NW_001838666.1</a>	Homo sapiens chromosome 20 genomic contig, alternate as	<a href="#">42.8</a>	42.8	16%	0.11	93%	
<a href="#">NT_025028.14</a>	Homo sapiens chromosome 18 genomic contig, GRCh37.p2 r	<a href="#">41.0</a>	78.3	28%	0.40	96%	
<a href="#">NW_001838469.1</a>	Homo sapiens chromosome 18 genomic contig, alternate as	<a href="#">41.0</a>	78.3	28%	0.40	96%	
<a href="#">NT_008046.16</a>	Homo sapiens chromosome 8 genomic contig, GRCh37.p2 re	<a href="#">39.2</a>	39.2	16%	1.4	89%	
<a href="#">NT_029289.11</a>	Homo sapiens chromosome 5 genomic contig, GRCh37.p2 re	<a href="#">39.2</a>	39.2	14%	1.4	92%	
<a href="#">NT_005403.17</a>	Homo sapiens chromosome 2 genomic contig, GRCh37.p2 re	<a href="#">39.2</a>	39.2	18%	1.4	87%	
<a href="#">NW_001839136.1</a>	Homo sapiens chromosome 8 genomic contig, alternate asse	<a href="#">39.2</a>	39.2	16%	1.4	89%	
<a href="#">NW_001838953.2</a>	Homo sapiens chromosome 5 genomic contig, alternate asse	<a href="#">39.2</a>	39.2	14%	1.4	92%	
<a href="#">NT_011512.11</a>	Homo sapiens chromosome 21 genomic contig, GRCh37.p2 r	<a href="#">37.4</a>	37.4	11%	4.8	100%	
<a href="#">NT_167190.1</a>	Homo sapiens chromosome 11 genomic contig, GRCh37.p2 r	<a href="#">37.4</a>	37.4	13%	4.8	95%	
<a href="#">NT_035014.4</a>	Homo sapiens chromosome 9 genomic contig, GRCh37.p2 re	<a href="#">37.4</a>	37.4	14%	4.8	92%	
<a href="#">NT_025741.15</a>	Homo sapiens chromosome 6 genomic contig, GRCh37.p2 re	<a href="#">37.4</a>	37.4	14%	4.8	92%	
<a href="#">NT_022135.16</a>	Homo sapiens chromosome 2 genomic contig, GRCh37.p2 re	<a href="#">37.4</a>	37.4	15%	4.8	89%	
<a href="#">NT_022184.15</a>	Homo sapiens chromosome 2 genomic contig, GRCh37.p2 re	<a href="#">37.4</a>	37.4	15%	4.8	89%	
<a href="#">NT_167186.1</a>	Homo sapiens chromosome 1 genomic contig, GRCh37.p2 re	<a href="#">37.4</a>	37.4	22%	4.8	82%	
<a href="#">NT_004487.19</a>	Homo sapiens chromosome 1 genomic contig, GRCh37.p2 re	<a href="#">37.4</a>	37.4	22%	4.8	80%	
<a href="#">NW_001842097.1</a>	Homo sapiens unplaced genomic contig, alternate assembly	<a href="#">37.4</a>	37.4	15%	4.8	89%	
<a href="#">NW_001838028.2</a>	Homo sapiens chromosome 11 genomic contig, alternate as	<a href="#">37.4</a>	37.4	13%	4.8	95%	
<a href="#">NW_001839240.1</a>	Homo sapiens chromosome 9 genomic contig, alternate asse	<a href="#">37.4</a>	37.4	14%	4.8	92%	
<a href="#">NW_001838990.2</a>	Homo sapiens chromosome 6 genomic contig, alternate asse	<a href="#">37.4</a>	37.4	14%	4.8	92%	
<a href="#">NW_001838543.2</a>	Homo sapiens chromosome 1 genomic contig, alternate asse	<a href="#">37.4</a>	37.4	22%	4.8	82%	
<a href="#">NW_001838533.2</a>	Homo sapiens chromosome 1 genomic contig, alternate asse	<a href="#">37.4</a>	37.4	22%	4.8	80%	
<a href="#">NW_001838573.1</a>	Homo sapiens chromosome 1 genomic contig, alternate asse	<a href="#">37.4</a>	37.4	19%	4.8	82%	

Figura 38. En la figura se muestran los alineamientos encontrados mediante Blastn para la secuencia 5' LTR del MMTV. En la figura se observan los valores del score asignados a cada alineamiento, porcentaje de similitud en la búsqueda, valor e, porcentaje de identidad entre las secuencias el numero de acceso para cada secuencia. Asimismo se resaltó el alineamiento NW\_001842097.1 el cual se encontró en una región cromosómica sin datos disponibles.

Las secuencias encontradas en cada alinamiento fueron las siguientes:

**NT\_011362.10**                    Score = 42.8 bits (46), Expect = 0.11  
                                 Identities = 28/30 (93%), Gaps = 1/30 (3%)  
                                 Strand=Plus/Plus

Regiones encontradas:    30417 bp at 5' side: zinc finger protein 334  
                                 isoform b  
                                 6181 bp at 3' side: transmembrane protein  
                                 C20orf123

**NW\_001838666.1**                Score = 42.8 bits (46), Expect = 0.11  
                                 Identities = 28/30 (93%), Gaps = 1/30 (3%)  
                                 Strand=Plus/Plus

Regiones encontradas:    30385 bp at 5' side: zinc finger protein 334  
                                 isoform b  
                                 6181 bp at 3' side: transmembrane protein  
                                 C20orf123

**NT\_025028.14**                 Score = 41.0 bits (44), Expect = 0.40  
                                 Identities = 24/25 (96%), Gaps = 0/25 (0%)  
                                 Strand=Plus/Plus

Regiones encontradas:    668426 bp at 5' side: serpin B8 isoform a  
                                 1107117 bp at 3' side: cadherin-7 preproprotein  
                                 preproprotein

**NW\_001838469.1**                Score = 41.0 bits (44), Expect = 0.40  
                                 Identities = 24/25 (96%), Gaps = 0/25 (0%)  
                                 Strand=Plus/Plus

Regiones encontradas:    670957 bp at 5' side: serpin B8 isoform b  
                                 1107344 bp at 3' side: cadherin-7 preproprotein  
                                 preproprotein

7. Para la región 3' LTR se encontraron 55 hits como se puede observar en la siguiente figura:

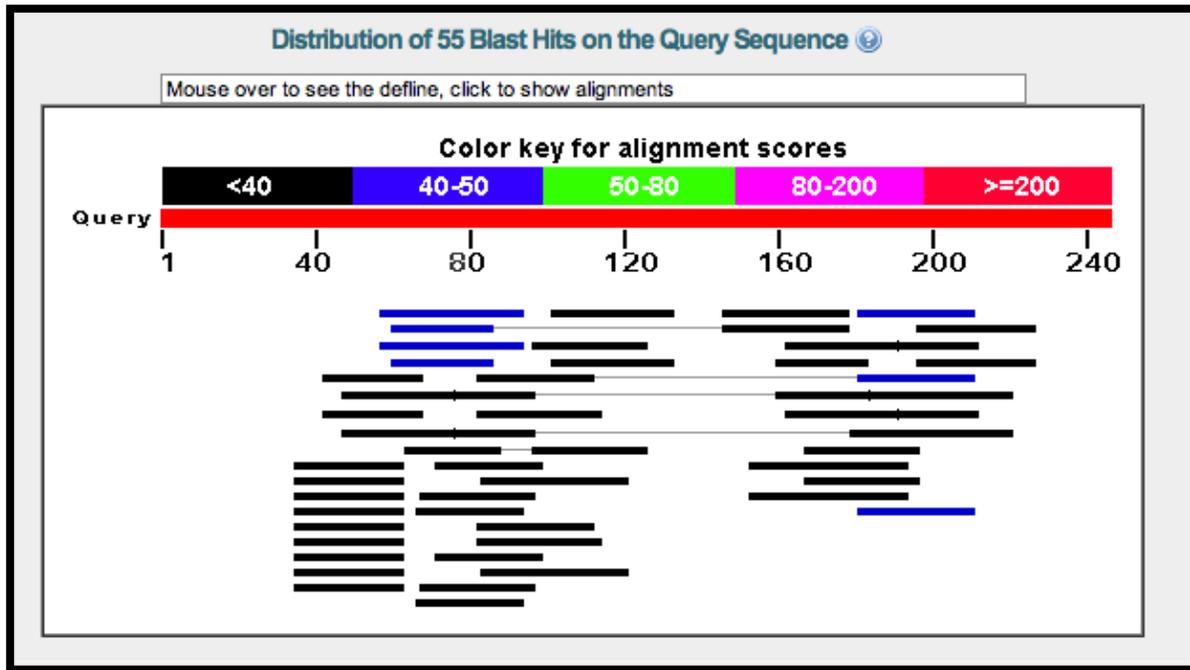
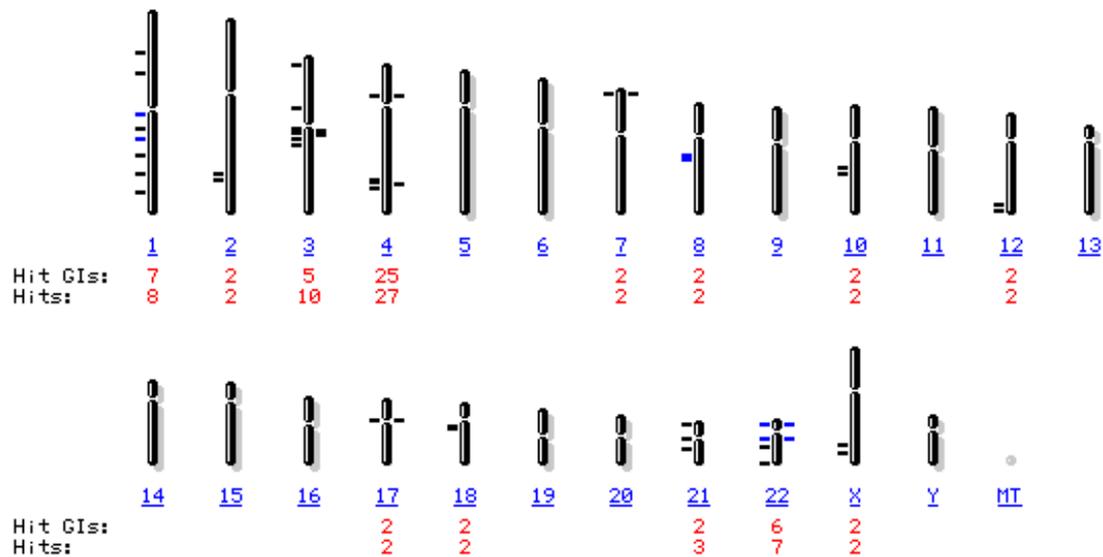


Figura 39. Distribución de los 55 hits encontrados mediante Blastn para la secuencia 3' LTR del MMTV. En la figura se observan 26 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 113.0

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con la región 3' LTR se muestra en la siguiente figura:



**Figura 40. Distribución genómica de los 26 hits encontrados mediante Blastn para la secuencia 3' LTR del MMTV.** En la figura se observan 55 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados fueron los siguientes:

	Score (Bits)	E Value
ref NT_008183.19  Homo sapiens chromosome 8 genomic contig, G...	42.8	0.17
ref NT_004487.19  Homo sapiens chromosome 1 genomic contig, G...	42.8	0.17
ref NW_001839132.1  Homo sapiens chromosome 8 genomic contig,...	42.8	0.17
ref NW_001838531.2  Homo sapiens chromosome 1 genomic contig,...	42.8	0.17
ref NM_015094.2  Homo sapiens hypermethylated in cancer 2 (HI...	41.0	0.59
ref NT_011520.12  Homo sapiens chromosome 22 genomic contig, ...	41.0	0.59
ref NW_001838744.1  Homo sapiens chromosome 22 genomic contig...	41.0	0.59
ref NT_011786.16  Homo sapiens chromosome X genomic contig, G...	39.2	2.0
ref NT_016354.19  Homo sapiens chromosome 4 genomic contig, G...	39.2	2.0
ref NT_005612.16  Homo sapiens chromosome 3 genomic contig, G...	39.2	2.0
ref NT_167186.1  Homo sapiens chromosome 1 genomic contig, GR...	39.2	2.0
ref NW_001842409.1  Homo sapiens chromosome X genomic contig,...	39.2	2.0
ref NW_001838921.1  Homo sapiens chromosome 4 genomic contig,...	39.2	2.0
ref NW_001838881.2  Homo sapiens chromosome 3 genomic contig,...	39.2	2.0
ref NW_001838880.2  Homo sapiens chromosome 3 genomic contig,...	39.2	2.0
ref NW_001838537.2  Homo sapiens chromosome 1 genomic contig,...	39.2	2.0
ref NM_001166052.1  Homo sapiens amyloid beta (A4) precursor ...	37.4	7.1
ref NM_001166051.1  Homo sapiens amyloid beta (A4) precursor ...	37.4	7.1
ref NM_001166054.1  Homo sapiens amyloid beta (A4) precursor ...	37.4	7.1
ref NM_001166053.1  Homo sapiens amyloid beta (A4) precursor ...	37.4	7.1
ref NM_001166050.1  Homo sapiens amyloid beta (A4) precursor ...	37.4	7.1
ref NM_004307.1  Homo sapiens amyloid beta (A4) precursor pro...	37.4	7.1



Strand=Plus/Minus

Regiones encontradas: **Homo sapiens hypermethylated in cancer 2 (HIC2), mRNA**

```
Query 181 TGTAATGCTTATGTAAACCATGATATAAA 210
        ||||| ||||| ||||| ||||| |||||
Sbjct 4898 TGTAAGCTTATGTAAACAATGAGATAAA 4869
```

**NT\_011520.12**

Score = 41.0 bits (44), Expect = 0.59  
Identities = 27/30 (90%), Gaps = 0/30 (0%)  
Strand=Plus/Minus

Regiones encontradas: **2791 bp at 5' side: hypermethylated in cancer 2 protein**

18242 bp at 3' side: hypothetical protein  
LOC100507089

```
Query 181 TGTAATGCTTATGTAAACCATGATATAAA 210
        ||||| ||||| ||||| ||||| |||||
Sbjct 1194421 TGTAAGCTTATGTAAACAATGAGATAAA 1194392
```

**NW\_001838744.1**

Score = 41.0 bits (44), Expect = 0.59  
Identities = 27/30 (90%), Gaps = 0/30 (0%)  
Strand=Plus/Minus

Regiones encontradas: Homo sapiens chromosome 22 genomic contig,  
alternate assembly HuRef SCAF\_1103279180195, whole  
env genome shotgun sequence

```
Query 181 TGTAATGCTTATGTAAACCATGATATAAA 210
        ||||| ||||| ||||| ||||| |||||
Sbjct 7331 TGTAAGCTTATGTAAACAATGAGATAAA 7302
```

8. La búsqueda bioinformática con la secuencia del gen *env* del MMTV se realizó de acuerdo a la siguiente estrategia:

- Una vez descargada la secuencia, se dividió la secuencia en fragmentos de 300 pb de acuerdo a la figura 41 y posteriormente se realizaron las búsquedas en la base de datos del NCBI para cada fragmento resultante bajo las condiciones antes mencionadas.

### Gen env del MMTV

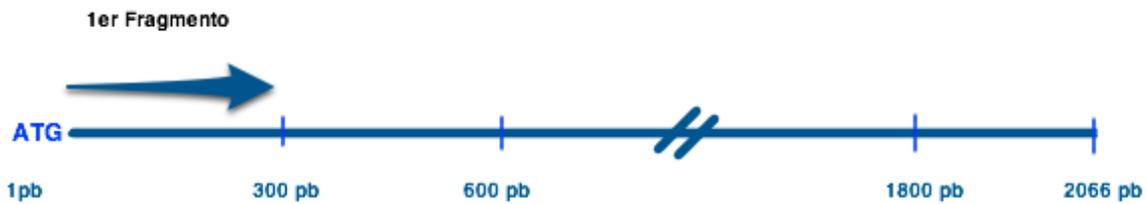


Figura 41. Estrategia de búsqueda del gen *env* del MMTV. La secuencia completa del gen *env* del MMTV fue dividida en fragmentos de 300 pb y posteriormente las búsquedas fueron realizadas.

Los fragmentos generados fueron los siguientes:

Fragmento 1 (1 – 300pb):

```
ATGCCGAAACACCAATCTGGGTCCCCGACCGATTCATCC
GACCTTTTACTGAGCGGAAAGAAGCAACGCCACACCTG
GCACTGCGGAGAAAACGCCGCCGAGATGAGAAAGATC
AACAGGAAAGTCCCAAGAATGAATCTAGTCCCCATCAAA
GAGAAGACGGCTTGGCAACATCTGCAGGCGTTGATCTCC
GAAGCGGAGGAGGTCCTTAAAACCTCACAACTCCCCAA
ACCTCTTTGACCTTATTTCTTGCTTTGTTGTCTGTCTC
GGCCCCCGCCTGTGACAGGGGAGAGTT
```

Fragmento 2 (300-600):

```
TATTGGGCCTACCTACCTAAACCACCTATTCTCCATCCC
GTGGGATGGGGAAGTACAGACCCCATTAGAGTTCTGACA
AATCAAACCATGTATTTGGGTGGTTCGCCTGACTTTCAT
GGGTTTCAGAAATATGTCTGGTAATGTACATTTTGAGGGG
AAGTCTGATACGCTCCCCATTTGCCTTTCTTCTCCTTT
TCTACCCCCACGGGCTGCTTTCAAGTAGACAAGCAAGTA
TTTCTTTCTGATACACCACGGTTGATAATAATAAACCT
GGGGGAAAGGGTGATAAAAAGGCGTATGT
```

Fragmento 3 (600-900):

```
TGGGAACTTTGGTTGACTACCTTGGGGAACTCAGGGGCC
AATACAAAACCTGGTCCCTATAAAAAAGAAGTTGCCCCC
AAATATCCTCAC TGCCAGATCGCCTTTAAGAAGGACGCC
TTCTGGGAGGGAGACGAGTCTGCTCCTCCACGGTGGTTG
CCTTGCGCCTTCCCTGACCAGGGGGTGAGTTTTTCTCCA
AAAGGGGCCCTTGGGTTACTTTGGGATTTCTCCCTTCCC
TCGCCTAGTGTAGATCAGTCAGATCAGATTAAGCAAAA
AAGAATCTATTTGGAAATTATACTCCCC
```

Fragmento 4 (900 -1200):

```
CCTGTCAATAAAGAGGTTTCATCGATGGTATGAAGCAGGA
TGGGTAGAACCTACTTGGTTCTGGGAAAATTCTCCTAAG
```

Fragmento 5 (1200 -1500):

GATCCCAATGATAGAGATTTCACTGCACTAGTCCCCCAT  
ACAGAATTGTTTCGCTTAGTCGCAGCCTCAAGACATCTT  
ATTCTCAAAGGCCAGGATTTCAAGAACATGAGATGATT  
CCTACATCTGCCTGTGTTACTTACCCTTATGCCATATTA  
TTAGGATTACCTCAGTTAATAGATATAGAGAAAAGAGGA  
TCTACTTTTCATATTTCCCTGTTCTTCTT  
TG TAGATTGACTAATTGTTT TAGACTCTTCTGCCTACGAC  
TATGCAGCGATCATAGTCAAGAGGCCGCCATATGTGCTG  
CTACCTGTAGATATTGGTGATGAACCATGGTTTGATGAT  
TCTGCCATTCAAACCTTTAGGTATGCCACAGATTTAATT  
CGAGCTAAGCGATTTCGTGCAGCCATTATCCTGGGCATA  
TCTGCTTTGATTGCTATTATTACTTCCCTTGCTGTAGCT  
ACTACTGCCTTAGTTAAGGAGATGCAAACCTGCTACGTTT  
GTTAATAATCTTCATAGGAATGTTACAT

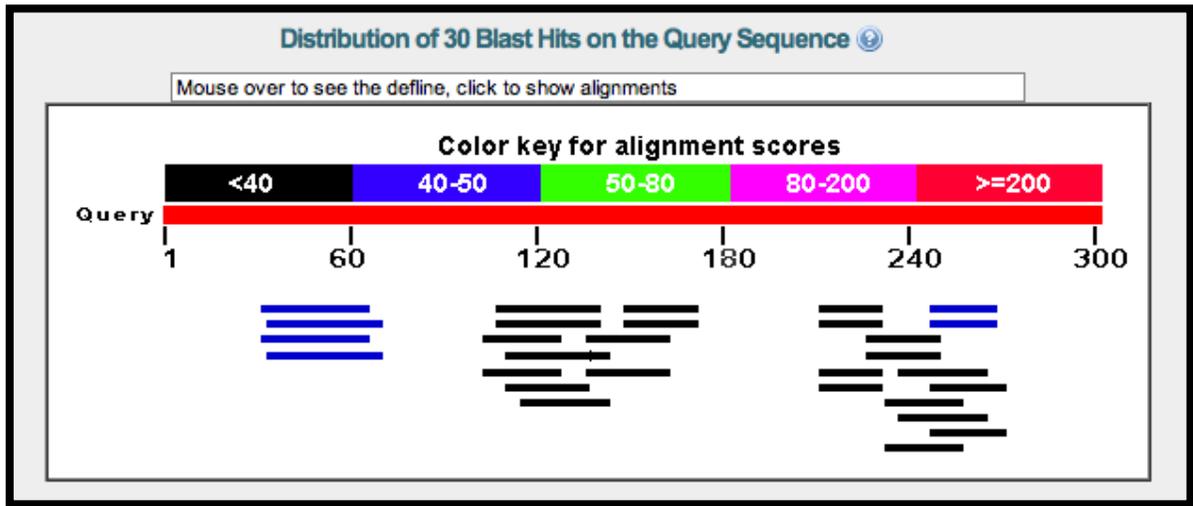
Fragmento 6 (1500 -1800):

TTAGCCTTATCTGAACAACGGATAATAGATTTAAAATTA  
GAAGCTAGACTTAATGCTTTAGAAGAAGTAGTTTTAGAG  
TTGGGACAAGATGTGGCCAATTTAAAGACCAGAATGTCC  
ACTAGGTGTCATGCAAATTATGACTTTATCTGCGTTACA  
CCTTTACCCTATAATGCTACTGAGAACTGGGAAAGAACC  
AGGGCTCATTTATTGGGCATTTGGAATGATAATGAGATT  
TCATATAACATACAGGAGTTAACCAACCTGATTAGTGAT  
ATGAGCAAACAACATATTGATGCAGTGG

Fragmento 7 (1800 -2066):

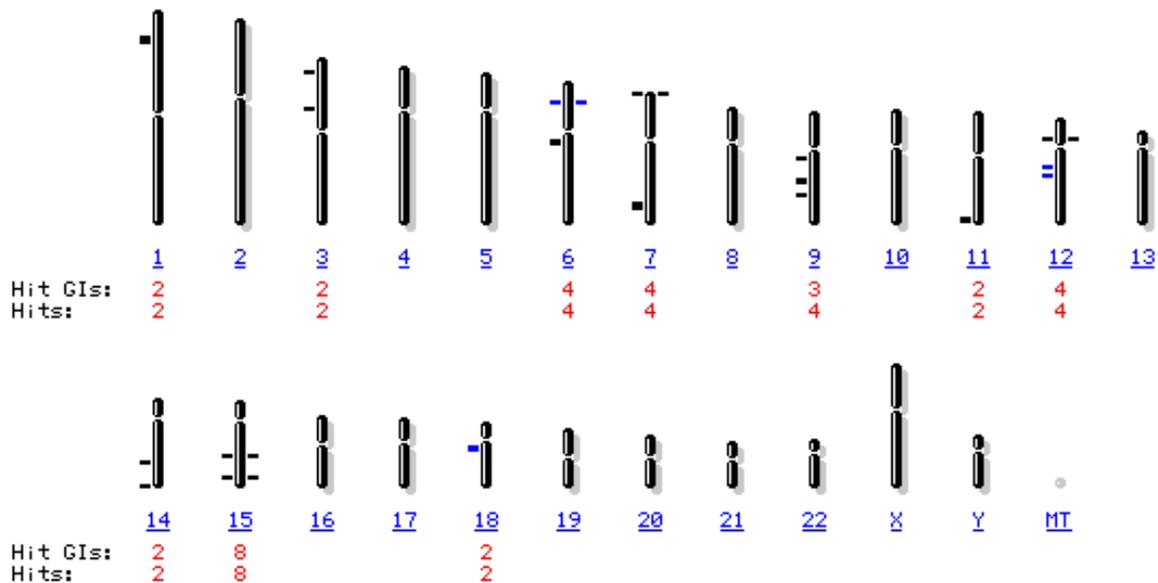
GACCTTAGTGGCTTGGCTCAGTCTTTTGCCAATGGAGTG  
AAGGCTTTAAATCCATTAGATTGGACACAATATTTTCATT  
TTTATAGGTGTTGGAGCCCTGCTTTTAGTCATTGTACTT  
ATGATTTTCCCCATTGTTTTCCAGTGCCTTGCGAAGAGC  
CTTGACCAAGTGCAGTCAGATCTTAACGTGCTTCTTTTA  
AAAAAGAAAAAGGGGGAAATGCCGCGCCTGCAGCAGAA  
ATGGTTGAACTCCCGAGAGTGTCTTACACTTAG

Para el fragmento 1 del gen *env* del MMTV se encontraron 30 hits como se puede observar en la siguiente figura:



**Figura 42. Distribución de los 30 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de la secuencias del gen *env* del MMTV para el fragmento 1.** En la figura se observan 26 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 41.0

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 1 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 43. Distribución genómica de los 30 hits encontrados mediante Blastn para el fragmento 1 del gen env del MMTV.** En la figura se observan 30 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados fueron los siguientes:

Sequences producing significant alignments:	(Bits)	Value
ref NT_010966.14  Homo sapiens chromosome 18 genomic contig, ...	41.0	0.74
ref NT_029419.12  Homo sapiens chromosome 12 genomic contig, ...	41.0	0.74
ref NT_007592.15  Homo sapiens chromosome 6 genomic contig, G...	41.0	0.74
ref NW_001838467.2  Homo sapiens chromosome 18 genomic contig...	41.0	0.74
ref NW_001838060.2  Homo sapiens chromosome 12 genomic contig...	41.0	0.74
ref NW_001838974.1  Homo sapiens chromosome 6 genomic contig,...	41.0	0.74
ref NT_033899.8  Homo sapiens chromosome 11 genomic contig, G...	39.2	2.6
ref NT_007933.15  Homo sapiens chromosome 7 genomic contig, G...	39.2	2.6
ref NT_007819.17  Homo sapiens chromosome 7 genomic contig, G...	39.2	2.6
ref NW_001838044.1  Homo sapiens chromosome 11 genomic contig...	39.2	2.6
ref NW_001839072.1  Homo sapiens chromosome 7 genomic contig,...	39.2	2.6
ref NW_001838998.1  Homo sapiens chromosome 7 genomic contig,...	39.2	2.6
ref NM_018193.2  Homo sapiens Fanconi anemia, complementation...	37.4	9.0
ref NM_001113378.1  Homo sapiens Fanconi anemia, complementat...	37.4	9.0
ref NT_010274.17  Homo sapiens chromosome 15 genomic contig, ...	37.4	9.0
ref NT_026437.12  Homo sapiens chromosome 14 genomic contig, ...	37.4	9.0
ref NT_009714.17  Homo sapiens chromosome 12 genomic contig, ...	37.4	9.0
ref NT_008470.19  Homo sapiens chromosome 9 genomic contig, G...	37.4	9.0
ref NT_007299.13  Homo sapiens chromosome 6 genomic contig, G...	37.4	9.0
Sequences producing significant alignments:	(Bits)	Value
ref NT_022517.18  Homo sapiens chromosome 3 genomic contig, G...	37.4	9.0
ref NT_032977.9  3Homo sapiens chromosome 1 genomic contig, GR...	37.4	9.0
ref NW_001838222.1  Homo sapiens chromosome 15 genomic contig...	37.4	9.0
ref NW_001838115.2  Homo sapiens chromosome 14 genomic contig...	37.4	9.0
ref NW_001838052.1  Homo sapiens chromosome 12 genomic contig...	37.4	9.0

```

ref|NW_001839236.2| Homo sapiens chromosome 9 genomic contig,... 37.4 9.0
ref|NW_001839221.1| Homo sapiens chromosome 9 genomic contig,... 37.4 9.0
ref|NW_001838987.1| Homo sapiens chromosome 6 genomic contig,... 37.4 9.0
ref|NW_001838877.2| Homo sapiens chromosome 3 genomic contig,... 37.4 9.0
ref|NW_001838577.2| Homo sapiens chromosome 1 genomic contig,... 37.4 9.0

```

**Las secuencias encontradas en cada alineamiento fueron las siguientes:**

**NT\_010966.14** Score = 41.0 bits (44), Expect = 0.74  
 Identities = 22/22 (100%), Gaps = 0/22 (0%)  
 Strand=Plus/Minus

Regiones encontradas: 4529 bp at 5' side: nucleolar protein 4  
 528174 bp at 3' side: dystrobrevin alpha isoform 9

```

Query 247 TTATTTCTTGCTTTGTTGTCTG 268
          |||
Sbjct 13296869 TTATTTCTTGCTTTGTTGTCTG 13296848

```

**NT\_029419.12** Score = 41.0 bits (44), Expect = 0.74  
 Identities = 30/35 (86%), Gaps = 0/35 (0%)  
 Strand=Plus/Minus

Regiones encontradas: monocarboxylate transporter 2

```

Query 33 TTCATCCGACCTTTTACTGAGCGGAAAGAAGCAAC 67
          |||
Sbjct 22283218 TTCATCCGACCTTTTACTGGACACAAAGAAGTAAC 22283184

```

**NT\_007592.15** Score = 41.0 bits (44), Expect = 0.74  
 Identities = 32/37 (86%), Gaps = 1/37 (3%)  
 Strand=Plus/Minus

Regiones encontradas: leucine-rich repeat-containing protein 16A isoform 1  
 leucine-rich repeat-containing protein 16A isoform 2

```

Query 35 CATCCGACCTTTTACTGAGCGGAAAGAAGCAACGCC 71
          |||
Sbjct 25319862 CATCC-ACCTTTTAGTGATCAGAAAGAAGCAAAGGCC 25319827

```

**NW\_001838467.2** Score = 41.0 bits (44), Expect = 0.74  
 Identities = 22/22 (100%), Gaps = 0/22 (0%)  
 Strand=Plus/Plus

Regiones encontradas: 527834 bp at 5' side: dystrobrevin alpha isoform 9  
 4529 bp at 3' side: nucleolar protein 4

```

Query 247 TTATTTCTTGCTTTGTTGTCTG 268

```

```

|||||
Sbjct 12733902 TTATTTCTTGCTTTGTTGTCTG 1273392
NW_001838060.2 Score = 41.0 bits (44), Expect = 0.74
Identities = 30/35 (86%), Gaps = 0/35 (0%)
Strand=Plus/Plus

```

Regiones encontradas: monocarboxylate transporter 2

```

Query 33 TTCATCCGACCTTTTACTGAGCGGAAAGAAGCAAC 67
|||||
Sbjct 16378202 TTCATCCGACCTTTTACTGGACACAAAGAAGTAAC 16378236

```

```

NW_001838974.1 Score = 41.0 bits (44), Expect = 0.74
Identities = 32/37 (86%), Gaps = 1/37 (3%)
Strand=Plus/Minus

```

Regiones encontradas: leucine-rich repeat-containing protein 16A isoform 2  
leucine-rich repeat-containing protein 16A isoform 1

```

Query 35 CATCCGACCTTTTACTGAGCGGAAAGAAGCAACGCC 71
|||||
Sbjct 1333804 CATCC-ACCTTTTAGTGATCAGAAAGAAGCAAAGGCC 1333769

```

Para el fragmento 2 del gen *env* del MMTV se encontraron 70 hits como se puede observar en la siguiente figura

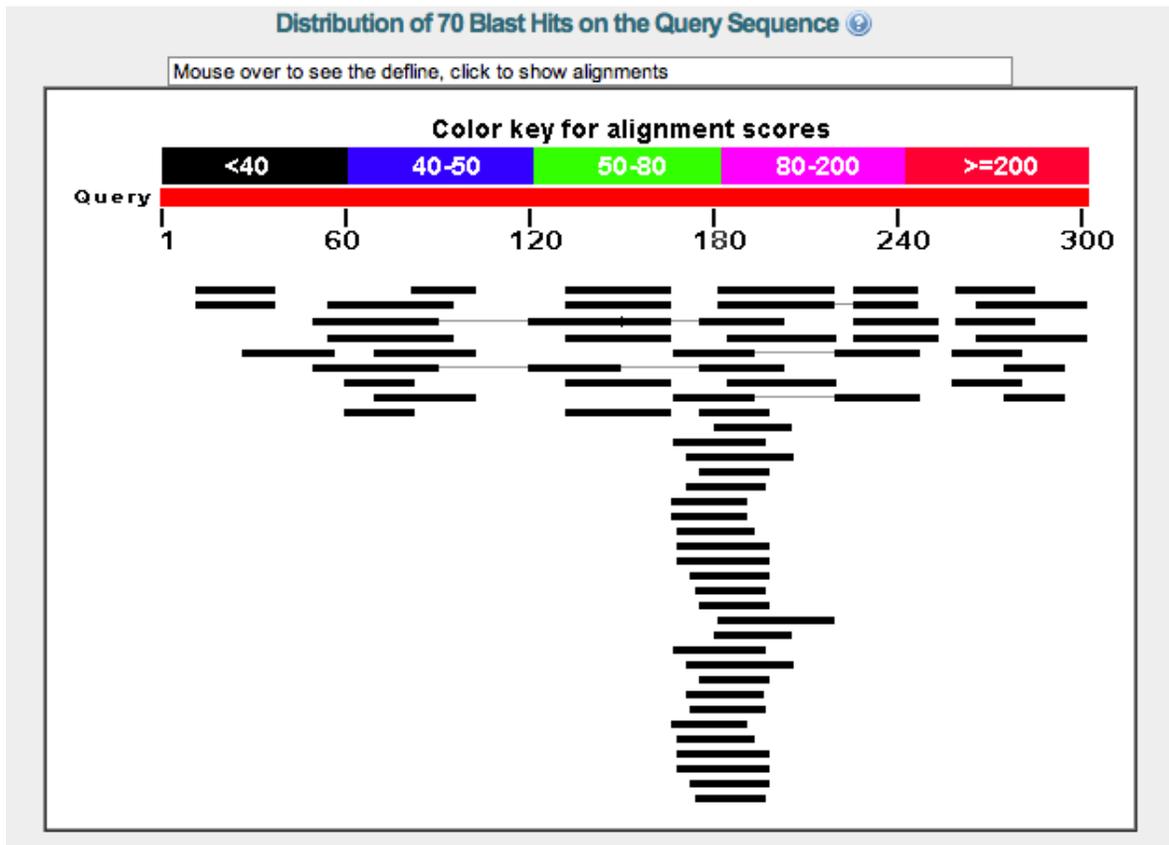
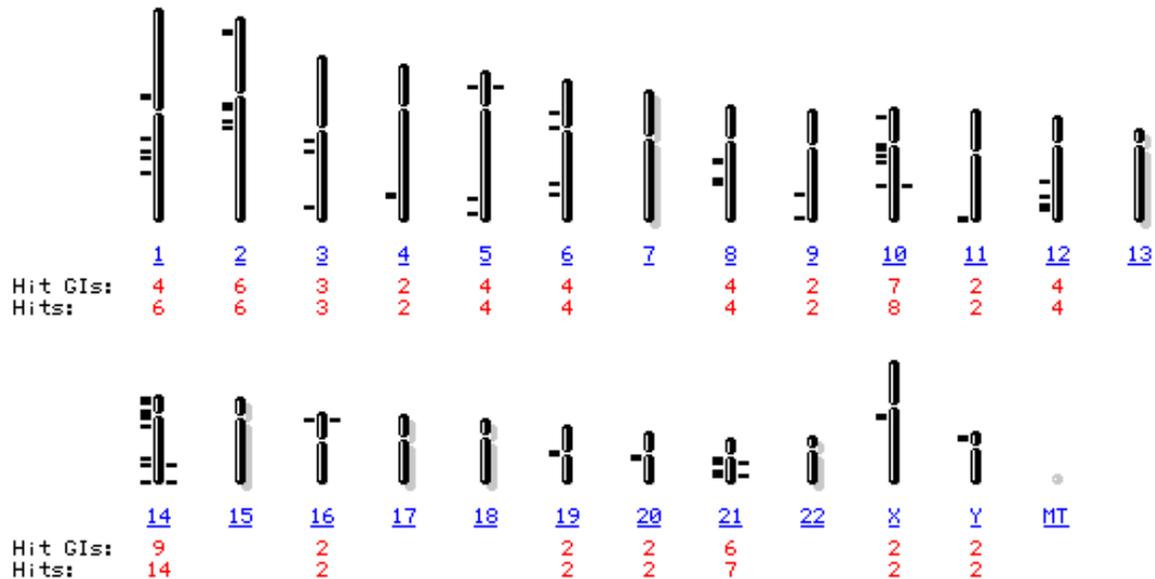


Figura 41. Distribución de los 70 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de la secuencias del gen *env* del MMTV para el fragmento 2. En la figura se observan 70 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 115.0

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 2 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 45. Distribución genómica de los 70 hits encontrados mediante Blastn para el fragmento 2 del gen *env* del MMTV.** En la figura se observan 70 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

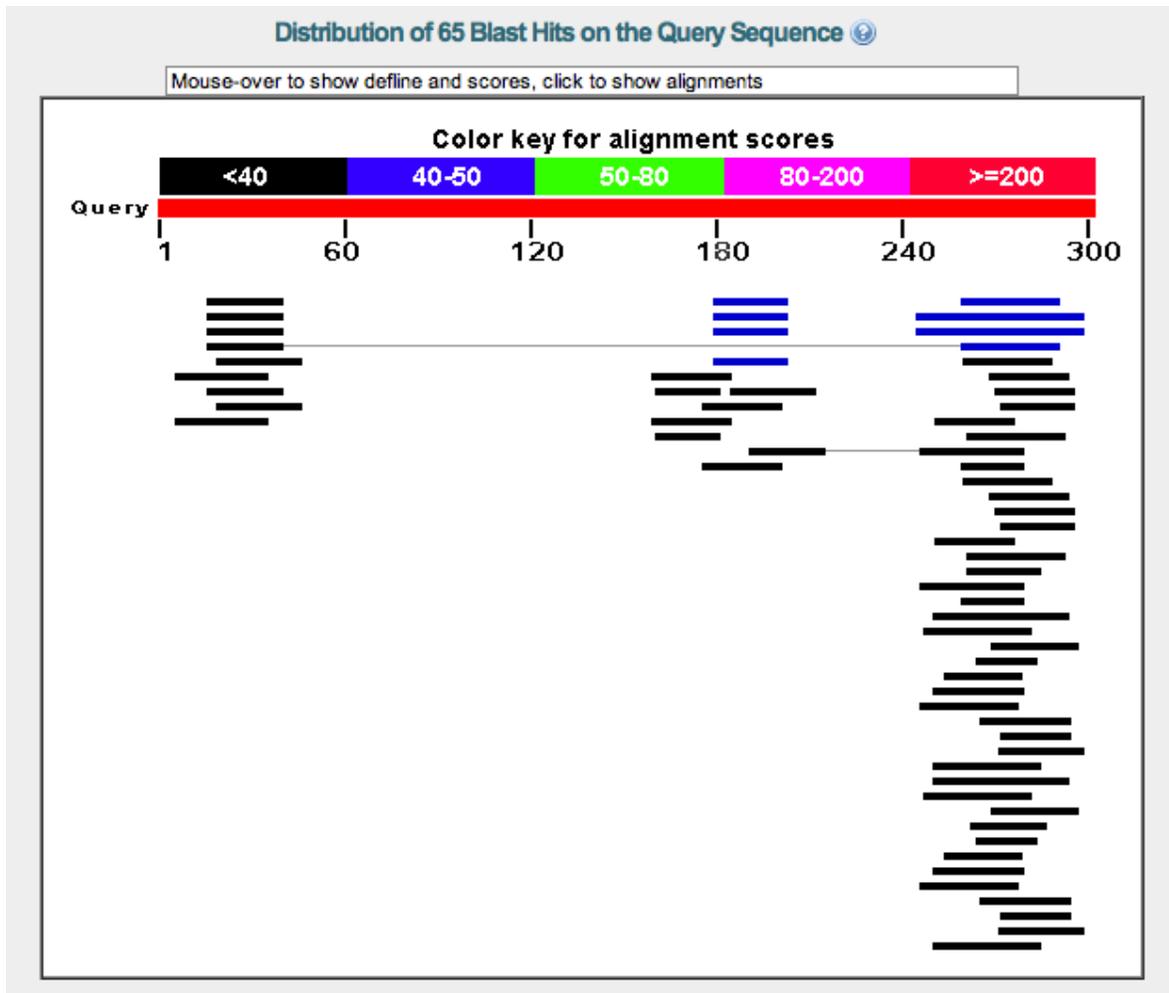
Los alineamientos encontrados fueron los siguientes:

Sequences producing significant alignments:	Score (Bits)	E Value
ref NT_011896.9  Homo sapiens chromosome Y genomic contig, GR...	39.2	2.6
ref NT_011512.11  Homo sapiens chromosome 21 genomic contig, ...	39.2	2.6
ref NT_026437.12  Homo sapiens chromosome 14 genomic contig, ...	39.2	2.6
ref NT_008705.16  Homo sapiens chromosome 10 genomic contig, ...	39.2	2.6
ref NT_008046.16  Homo sapiens chromosome 8 genomic contig, G...	39.2	2.6
ref NT_023133.13  Homo sapiens chromosome 5 genomic contig, G...	39.2	2.6
ref NT_015926.15  Homo sapiens chromosome 2 genomic contig, G...	39.2	2.6
ref NT_004487.19  Homo sapiens chromosome 1 genomic contig, G...	39.2	2.6
ref NT_032977.9  Homo sapiens chromosome 1 genomic contig, GR...	39.2	2.6
ref NW_001842427.1  Homo sapiens chromosome Y genomic contig,...	39.2	2.6
ref NW_001838706.1  Homo sapiens chromosome 21 genomic contig...	39.2	2.6
ref NW_001838110.1  Homo sapiens chromosome 14 genomic contig...	39.2	2.6
ref NW_001839136.1  Homo sapiens chromosome 8 genomic contig,...	39.2	2.6
ref NW_001838954.2  Homo sapiens chromosome 5 genomic contig,...	39.2	2.6
ref NW_001838767.2  Homo sapiens chromosome 2 genomic contig,...	39.2	2.6
ref NW_001838533.2  Homo sapiens chromosome 1 genomic contig,...	39.2	2.6
ref NW_001838590.2  Homo sapiens chromosome 1 genomic contig,...	39.2	2.6

ref NM_152505.3	Homo sapiens Leber congenital amaurosis 5-li...	37.4	9.0
ref NM_183004.3	Homo sapiens eukaryotic translation initiati...	37.4	9.0
ref NM_001969.3	Homo sapiens eukaryotic translation initiati...	37.4	9.0
ref NT_011669.17	Homo sapiens chromosome X genomic contig, G...	37.4	9.0
ref NT_011362.10	Homo sapiens chromosome 20 genomic contig, ...	37.4	9.0
ref NT_011109.16	Homo sapiens chromosome 19 genomic contig, ...	37.4	9.0
ref NT_010393.16	Homo sapiens chromosome 16 genomic contig, ...	37.4	9.0
ref NT_009775.17	Homo sapiens chromosome 12 genomic contig, ...	37.4	9.0
ref NT_029419.12	Homo sapiens chromosome 12 genomic contig, ...	37.4	9.0
ref NT_033899.8	Homo sapiens chromosome 11 genomic contig, G...	37.4	9.0
ref NT_030059.13	Homo sapiens chromosome 10 genomic contig, ...	37.4	9.0
ref NT_030772.10	Homo sapiens chromosome 10 genomic contig, ...	37.4	9.0
ref NT_031847.8	Homo sapiens chromosome 10 genomic contig, G...	37.4	9.0
ref NT_035014.4	Homo sapiens chromosome 9 genomic contig, GR...	37.4	9.0
ref NT_008183.19	Homo sapiens chromosome 8 genomic contig, G...	37.4	9.0
ref NT_025741.15	Homo sapiens chromosome 6 genomic contig, G...	37.4	9.0
ref NT_007592.15	Homo sapiens chromosome 6 genomic contig, G...	37.4	9.0
ref NT_006576.16	Homo sapiens chromosome 5 genomic contig, G...	37.4	9.0
ref NT_016354.19	Homo sapiens chromosome 4 genomic contig, G...	37.4	9.0
ref NT_005612.16	Homo sapiens chromosome 3 genomic contig, G...	37.4	9.0
ref NT_022135.16	Homo sapiens chromosome 2 genomic contig, G...	37.4	9.0
ref NT_022171.15	Homo sapiens chromosome 2 genomic contig, G...	37.4	9.0
ref NW_001842375.2	Homo sapiens chromosome X genomic contig,...	37.4	9.0
ref NW_001838709.2	Homo sapiens chromosome 21 genomic contig...	37.4	9.0
ref NW_001838664.2	Homo sapiens chromosome 20 genomic contig...	37.4	9.0
ref NW_001838491.1	Homo sapiens chromosome 19 genomic contig...	37.4	9.0
ref NW_001838342.1	Homo sapiens chromosome 16 genomic contig...	37.4	9.0
ref NW_001838115.2	Homo sapiens chromosome 14 genomic contig...	37.4	9.0
ref NW_001838063.1	Homo sapiens chromosome 12 genomic contig...	37.4	9.0
ref NW_001838061.2	Homo sapiens chromosome 12 genomic contig...	37.4	9.0
ref NW_001838045.1	Homo sapiens chromosome 11 genomic contig...	37.4	9.0
ref NW_001838005.2	Homo sapiens chromosome 10 genomic contig...	37.4	9.0
ref NW_001837986.1	Homo sapiens chromosome 10 genomic contig...	37.4	9.0
ref NW_001837962.1	Homo sapiens chromosome 10 genomic contig...	37.4	9.0
ref NW_001839241.2	Homo sapiens chromosome 9 genomic contig,...	37.4	9.0
ref NW_001839132.1	Homo sapiens chromosome 8 genomic contig,...	37.4	9.0
ref NW_001838990.2	Homo sapiens chromosome 6 genomic contig,...	37.4	9.0
ref NW_001838981.2	Homo sapiens chromosome 6 genomic contig,...	37.4	9.0
ref NW_001838928.2	Homo sapiens chromosome 5 genomic contig,...	37.4	9.0
ref NW_001838921.1	Homo sapiens chromosome 4 genomic contig,...	37.4	9.0
ref NW_001838884.2	Homo sapiens chromosome 3 genomic contig,...	37.4	9.0
ref NW_001838881.2	Homo sapiens chromosome 3 genomic contig,...	37.4	9.0
ref NW_001838850.1	Homo sapiens chromosome 2 genomic contig,...	37.4	9.0
ref NW_001838822.1	Homo sapiens chromosome 2 genomic contig,...	37.4	9.0

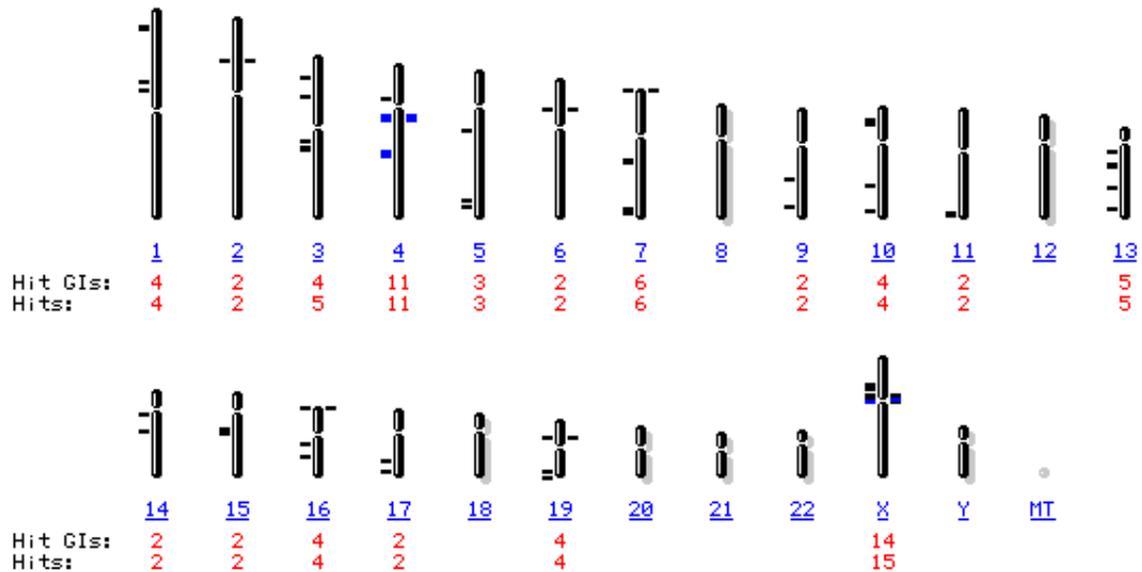
No fue posible encontrar alineamientos con un Valor de e menor a 1.0

Para el fragmento 3 del gen *env* del MMTV se encontraron 65 hits como se puede observar en la siguiente figura



**Figura 46. Distribución de los 65 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de la secuencias del gen *env* del MMTV para el fragmento 3.** En la figura se observan 65 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 115.0

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 3 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 47. Distribución de los 70 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de las secuencias del gen *env* del MMTV para el fragmento 3.** En la figura se observan 70 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 44.6 y 37.4

Los alineamientos encontrados fueron los siguientes:

Sequences producing significant alignments:	Score (Bits)	E Value
ref NM_001012763.1  Homo sapiens gonadotropin-releasing hormo...	44.6	0.061
ref NM_000406.2  Homo sapiens gonadotropin-releasing hormone ...	44.6	0.061
ref NT_011630.14  Homo sapiens chromosome X genomic contig, G...	44.6	0.061
ref NT_022778.16  Homo sapiens chromosome 4 genomic contig, G...	44.6	0.061
ref NW_001842371.1  Homo sapiens chromosome X genomic contig,...	44.6	0.061
ref NW_001838914.1  Homo sapiens chromosome 4 genomic contig,...	44.6	0.061
ref NT_016354.19  Homo sapiens chromosome 4 genomic contig, G...	41.0	0.74
ref NW_001838915.1  Homo sapiens chromosome 4 genomic contig,...	41.0	0.74
ref NT_010393.16  Homo sapiens chromosome 16 genomic contig, ...	39.2	2.6
ref NT_010194.17  Homo sapiens chromosome 15 genomic contig, ...	39.2	2.6
ref NT_026437.12  Homo sapiens chromosome 14 genomic contig, ...	39.2	2.6
ref NT_009952.14  Homo sapiens chromosome 13 genomic contig, ...	39.2	2.6
ref NT_024524.14  Homo sapiens chromosome 13 genomic contig, ...	39.2	2.6
ref NT_030059.13  Homo sapiens chromosome 10 genomic contig, ...	39.2	2.6
ref NT_008470.19  Homo sapiens chromosome 9 genomic contig, G...	39.2	2.6
ref NT_007914.15  Homo sapiens chromosome 7 genomic contig, G...	39.2	2.6
ref NT_023133.13  Homo sapiens chromosome 5 genomic contig, G...	39.2	2.6
ref NT_005612.16  Homo sapiens chromosome 3 genomic contig, G...	39.2	2.6
ref NT_022184.15  Homo sapiens chromosome 2 genomic contig, G...	39.2	2.6

ref NW_001838339.2	Homo sapiens chromosome 16 genomic contig...	39.2	2.6
ref NW_001838218.2	Homo sapiens chromosome 15 genomic contig...	39.2	2.6
ref NW_001838111.1	Homo sapiens chromosome 14 genomic contig...	39.2	2.6
ref NW_001838084.2	Homo sapiens chromosome 13 genomic contig...	39.2	2.6
ref NW_001838074.1	Homo sapiens chromosome 13 genomic contig...	39.2	2.6
ref NW_001838006.2	Homo sapiens chromosome 10 genomic contig...	39.2	2.6
ref NW_001839237.2	Homo sapiens chromosome 9 genomic contig,...	39.2	2.6
ref NW_001839088.2	Homo sapiens chromosome 7 genomic contig,...	39.2	2.6
ref NW_001838954.2	Homo sapiens chromosome 5 genomic contig,...	39.2	2.6
ref NW_001838903.1	Homo sapiens chromosome 4 genomic contig,...	39.2	2.6
ref NW_001838881.2	Homo sapiens chromosome 3 genomic contig,...	39.2	2.6
ref NW_001838769.1	Homo sapiens chromosome 2 genomic contig,...	39.2	2.6
ref NM_015107.2	Homo sapiens PHD finger protein 8 (PHF8), tr...	37.4	9.0
ref NM_001184897.1	Homo sapiens PHD finger protein 8 (PHF8),...	37.4	9.0
ref NM_001184896.1	Homo sapiens PHD finger protein 8 (PHF8),...	37.4	9.0
ref NT_079573.4	Homo sapiens chromosome X genomic contig, GR...	37.4	9.0
ref NT_011109.16	Homo sapiens chromosome 19 genomic contig, ...	37.4	9.0
ref NT_011295.11	Homo sapiens chromosome 19 genomic contig, ...	37.4	9.0
ref NT_010783.15	Homo sapiens chromosome 17 genomic contig, ...	37.4	9.0
ref NT_010498.15	Homo sapiens chromosome 16 genomic contig, ...	37.4	9.0
ref NT_033899.8	Homo sapiens chromosome 11 genomic contig, G...	37.4	9.0
ref NT_008705.16	Homo sapiens chromosome 10 genomic contig, ...	37.4	9.0
ref NT_007933.15	Homo sapiens chromosome 7 genomic contig, G...	37.4	9.0
ref NT_007819.17	Homo sapiens chromosome 7 genomic contig, G...	37.4	9.0
ref NT_007592.15	Homo sapiens chromosome 6 genomic contig, G...	37.4	9.0
ref NT_022517.18	Homo sapiens chromosome 3 genomic contig, G...	37.4	9.0
ref NT_032977.9	Homo sapiens chromosome 1 genomic contig, GR...	37.4	9.0
ref NT_004610.19	Homo sapiens chromosome 1 genomic contig, G...	37.4	9.0
ref NW_001842370.2	Homo sapiens chromosome X genomic contig,...	37.4	9.0
ref NW_001842361.2	Homo sapiens chromosome X genomic contig,...	37.4	9.0
ref NW_001838501.2	Homo sapiens chromosome 19 genomic contig...	37.4	9.0
ref NW_001838484.1	Homo sapiens chromosome 19 genomic contig...	37.4	9.0
ref NW_001838454.2	Homo sapiens chromosome 17 genomic contig...	37.4	9.0
ref NW_001838290.1	Homo sapiens chromosome 16 genomic contig...	37.4	9.0
ref NW_001838081.1	Homo sapiens chromosome 13 genomic contig...	37.4	9.0
ref NW_001838044.1	Homo sapiens chromosome 11 genomic contig...	37.4	9.0
ref NW_001837931.2	Homo sapiens chromosome 10 genomic contig...	37.4	9.0
ref NW_001839064.2	Homo sapiens chromosome 7 genomic contig,...	37.4	9.0
ref NW_001838998.1	Homo sapiens chromosome 7 genomic contig,...	37.4	9.0
ref NW_001838980.1	Homo sapiens chromosome 6 genomic contig,...	37.4	9.0
ref NW_001838951.1	Homo sapiens chromosome 5 genomic contig,...	37.4	9.0
ref NW_001838877.2	Homo sapiens chromosome 3 genomic contig,...	37.4	9.0
ref NW_001838589.2	Homo sapiens chromosome 1 genomic contig,...	37.4	9.0
ref NW_001838573.1	Homo sapiens chromosome 1 genomic contig,...	37.4	9.0

Las secuencias encontradas en cada alineamiento fueron las siguientes:

**NM\_001012763.1**      Score = 44.6 bits (48),    Expect = 0.061  
 Identities = 24/24 (100%), Gaps = 0/24 (0%)  
 Strand=Plus/Plus

Regiones encontradas:    Homo sapiens gonadotropin-releasing hormone  
 receptor (GNRHR), transcript variant 2, mRNA  
 Length=5715

Query    180      GGTGAGTTTTTCTCCAAAAGGGGC    203  
 ||||||||||||||||||||||||||||

Sbjct 3002 GGTGAGTTTTTCTCCAAAAGGGGC 3025

**NM\_000406.2**

Score = 44.6 bits (48), Expect = 0.061  
Identities = 24/24 (100%), Gaps = 0/24 (0%)  
Strand=Plus/Plus

Regiones encontradas: Homo sapiens gonadotropin-releasing hormone  
receptor (GNRHR),  
transcript variant 1, mRNA  
Length=5843

Query 180 GGTGAGTTTTTCTCCAAAAGGGGC 203  
|||||  
Sbjct 3130 GGTGAGTTTTTCTCCAAAAGGGGC 3153

**NT\_011630.14**

Homo sapiens chromosome X genomic contig,  
GRCh37.p2 reference  
primary assembly  
Length=6136098

Regiones encontradas: 398504 bp at 5' side: zinc finger X-linked protein  
ZXDA

Score = 44.6 bits (48), Expect = 0.061  
Identities = 29/32 (91%), Gaps = 0/32 (0%)  
Strand=Plus/Plus

Query 259 CAGATTAAGCAAAAAGAATCTATTTGGAAA 290  
|||||  
Sbjct 5889444 CAGATAAAAAGCAAAAAGAAGCTATTTGTAAA 5889475

**NT\_022778.16**

Score = 44.6 bits (48), Expect = 0.061  
Identities = 24/24 (100%), Gaps = 0/24 (0%)  
Strand=Plus/Minus

Regiones encontradas: 38946 bp at 5' side: ubiquitin-like modifier-  
activating enzyme 6  
392 bp at 3' side: gonadotropin-releasing hormone  
receptor isoform 1

Query 180 GGTGAGTTTTTCTCCAAAAGGGGC 203  
|||||  
Sbjct 8816473 GGTGAGTTTTTCTCCAAAAGGGGC 8816450

**NW\_001842371.1**

Score = 44.6 bits (48), Expect = 0.061  
Identities = 29/32 (91%), Gaps = 0/32 (0%)  
Strand=Plus/Plus

Regiones encontradas: 395819 bp at 5' side: zinc finger X-linked protein  
ZXDA

Query 259 CAGATTAAAAGCAAAAAGAATCTATTTGGAAA 290  
 ||||| |||||||||||||||| ||||||| |||  
 Sbjct 607889 CAGATAAAAAGCAAAAAGAAGCTATTTGTAAA 607920  
**NW\_001838914.1** Score = 44.6 bits (48), Expect = 0.061  
 Identities = 24/24 (100%), Gaps = 0/24 (0%)  
 Strand=Plus/Minus

Regiones encontradas: 38911 bp at 5' side: ubiquitin-like modifier-  
 activating enzyme 6  
 392 bp at 3' side: gonadotropin-releasing hormone  
 receptor isoform 1

Query 180 GGTGAGTTTTTCTCCAAAAGGGGC 203  
 ||||||||||||||||||||  
 Sbjct 8820705 GGTGAGTTTTTCTCCAAAAGGGGC 8820682  
**NT\_016354.19** Score = 41.0 bits (44), Expect = 0.74  
 Identities = 44/56 (79%), Gaps = 2/56 (4%)  
 Strand=Plus/Plus

Regiones encontradas: pituitary homeobox 2 isoform a  
 pituitary homeobox 2 isoform b

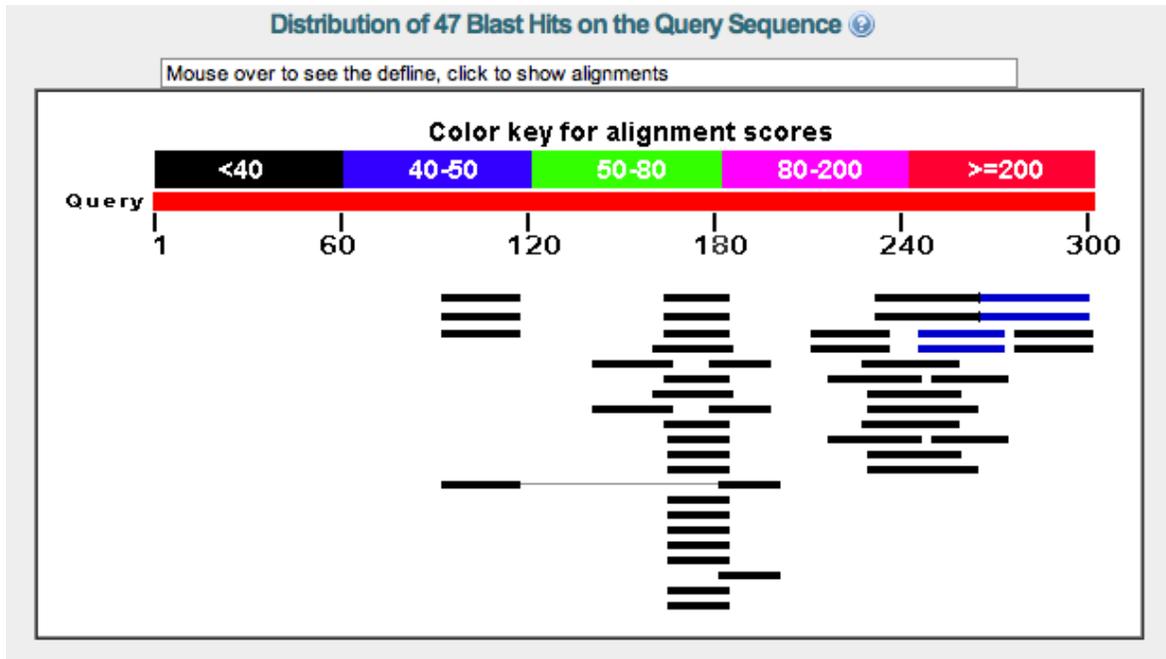
Query 245 TAGATCAGTCAGATCAGATTAAAAGCAAAAAGA-ATCTATTTGGAAAT-TATACTC 298  
 ||||| | ||||| | | ||||||||| | ||||||| | || ||||| ||  
 Sbjct TAGATAATTCAGATTAAAGTAAAAGCAAAACCATATCTATTTGTATATATATATTC  
 36094786 36094841

**NW\_001838915.1** Score = 41.0 bits (44), Expect = 0.74  
 Identities = 44/56 (79%), Gaps = 2/56 (4%)  
 Strand=Plus/Plus

Regiones encontradas: pituitary homeobox 2 isoform a  
 pituitary homeobox 2 isoform b

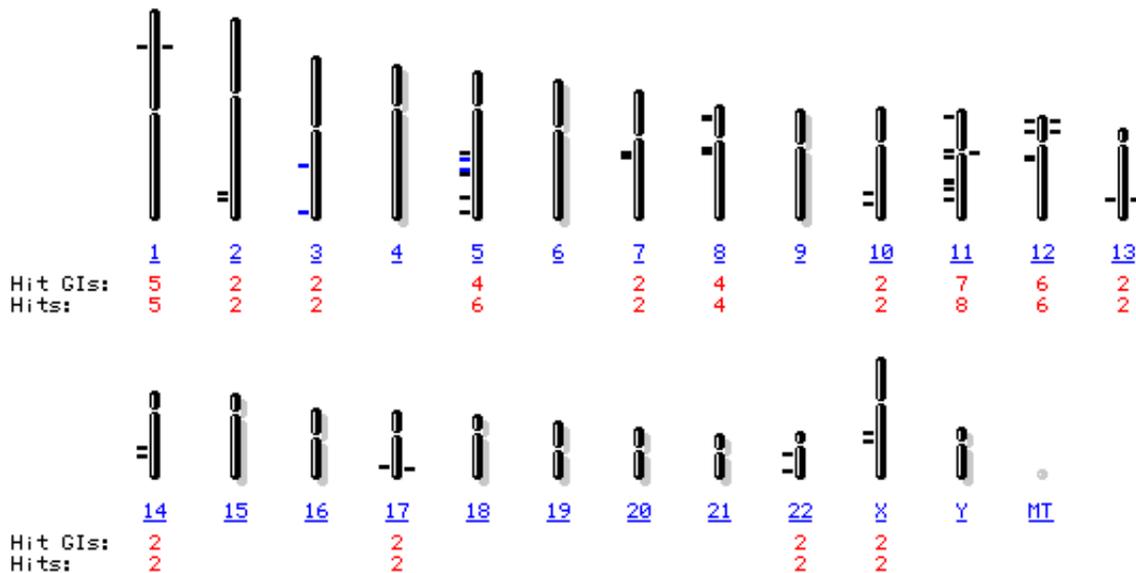
Query 245 TAGATCAGTCAGATCAGATTAAAAGCAAAAAGA-ATCTATTTGGAAAT-TATACTC 298  
 ||||| | ||||| | | ||||||||| | ||||||| | || ||||| ||  
 Sbjct TAGATAATTCAGATTAAAGTAAAAGCAAAACCATATCTATTTGTATATATATATTC  
 36040154 36040209

Para el fragmento 4 del gen *env* del MMTV se encontraron 65 hits como se puede observar en la siguiente figura



**Figura 48. Distribución de los 47 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de la secuencia del gen *env* del MMTV para el fragmento 4.** En la figura se observan 70 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 42.8

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 4 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 49. Distribución genómica de los 47 hits encontrados mediante Blastn para el fragmento 4 del gen *env* del MMTV.** En la figura se observan 47 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados fueron los siguientes:

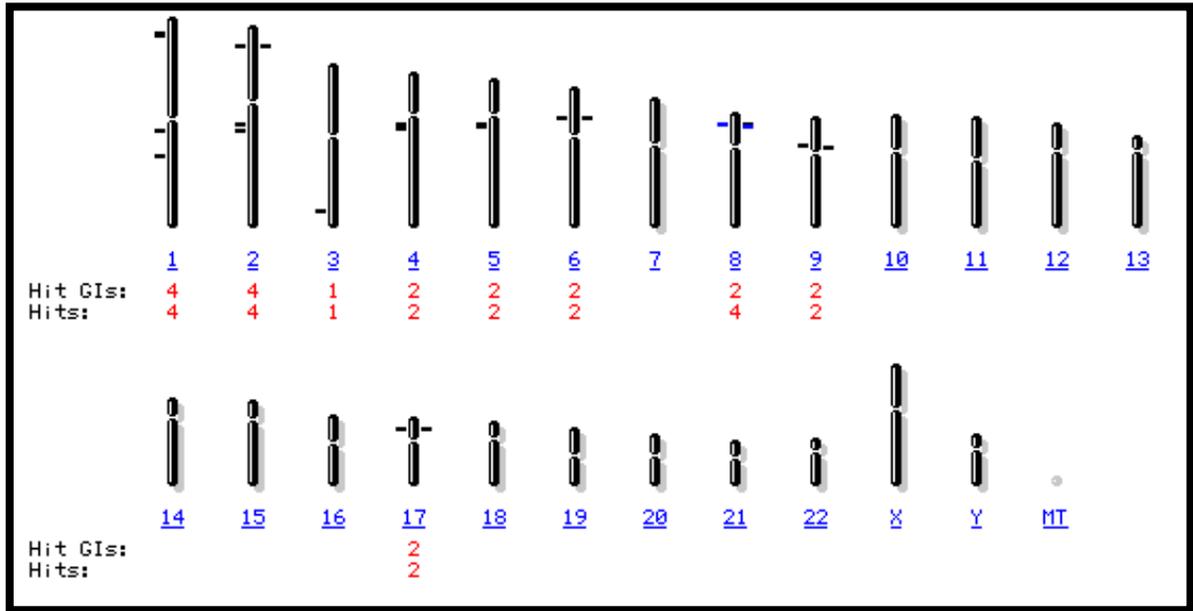
Sequences producing significant alignments:	Score (Bits)	E Value
ref NT_034772.6  Homo sapiens chromosome 5 genomic contig, GR...	42.8	0.21
ref NW_001838952.2  Homo sapiens chromosome 5 genomic contig,...	42.8	0.21
ref NT_005612.16  Homo sapiens chromosome 3 genomic contig, G...	41.0	0.74
ref NW_001838884.2  Homo sapiens chromosome 3 genomic contig,...	41.0	0.74
ref XR_113353.1  PREDICTED: Homo sapiens hypothetical LOC1005...	39.2	2.6
ref XR_112067.1  PREDICTED: Homo sapiens hypothetical LOC1005...	39.2	2.6
ref XR_110507.1  PREDICTED: Homo sapiens hypothetical LOC1005...	39.2	2.6
ref NT_011520.12  Homo sapiens chromosome 22 genomic contig, ...	39.2	2.6
ref NT_005403.17  Homo sapiens chromosome 2 genomic contig, G...	39.2	2.6
ref NT_032977.9  Homo sapiens chromosome 1 genomic contig, GR...	39.2	2.6
ref NW_001838745.1  Homo sapiens chromosome 22 genomic contig,...	39.2	2.6
ref NW_001838863.1  Homo sapiens chromosome 2 genomic contig,...	39.2	2.6
ref NW_001838578.2  Homo sapiens chromosome 1 genomic contig,...	39.2	2.6
ref NT_011651.17  Homo sapiens chromosome X genomic contig, G...	37.4	9.0
ref NT_010783.15  Homo sapiens chromosome 17 genomic contig, ...	37.4	9.0
ref NT_026437.12  Homo sapiens chromosome 14 genomic contig, ...	37.4	9.0
ref NT_009952.14  Homo sapiens chromosome 13 genomic contig, ...	37.4	9.0
ref NT_029419.12  Homo sapiens chromosome 12 genomic contig, ...	37.4	9.0
ref NT_009714.17  Homo sapiens chromosome 12 genomic contig, ...	37.4	9.0
ref NT_009759.16  Homo sapiens chromosome 12 genomic contig, ...	37.4	9.0
ref NT_033899.8  Homo sapiens chromosome 11 genomic contig, G...	37.4	9.0
ref NT_167190.1  Homo sapiens chromosome 11 genomic contig, G...	37.4	9.0





scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 42.8

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 5 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 51. Distribución genómica de los 23 hits encontrados mediante Blastn para el fragmento 5 del gen *env* del MMTV.** En la figura se observan 23 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados fueron los siguientes

Sequences producing significant alignments:	Score (Bits)	E Value
ref NT_167187.1  Homo sapiens chromosome 8 genomic contig, GR...	42.8	0.21
ref NW_001839126.2  Homo sapiens chromosome 8 genomic contig,...	42.8	0.21
ref NT_006713.15  Homo sapiens chromosome 5 genomic contig, G...	39.2	2.6
ref NT_005612.16  Homo sapiens chromosome 3 genomic contig, G...	39.2	2.6
ref NT_022184.15  Homo sapiens chromosome 2 genomic contig, G...	39.2	2.6
ref NW_001838934.1  Homo sapiens chromosome 5 genomic contig,...	39.2	2.6
ref NW_001838768.1  Homo sapiens chromosome 2 genomic contig,...	39.2	2.6
ref NT_010718.16  Homo sapiens chromosome 17 genomic contig, ...	37.4	9.0
ref NT_008413.18  Homo sapiens chromosome 9 genomic contig, G...	37.4	9.0
ref NT_007592.15  Homo sapiens chromosome 6 genomic contig, G...	37.4	9.0
ref NT_022778.16  Homo sapiens chromosome 4 genomic contig, G...	37.4	9.0
ref NT_022135.16  Homo sapiens chromosome 2 genomic contig, G...	37.4	9.0
ref NT_004487.19  Homo sapiens chromosome 1 genomic contig, G...	37.4	9.0
ref NT_004610.19  Homo sapiens chromosome 1 genomic contig, G...	37.4	9.0
ref NW_001838403.1  Homo sapiens chromosome 17 genomic contig...	37.4	9.0

```

ref|NW_001839151.2| Homo sapiens chromosome 9 genomic contig,... 37.4 9.0
ref|NW_001838980.1| Homo sapiens chromosome 6 genomic contig,... 37.4 9.0
ref|NW_001838914.1| Homo sapiens chromosome 4 genomic contig,... 37.4 9.0
ref|NW_001838848.1| Homo sapiens chromosome 2 genomic contig,... 37.4 9.0
ref|NW_001838532.2| Homo sapiens chromosome 1 genomic contig,... 37.4 9.0
ref|NW_001838573.1| Homo sapiens chromosome 1 genomic contig,... 37.4 9.0

```

**Las secuencias encontradas en cada alineamiento fueron las siguientes:**

```

NT_167187.1          Score = 42.8 bits (46), Expect = 0.21
                    Identities = 35/43 (81%), Gaps = 0/43 (0%)
                    Strand=Plus/Plus

```

```

Regiones encontradas: 343206 bp at 5' side: tumor suppressor candidate 3
                    isoform a
                    9028 bp at 3' side: macrophage scavenger receptor
                    types I and II isoform type 1

```

```

Query 192          CATATCTGCTTTGATTGCTATTATTACTTCCTTTGCTGTAGCT 234
                    ||| | | | | | | | | | | | | | | | | | | | | |
Sbjct 3816670     CATATCTGCTCTAATTGCTATTATTTCTTTCCCTCCGCTAGCT 3816712

```

```

NW_001839126.2     Score = 42.8 bits (46), Expect = 0.21
                    Identities = 35/43 (81%), Gaps = 0/43 (0%)
                    Strand=Plus/Minus

```

```

Regiones encontradas: 39952 bp at 5' side: macrophage scavenger receptor
                    types I and II isoform type 2
                    337036 bp at 3' side: tumor suppressor candidate 3
                    isoform b

```

```

Query 192          CATATCTGCTTTGATTGCTATTATTACTTCCTTTGCTGTAGCT 234
                    ||| | | | | | | | | | | | | | | | | | | | | |
Sbjct 5663757     CATATCTGCTCTAATTGCTATTATTTCTTTCCCTCCGCTAGCT 5663715

```

Para el fragmento 6 del gen *env* del MMTV se encontraron 55 hits como se puede observar en la siguiente figura

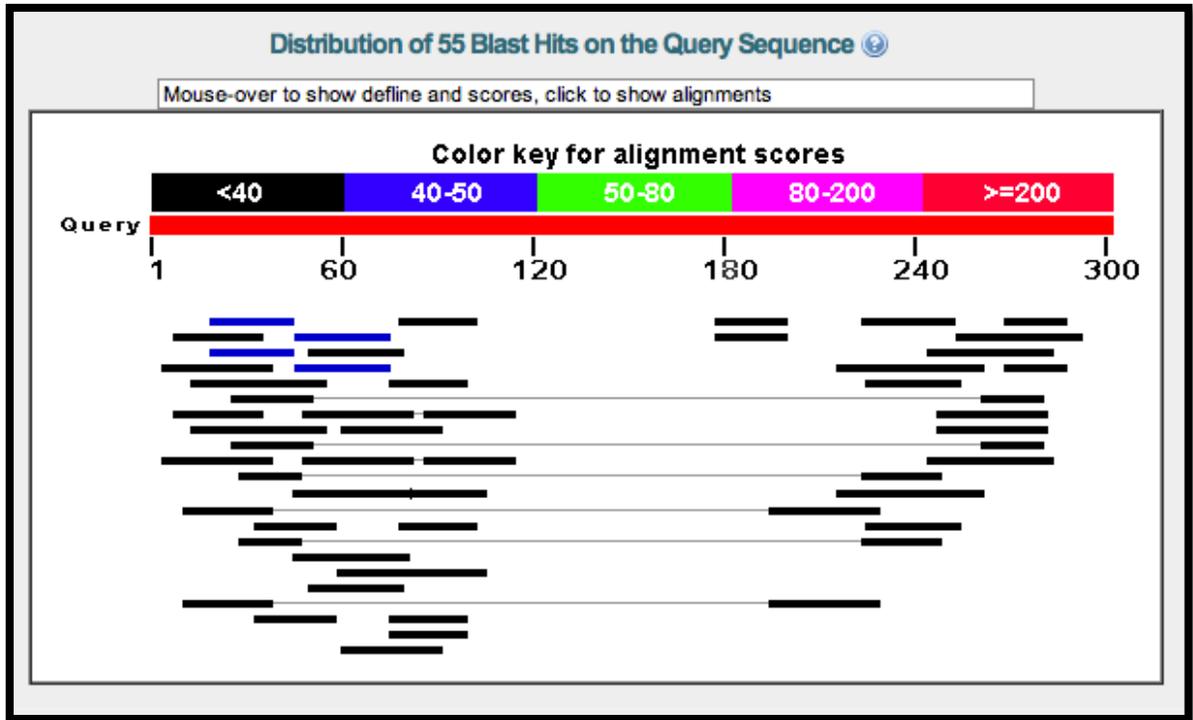
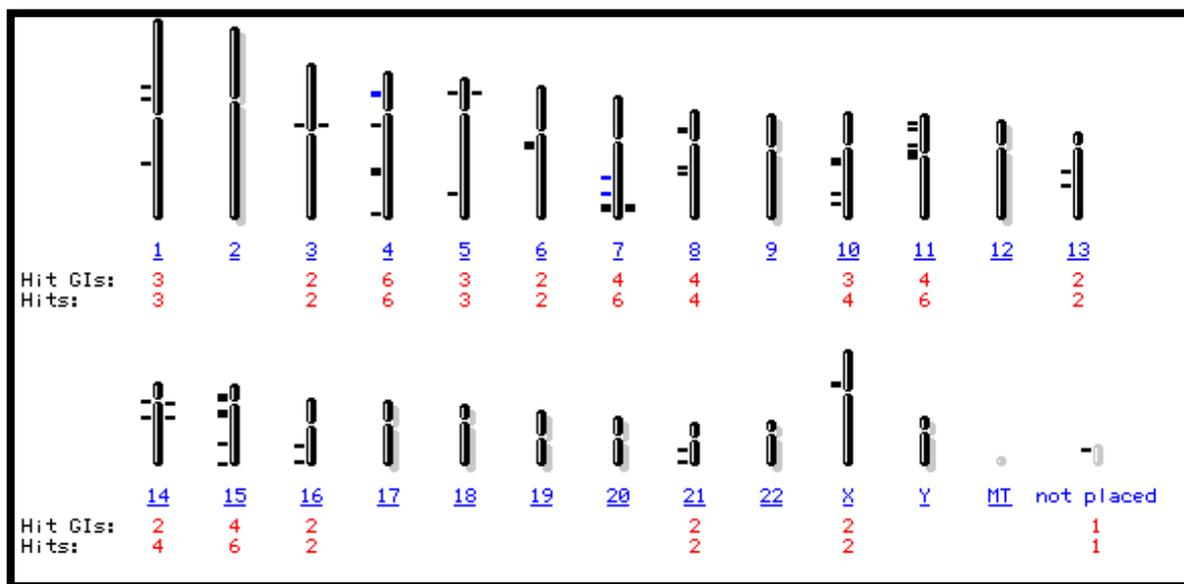


Figura 52. Distribución de los 55 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de la secuencia del gen *env* del MMTV para el fragmento 6. En la figura se observan 23 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 41.0

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 6 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 53. Distribución genómica de los 55 hits encontrados mediante Blastn para el fragmento 6 del gen *env* del MMTV.** En la figura se observan 55 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados fueron los siguientes:

Sequences producing significant alignments:	Score (Bits)	E Value
ref NT_007933.15  Homo sapiens chromosome 7 genomic contig, G...	41.0	0.74
ref NT_006316.16  Homo sapiens chromosome 4 genomic contig, G...	41.0	0.74
ref NW_001839071.2  Homo sapiens chromosome 7 genomic contig,...	41.0	0.74
ref NW_001838900.1  Homo sapiens chromosome 4 genomic contig,...	41.0	0.74
ref NT_010498.15  Homo sapiens chromosome 16 genomic contig, ...	39.2	2.6
ref NT_010194.17  Homo sapiens chromosome 15 genomic contig, ...	39.2	2.6
ref NT_009237.18  Homo sapiens chromosome 11 genomic contig, ...	39.2	2.6
ref NT_004487.19  Homo sapiens chromosome 1 genomic contig, G...	39.2	2.6
ref NW_001838328.2  Homo sapiens chromosome 16 genomic contig...	39.2	2.6
ref NW_001838214.2  Homo sapiens chromosome 15 genomic contig...	39.2	2.6
ref NW_001838022.2  Homo sapiens chromosome 11 genomic contig...	39.2	2.6
ref NW_001838921.1  Homo sapiens chromosome 4 genomic contig,...	39.2	2.6
ref NT_079573.4  Homo sapiens chromosome X genomic contig, GR...	37.4	9.0
ref NT_011515.12  Homo sapiens chromosome 21 genomic contig, ...	37.4	9.0
ref NT_010274.17  Homo sapiens chromosome 15 genomic contig, ...	37.4	9.0
ref NT_026437.12  Homo sapiens chromosome 14 genomic contig, ...	37.4	9.0

```

ref|NT_024524.14| Homo sapiens chromosome 13 genomic contig, ... 37.4 9.0
ref|NT_167190.1| Homo sapiens chromosome 11 genomic contig, G... 37.4 9.0
ref|NT_030059.13| Homo sapiens chromosome 10 genomic contig, ... 37.4 9.0
ref|NT_008183.19| Homo sapiens chromosome 8 genomic contig, G... 37.4 9.0
ref|NT_167187.1| Homo sapiens chromosome 8 genomic contig, GR... 37.4 9.0
ref|NT_007914.15| Homo sapiens chromosome 7 genomic contig, G... 37.4 9.0
ref|NT_007299.13| Homo sapiens chromosome 6 genomic contig, G... 37.4 9.0
ref|NT_006576.16| Homo sapiens chromosome 5 genomic contig, G... 37.4 9.0
ref|NT_016354.19| Homo sapiens chromosome 4 genomic contig, G... 37.4 9.0
ref|NT_022778.16| Homo sapiens chromosome 4 genomic contig, G... 37.4 9.0
ref|NT_022459.15| Homo sapiens chromosome 3 genomic contig, G... 37.4 9.0
ref|NT_032977.9| Homo sapiens chromosome 1 genomic contig, GR... 37.4 9.0
ref|NW_001842362.1| Homo sapiens chromosome X genomic contig,... 37.4 9.0
ref|NW_001839914.1| Homo sapiens unplaced genomic contig, alt... 37.4 9.0
ref|NW_001838716.1| Homo sapiens chromosome 21 genomic contig... 37.4 9.0
ref|NW_001838224.2| Homo sapiens chromosome 15 genomic contig... 37.4 9.0
ref|NW_001838111.1| Homo sapiens chromosome 14 genomic contig... 37.4 9.0
ref|NW_001838081.1| Homo sapiens chromosome 13 genomic contig... 37.4 9.0
ref|NW_001838023.1| Homo sapiens chromosome 11 genomic contig... 37.4 9.0
ref|NW_001838006.2| Homo sapiens chromosome 10 genomic contig... 37.4 9.0
ref|NW_001837986.1| Homo sapiens chromosome 10 genomic contig... 37.4 9.0
ref|NW_001839133.2| Homo sapiens chromosome 8 genomic contig,... 37.4 9.0
ref|NW_001839127.1| Homo sapiens chromosome 8 genomic contig,... 37.4 9.0
ref|NW_001839078.1| Homo sapiens chromosome 7 genomic contig,... 37.4 9.0
ref|NW_001838987.1| Homo sapiens chromosome 6 genomic contig,... 37.4 9.0
ref|NW_001838953.2| Homo sapiens chromosome 5 genomic contig,... 37.4 9.0
ref|NW_001838928.2| Homo sapiens chromosome 5 genomic contig,... 37.4 9.0
ref|NW_001838920.1| Homo sapiens chromosome 4 genomic contig,... 37.4 9.0
ref|NW_001838878.1| Homo sapiens chromosome 3 genomic contig,... 37.4 9.0
ref|NW_001838589.2| Homo sapiens chromosome 1 genomic contig,... 37.4 9.0

```

**Las secuencias encontradas en cada alineamiento fueron las siguientes:**

**NT\_007933.15** Score = 41.0 bits (44), Expect = 0.74  
Identities = 26/27 (96%), Gaps = 1/27 (4%)  
Strand=Plus/Plus

Regiones encontradas: 77350 bp at 5' side: neural Wiskott-Aldrich  
syndrome protein  
42165 bp at 3' side: hyaluronidase-4

```

Query 20          GGATAATAGATTTAAAATTAGAAGCTA 46
          |||||  |||
Sbjct 61498981  GGATAA-AGATTTAAAATTAGAAGCTA 61499006

```

**NT\_006316.16** Score = 41.0 bits (44), Expect = 0.74  
Identities = 27/30 (90%), Gaps = 0/30 (0%)  
Strand=Plus/Plus

Regiones encontradas: 1965812 bp at 5' side: stromal interaction  
molecule 2 isoform 3 precursor  
1734545 bp at 3' side: protocadherin-7 isoform c  
precursor

```

Query 47          GACTTAATGCTTTAGAAGAAGTAGTTTTAG 76
          ||||  ||  |||
Sbjct 20170268  GACTTTATTCTTTAGAAGAAGTTGTTTTAG 20170297

```

**NW\_001839071.2** Score = 41.0 bits (44), Expect = 0.74  
Identities = 26/27 (96%), Gaps = 1/27 (4%)  
Strand=Plus/Minus

Regiones encontradas: 42167 bp at 5' side: hyaluronidase-4  
77485 bp at 3' side: neural Wiskott-Aldrich  
syndrome protein

```
Query 20          GGATAATAGATTTAAAATTAGAAGCTA 46
          |||||  |||||||||||||||||||
Sbjct 7825045     GGATAA-AGATTTAAAATTAGAAGCTA 7825020
```

**NW\_001838900.1** Score = 41.0 bits (44), Expect = 0.74  
Identities = 27/30 (90%), Gaps = 0/30 (0%)  
Strand=Plus/Plus

Regiones encontradas: 1966024 bp at 5' side: stromal interaction  
molecule 2 isoform 3 precursor  
1740141 bp at 3' side: protocadherin-7 isoform c  
precursor

```
Query 47          GACTTAATGCTTTAGAAGAAGTAGTTTTAG 76
          |||||  ||  |||||||||||||||  |||||
Sbjct 19605778     GACTTTATTCTTTAGAAGAAGTTGTTTTAG 19605807
```

Para el fragmento 7 del gen *env* del MMTV se encontraron 50 hits como se puede observar en la siguiente figura

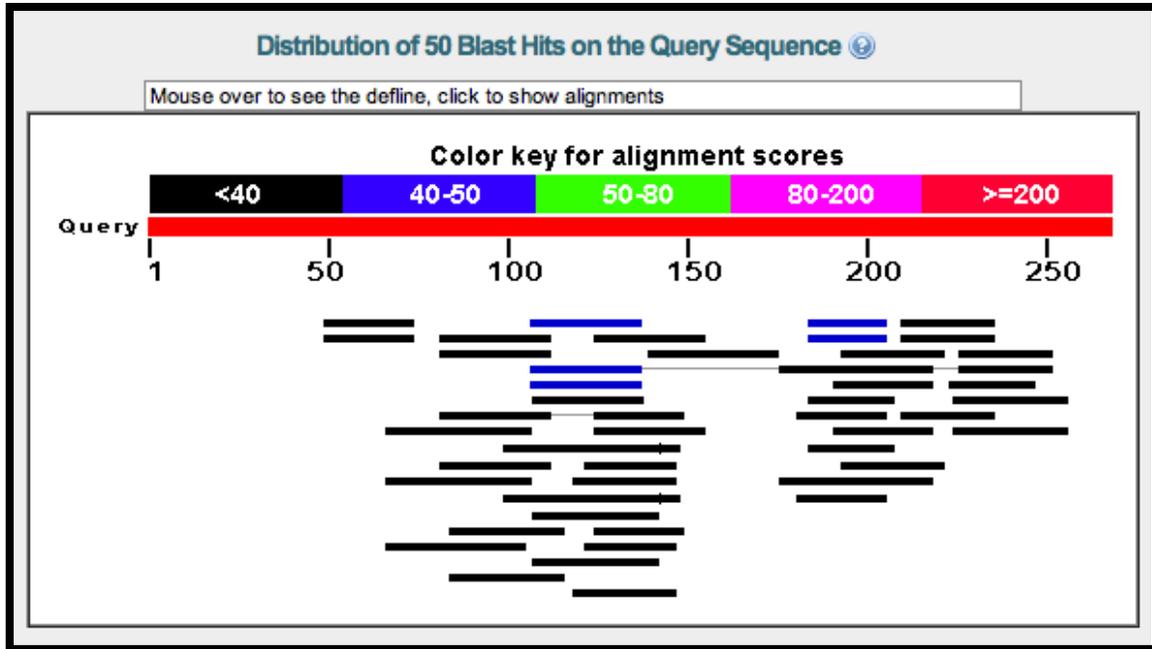
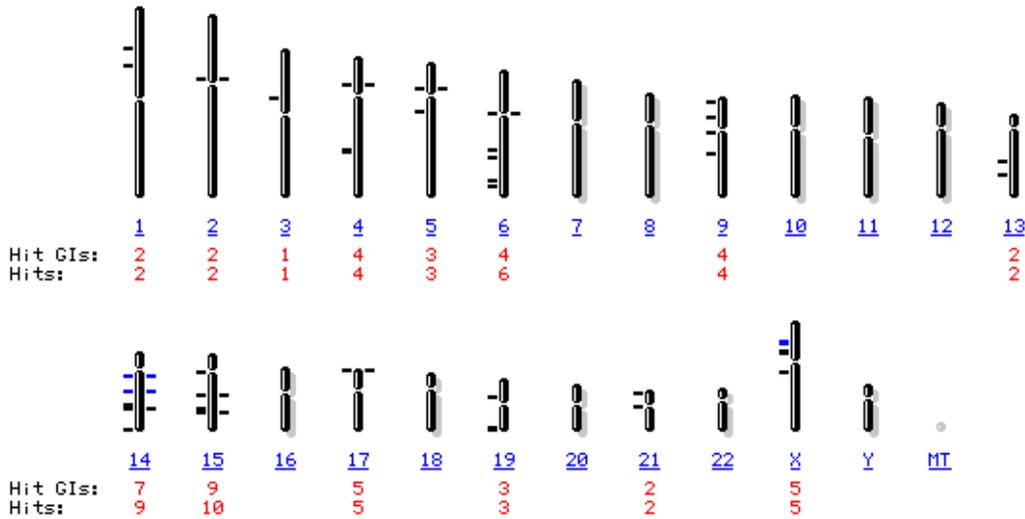


Figura 54. Distribución de los 50 hits encontrados mediante Blastn para la secuencia obtenida de la fragmentación de la secuencia del gen *env* del MMTV para el fragmento 7. En la figura se observan 50 scores relacionados con los alineamientos encontrados mediante Blastn, los scores para cada alineamientos se encuentran entre los valores de 37.4 y 42.8

La distribución genómica de los alineamientos encontrados en la búsqueda realizada con el fragmento 7 del gen *env* del MMTV se muestra en la siguiente figura:



**Figura 55. Distribución genómica de los 50 hits encontrados mediante Blastn para el fragmento 7 del gen env del MMTV.** En la figura se observan 50 hits relacionados con los alineamientos encontrados en cada uno de los cromosomas. Cada cromosoma humano se encuentra representado mediante barras y el número esta indicado en color azul, así mismo el número de hits encontrados en cada cromosoma esta indicado en color rojo.

Los alineamientos encontrados fueron los siguientes:

Sequences producing significant alignments:	Score (Bits)	E Value
ref NM_030755.4  Homo sapiens thioredoxin-related transmembra...	42.8	0.19
ref NT_026437.12  Homo sapiens chromosome 14 genomic contig, ...	42.8	0.19
ref NW_001838111.1  Homo sapiens chromosome 14 genomic contig...	42.8	0.19
ref NT_167197.1  Homo sapiens chromosome X genomic contig, GR...	41.0	0.65
ref NW_001842360.1  Homo sapiens chromosome X genomic contig,...	41.0	0.65
ref NM_024086.3  Homo sapiens methyltransferase 10 domain con...	39.2	2.3
ref NM_198902.2  Homo sapiens tetraspanin 3 (TSPAN3), transcr...	39.2	2.3
ref NM_005724.5  Homo sapiens tetraspanin 3 (TSPAN3), transcr...	39.2	2.3
ref NT_011669.17  Homo sapiens chromosome X genomic contig, G...	39.2	2.3
ref NT_011512.11  Homo sapiens chromosome 21 genomic contig, ...	39.2	2.3
ref NT_010718.16  Homo sapiens chromosome 17 genomic contig, ...	39.2	2.3
ref NT_010194.17  Homo sapiens chromosome 15 genomic contig, ...	39.2	2.3
ref NT_024524.14  Homo sapiens chromosome 13 genomic contig, ...	39.2	2.3
ref NT_025741.15  Homo sapiens chromosome 6 genomic contig, G...	39.2	2.3
ref NT_006576.16  Homo sapiens chromosome 5 genomic contig, G...	39.2	2.3
ref NT_016297.16  Homo sapiens chromosome 4 genomic contig, G...	39.2	2.3
ref NT_032977.9  Homo sapiens chromosome 1 genomic contig, GR...	39.2	2.3
ref NW_001838403.1  Homo sapiens chromosome 17 genomic contig...	39.2	2.3
ref NW_001838219.1  Homo sapiens chromosome 15 genomic contig...	39.2	2.3
ref NW_001838115.2  Homo sapiens chromosome 14 genomic contig...	39.2	2.3
ref NW_001838082.1  Homo sapiens chromosome 13 genomic contig...	39.2	2.3
ref NW_001838990.2  Homo sapiens chromosome 6 genomic contig,...	39.2	2.3
ref NW_001838951.1  Homo sapiens chromosome 5 genomic contig,...	39.2	2.3
ref NW_001838932.2  Homo sapiens chromosome 5 genomic contig,...	39.2	2.3
ref NW_001838901.1  Homo sapiens chromosome 4 genomic contig,...	39.2	2.3
ref NW_001838877.2  Homo sapiens chromosome 3 genomic contig,...	39.2	2.3
ref NW_001838579.2  Homo sapiens chromosome 1 genomic contig,...	39.2	2.3
ref NT_079573.4  Homo sapiens chromosome X genomic contig, GR...	37.4	7.9
ref NT_011109.16  Homo sapiens chromosome 19 genomic contig, ...	37.4	7.9

```

ref|NT_011295.11| Homo sapiens chromosome 19 genomic contig, ... 37.4 7.9
ref|NT_008470.19| Homo sapiens chromosome 9 genomic contig, G... 37.4 7.9
ref|NT_008413.18| Homo sapiens chromosome 9 genomic contig, G... 37.4 7.9
ref|NT_007592.15| Homo sapiens chromosome 6 genomic contig, G... 37.4 7.9
ref|NT_016354.19| Homo sapiens chromosome 4 genomic contig, G... 37.4 7.9
ref|NT_022184.15| Homo sapiens chromosome 2 genomic contig, G... 37.4 7.9
ref|NW_001842362.1| Homo sapiens chromosome X genomic contig,... 37.4 7.9
ref|NW_001838706.1| Homo sapiens chromosome 21 genomic contig... 37.4 7.9
ref|NW_001838501.2| Homo sapiens chromosome 19 genomic contig... 37.4 7.9
ref|NW_001838218.2| Homo sapiens chromosome 15 genomic contig... 37.4 7.9
ref|NW_001838113.2| Homo sapiens chromosome 14 genomic contig... 37.4 7.9
ref|NW_001839221.1| Homo sapiens chromosome 9 genomic contig,... 37.4 7.9
ref|NW_001839149.2| Homo sapiens chromosome 9 genomic contig,... 37.4 7.9
ref|NW_001838983.1| Homo sapiens chromosome 6 genomic contig,... 37.4 7.9
ref|NW_001838920.1| Homo sapiens chromosome 4 genomic contig,... 37.4 7.9
ref|NW_001838769.1| Homo sapiens chromosome 2 genomic contig,... 37.4 7.9

```

**Las secuencias encontradas en cada alineamiento fueron las siguientes:**

**NM\_030755.4** Score = 42.8 bits (46), Expect = 0.19  
Identities = 28/31 (90%), Gaps = 0/31 (0%)  
Strand=Plus/Minus

Regiones encontradas: Homo sapiens thioredoxin-related transmembrane protein 1 (TMX1), mRNA  
Length=4119

```

Query 107 TCATTGTA CTTATGATTTTCCCCATTGTTTT 137
          |||||  |||||  |||||  ||  ||||
Sbjct 2263 TCATTGTA TTTATGATTTTCCCCCTTCTTTT 2233

```

**NT\_026437.12** Score = 42.8 bits (46), Expect = 0.19  
Identities = 28/31 (90%), Gaps = 0/31 (0%)  
Strand=Plus/Minus

Regiones encontradas: 1265 bp at 5' side: thioredoxin-related transmembrane protein 1 precursor  
395420 bp at 3' side: putative uncharacterized protein C14orf82

```

Query 107 TCATTGTA CTTATGATTTTCCCCATTGTTTT 137
          |||||  |||||  |||||  ||  ||||
Sbjct 32722532 TCATTGTA TTTATGATTTTCCCCCTTCTTTT 32722502

```

**NW\_001838111.1** Score = 42.8 bits (46), Expect = 0.19  
Identities = 28/31 (90%), Gaps = 0/31 (0%)  
Strand=Plus/Minus

Regiones encontradas: 1260 bp at 5' side: thioredoxin-related transmembrane protein 1 precursor  
395943 bp at 3' side: putative uncharacterized protein C14orf82

```
Query 107          TCATTGTA CTTATGATTTTCCCCATTGTTTT 137
||||| |||||
Sbjct 10583539    TCATTGTATTTATGATTTTCCCCCTTCTTTT 10583509
```

**NT\_167197.1**

Score = 41.0 bits (44), Expect = 0.65  
Identities = 22/22 (100%), Gaps = 0/22 (0%)  
Strand=Plus/Plus

Regiones encontradas: dystrophin Dp260-2 isoform  
dystrophin Dp260-1 isoform

```
Query 184          GTGCTTCTTTTaaaaaagaaaa 205
|||||
Sbjct 29786730    GTGCTTCTTTTAAAAAAGAAAA 29786751
```

**NW\_001842360.1**

Score = 41.0 bits (44), Expect = 0.65  
Identities = 22/22 (100%), Gaps = 0/22 (0%)  
Strand=Plus/Plus

Regiones encontradas: dystrophin Dp260-2 isoform  
dystrophin Dp260-1 isoform

```
Query 184          GTGCTTCTTTTaaaaaagaaaa 205
|||||
Sbjct 23433345    GTGCTTCTTTTAAAAAAGAAAA 23433366
```

## 8. Discusión

Los primeros reportes en los cuales se evidenciaba la presencia de secuencias virales parecidas a las secuencias del gen *env* del MMTV en adenocarcinomas mamarios de humanos aparecieron hace poco más de 15 años en los reportes publicados por Wang Y y colaboradores (Wang Y y col., 1995). A partir de ahí diversos trabajos han sido publicados evidenciando la presencia de secuencias de DNA del MMTV en muestras de adenocarcinomas mamarios en humanos (Etkind P y col., 2000, Melana S y col., 2001, Lui B y col., 2001, Melana y col., 2002, Ford C y col., 2003, Mant C y col., 2004, Levine P y col., 2004, Luo T y col., 2006, Brinda A y col., 2007, Zapata-Benavides y col., 2007). La gran mayoría de éstos estudios se enfocan en la amplificación de una secuencia de 685 pb del gen *env* del MMTV y una secuencia interna de 250pb empleando una PCR anidada. La principal ventaja de realizar una PCR anidada es aumentar la especificidad de la reacción y lograr detectar cantidades muy pequeñas del genoma del retrovirus (probablemente hasta  $10^3$  células). Estas secuencias amplificadas contienen muy baja similitud con regiones de retrovirus endógenos conocidos y se han utilizado para detectar DNA proviral y secuencias de expresión del retrovirus MMTV.

En México, los datos de incidencia reportados por Zapata Benavides y colaboradores en 2007 con relación a la detección de las secuencias parecidas al gen *env* del MMTV suponen una relación mucho menor a otros estudios reportados incluyendo el nuestro, ellos reportan haber encontrado secuencias parecidas al gen *env* del MMTV en 5 de 119 (4.2%) tumores de mama y en ninguno de los tejidos no afectados estudiados. Nuestros datos encontrados en 16 (40% de las 40 muestras de adenocarcinomas estudiados), cero muestras positivas en los 14 tejidos no afectados y cero muestras positivas en los 2 fibroadenomas mamarios mantienen la tendencia de los datos reportados por Wang Y en Estados Unidos en 1998 (36% de muestras *env* positivas), Melana S en Argentina en 2002 (31% de muestras positivas), Levine H en Italia en 2004 (38% muestras *env* positivas).

Nosotros consideramos que esta diferencia en los porcentajes reportados entre los dos estudios realizados en México radica principalmente en el origen de las muestras empleadas, la población de estudio en el trabajo publicado por Zapata Benavides pertenece a la región noreste del país (una región geolocalizada, un solo tipo de población) mientras que las muestras empleadas en nuestro estudio provienen de una clínica de referencia, sin una población definida (las muestras provienen de personas de diversos estados de la república mexicana) probablemente esta diferencia entre el origen de las muestras de estudio haya incrementado los porcentajes de detección de las secuencias *env* positivas en nuestro estudio.

Como se puede observar en las figuras 14 -17, los amplificadores correspondientes a cada una de las muestras de adenocarcinomas mamarios tienen un tamaño aproximado de 250pb de igual forma que el control positivo (muestra amplificada del plásmido construido con el gen *env* del MMTV C3H) de acuerdo al diseño de realizado con los oligonucleótidos sin embargo, es notorio que la concentración de los amplificadores no es la misma como se puede observar en los amplificadores de las muestras 007CM y 004CM (figura 14). Esto también es notorio en las muestras 031CM y 032CM (figura 17). Estos datos sugieren una mayor concentración del amplificador en las muestras 007 y 032, al cargar en cada pozo el mismo volumen de reacción. Este comportamiento no se observa para los amplificadores del gen GAPDH de las muestras 007 y 004 (figura 18). Esta relación implicaría que las secuencias MMTV integradas en el genoma humano no son integradas en la misma cantidad en cada una de las muestras mamarias (es necesario utilizar otra metodología para corroborar dichos datos, qPCR). Ford C y col., en el 2004 sugirieron la idea de que existe una relación entre el grado de malignidad del adenocarcinoma de mama y el porcentaje de secuencias relacionadas con el MMTV sin embargo, estos datos no pudieron ser verificados en el presente estudio.

Los análisis realizados a las secuencias amplificadas del gen *env* del MMTV en los adenocarcinomas mamarios muestran una similitud de hasta un 99% con respecto

a la secuencia depositada en el NCBI para el gen *env* del MMTV ver figuras 26 y 28. Estos porcentajes de similitud entre las secuencias encontradas a partir de los amplificadas de los adenocarcinomas mamarios estudiados y las secuencias depositadas en el NCBI para el MMTV se encuentran en el mismo rango que los datos publicados por Wang y col., en 1995, Melana S y col., 2001, Lui B y col., 2001, Melana y col., 2002, Ford C y col., 2003, , Levine P y col., 2004 con datos entre 95 – 98% de similitud entre secuencias.

Adicionalmente, nosotros encontramos un amplificado cuya secuencia correspondía con un 99% de similitud con respecto al gen *env* del HMTV (EF495356.1) ver figura 28. Esta similitud cercana al 100% entre las secuencias obtenidas nos llevó a preguntarnos si existe homología entre dichas secuencias. El árbol filogenético construido a partir de las distancias calculadas por el alineamiento hecho con las secuencias amplificadas de los adenocarcinomas muestra claramente tres principales ramas en las que son organizadas las secuencias, una de ellas contiene a la secuencia depositada en el NCBI para el MMTV (MMTV NC 001503.1) y las secuencias obtenida a partir de las muestras tumorales 027CM y 007CM (ambas muestras con un diagnóstico de Carcinoma ductal infiltrante), y en otra rama se ven organizadas las secuencias obtenidas de los adenocarcinomas mamarios 032CM, 031CM y la secuencia del control positivo (construcción hecha con el gen *env* del MMTVC3H) con una distancia muy parecida entre ellas, lo cual indicaría una distancia evolutiva más cercana entre ellas con respecto a las secuencia MMTV\_ NC001503.1. Estos datos confirman los datos publicados anteriormente, las secuencias amplificadas son similares pero no iguales al MMTV y manifiestan homología. Este hecho sustentaría la idea de un homólogo humano para el MMTV el cual ha sido denominado con anterioridad Virus del tumor mamario en humanos (HMTV de sus siglas en inglés *Human Mammary Tumor Virus*). Estos hechos podrían tener implicaciones mucho más serias al sugerir un evento evolutivo conjunto entre los genomas del MMTV y el HMTV mediante un ancestro común y como resultado la aparición de genes ortólogos para el gen *env* del MMTV y el HMTV.

Aunque existe la misma tendencia de encontrar tumores mamarios con secuencias de alta homología (98%-99% de similitud entre secuencias) relacionadas al gen *env* del MMTV los porcentajes de detección son heterogeneos como se puede observar en los datos de incidencia publicados por Zapata P en 2006, en dicho trabajo es posible analizar desde una perspectiva geográfica los datos reportados por Wang y col. en 1988, Melana y col. en el 2002, Mant y col. en el 2004, Levine y col. en el 2004 y Ford y col. en el 2003. Los porcentajes de incidencia son bastante heterogeneos desde un 74% reportado en Túnez hasta un 0% reportado en Reino unido y un 4.2% reportado en México en un primer trabajo por el grupo de Zapata-Benavides y col. Estos datos demuestran que los porcentajes de incidencia son heterogeneos y no presentan aparentemente

Los reportes anteriores destacan los porcentajes de detección de las secuencias *env* en los tejidos de mama no afectados (1-4% de las muestras analizadas). Nuestros datos encontrados en la detección de secuencias *env* del MMTV siguen esta misma tendencia (0% de muestras *env* positivas en los tejidos no afectados estudiados). Estos datos sugieren fuertemente una posible asociación entre la presencia de las secuencias *env* en los tumores mamarios y no en los tejidos no afectados del mismo paciente, lo que conllevaría una infección por el retrovirus MMTV al no encontrarse dicha secuencia en los tejidos no afectados del mismo paciente. Sugerentemente la secuencia permanece de forma exogena tal y como lo habia indicado Melana S y col. en 2001.

Analizando esta aparente asociación geográfica Steward T y colaboradores en el año 2000, encontró que la mayor tasa de incidencia de cáncer de mama en humanos se localizaba en lugares donde *Mus domesticus* era un residente común con los humanos. Notando que en el 50% de los ratones *Mus domesticus*, provenientes del sur California (USA), eran detectables regiones MMTV exógenas. Stewart y colaboradores sugirieron que una alta concentración de *Mus domesticus* en la población humana puede estar relacionado con una mayor incidencia del MMTV en los tumores mamarios de humanos. Lo que colocaría necesariamente a *Mus domesticus* como agente infeccioso del MMTV en humanos.

Sin embargo, aún es necesario aclarar los mecanismos por los cuales el MMTV pudiera estar infectando células humanas. Las principales dudas sobre esta asociación entre la infección con MMTV y los adenocarcinomas mamarios radican en:

- El tumor murino debido al MMTV como modelo para los carcinomas de mama humanos está lejos de ser ideal, ya que la histopatología de estas lesiones es diferente: el primero se asemeja más una hiperplasia que a un carcinoma ductal infiltrante (el tipo de cáncer mamario más común en humanos).
- En segundo lugar, si el virus humano MMTV(HMTV) tiene un ciclo de vida similar al MMTV, se esperaría que la transmisión a través de la leche materna debería ocurrir (como ocurre de la forma endógena en los ratones infectados con MMTV). Titus-Ernstoff L en 1998 demostró en un amplio estudio de casos y controles en 8 300 mujeres con cáncer de mama que la lactancia materna por sus madres no fue un factor de riesgo en ellas, sin embargo partículas virales del MMTV han sido descritas en la leche materna humana (Titus-Ernstoff L y col., 1998, Johal H y col., 2010) .

Los efectos mutagénicos del MMTV dependen del sitio de integración del DNA proviral en el genoma humano, lo que da como resultado múltiples mutaciones y la desregulación de WNT (una familia de genes que codifican factores de señalización extracelular y participan en la regulación del crecimiento de las células embrionarias y en procesos de diferenciación). En las células humanas el DNA del MMTV se integra en múltiples sitios en el genoma: en los cromosomas 1, 2, 3, 6, 7, 10, 11, 12, 17 y 22. Resulta de particular interés una región en el cromosoma 17, donde ocurren dos sitios de integración (WNT15 y WNT3), estos sitios de inserción flanquean estrechamente un gen de susceptibilidad a cáncer mamario BRCA-1. Esta pequeña región en 17q21 también contiene varios oncogenes, incluyendo dos que se asocian con cáncer de mama: HER-2/neu [v-erb-b2 aviar eritroblásticos leucemia homólogo del oncogén viral 2 y RAB5C (un miembro de la familia de oncoproteínas RAS), quienes conjuntamente con el factor

de crecimiento de fibroblastos 3 (FGF3) resultan fundamentales para la inducción de tumores mamarios murinos.

Recientemente en 2007 el grupo de trabajo de Theodorou V, reportaron 33 sitios de inserción comunes los cuales fueron identificados mediante un “screening masivo” en tumores mamarios inducidos por MMTV (Theodorou V y col, 1997).

Los genes encontrados en las búsquedas bioinformáticas realizadas y que posiblemente estén involucrados en inserciones virales del MMTV sugieren diversos procesos celulares y distintas rutas metabólicas involucradas. Los genes precursores de Caderina 7 (cadherin-7 NW\_00183469.1) fueron encontrados 4 veces (Valor  $e = 0.40$ ) situados sobre el cromosoma 18 y que al igual que el transcrito NT\_004487.19 (Molécula de adhesión celular isoformas 1 y 2, valor  $e = 0.17$ ) están implicados en los procesos de adhesión celular, mecanismo implicado fuertemente en el desarrollo del cáncer y metástasis como consecuencia de la pérdida del señalamiento celular y la inflamación. Otra región encontrada fue la correspondiente al gen de la proteína Serpin B8 isoformas a y b (serin proteinase inhibitors, valor  $e = 0.40$ ) las cuales corresponde a un grupo de proteínas que regulan un conjunto de procesos intra y extracelulares tales como la fibrinólisis, coagulación, diferenciación celular, supresión celular, apoptosis y migración celular. También fueron encontradas las regiones correspondientes al gen HIC2 (Homo sapiens hypermethylated in cancer 2, valor de  $e = 0.59$ ). De acuerdo a los datos reportados por Theodorou V y col. en 2007 ellos reportan 33 sitios comunes de inserción y genes de las familias *FGT*, *RSPO3*, *WNT*, *ASTN2*, *ATP2b1* susceptibles de inserciones por MMTV. Nosotros logramos amplificar una región correspondiente al proto-oncogen *WNT-1* (NT\_029419.12 ver secuencia seq012) y el *FGF* (NT\_030059.13, secuencia seq003) lo cual corresponde a los sitios reportados anteriormente como sitios susceptibles de una inserción por MMTV.

Los genes encontrados mediante PCR splinkerette en los adenocarcinomas mamarios *env* positivos (NM\_001012763.1 y NT\_034772.6 correspondientes al receptor de la hormona liberadora de gonadotropina (GNRHR) y a la Pirofosfatasa

peroxisomal de NADH (NUDT12) implicada en el metabolismo energético) también fueron encontrados mediante las búsquedas informáticas realizadas, lo cual corrobora los datos obtenidos en los adenocarcinomas y sugiere la posibilidad de complementar el protocolo de PCR splinkerette y la búsqueda bioinformática.

Otra región amplificada en los adenocarcinomas mamarios corresponde a la NM\_005204.2 correspondiente a la MAP3K8 (mitogen-activated protein kinase) la cual pertenece a la familia de las proteincinasas que se activan en respuesta a estímulos extracelulares (mitógenos, estrés osmótico, calor y citosinas pro-inflamatorias) y regulan de manera rigurosa diversos mecanismos celulares tales como diferenciación celular, proliferación y apoptosis.

Finalmente, el conjunto de datos obtenidos correspondientes a las búsquedas informáticas empleando las secuencias LTRs y la secuencia del gen *env* del MMTV nos han arrojado datos interesantes que corroboran los estudios publicados anteriormente y confirman la participación del MMTV en el proceso de carcinogenesis debido a la mutagenesis insercional que provoca. La mutagénesis insercional provocada por el MMTV en el genoma humano pudiera estar participando conjuntamente con los factores de riesgo conocidos, confirmando la teoría del daño múltiple y coadyuvando con el diagnóstico de una enfermedad multifactorial. Nosotros coincidimos que aún es necesario responder diversas preguntas relacionadas con la forma en que la infección por MMTV se lleva a cabo en los humanos para sugerir una etiología viral del cáncer de mama, sin embargo también aportamos evidencia clara de que secuencias con similitud cercana al 99% con respecto al gen *env* del MMTV están presente y pueden ser amplificadas mediante PCR en tumores mamarios humanos y no en tejidos no afectados de pacientes con cáncer mamario. En adición, los genes encontrados mediante PCR splinkerette en las muestras tumorales, se ha documentado anteriormente que participan en diversos procesos relacionados en la iniciación y mantenimiento de los tumores mamarios en humanos. Esta evidencia acumulada en los últimos 15 años indica esa fuerte asociación entre una infección viral por MMTV y la manifestación de cáncer de mama en humanos.

## 9. Conclusiones

- Las secuencias *env* del MMTV son posibles detectarlas en muestras de pacientes mexicanas con cáncer de mama.
- Los porcentajes de similitud entre las secuencias amplificadas de las muestras de adenocarcinomas mamarios de mujeres mexicanas y las secuencias depositadas en el NCBI son muy parecidos a los porcentajes reportados con anterioridad (98 y 99% de similitud).
- Es posible detectar regiones de similitud de secuencias entre el gen *env*, las LTRs del MMTV y el genoma humano mediante búsquedas bioinformáticas.
- Los genes o regiones detectadas como posibles sitios de inserción del MMTV en el genoma humano son genes de las familias WTN, FGT.
- Los genes encontrados participan estrechamente en inicio y el mantenimiento de la carcinogénesis mamaria así como en procesos de diferenciación y apoptosis.
- Los datos encontrados mediante PCR splinkerette corroboran los genes encontrados en las búsquedas bioinformáticas.

## 10. Referencias

1. Adams, J., and Cory, S. (2007). The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* 26, 1324–1337.
2. Amit, I., Citri, A., Shay, T., Lu, Y., Katz, M., Zhang, F., Tarcic, G., Siwak, D., Lahad, J., Jacob-Hirsch, J., et al. (2007). A module of negative feedback regulators defines growth factor signaling. *Nat. Genet.* 39, 503–512.
3. Kesteloot H, Zhang J (2006): Differences in breast cancer mortality worldwide: unsolved problems *Eur J Cancer Prev* 15: 416-423.
4. Hanahan D, Weinberg R (2011): Hallmarks of Cancer: The Next Generation. *Cell* DOI 10.1016/j.cell.2011.02.013
5. Isselbacher K, Braunwald E, Wilson J, Martín J, Fauci As, Kasper D. Harrison: (1994). *Principios de Medicina Interna*. 13ª ed. Madrid: McGRAW-Hill-Interamericana de España; P. 2089-2104
6. Osin P, Lakhani S (1999). The pathology of familial breast cancer: immunohistochemistry and molecular analysis. *Breast cancer research* 1:36:-40
7. Stratton M, Campbell P, Futreal A (2009): The Cancer genome. *Nature Reviews*, doi:10.1038/nature07943
8. Desai K, Kavanaugh C, Calvo A, Green J (2002): Chipping away at breast cancer: insights from microarray studies of human and mouse mammary cancer. *Endocr Relat Cancer* 9: 207-220.
9. Talbot S., & Crawford, D. H (2004) *Viruses and tumours—an update*. *Eur. J. Cancer* 40,1998–2005
10. Chatterjee A., Mambo, E. & Sidransky, D. (2006) Mitochondrial DNA mutations in human cancer. *Oncogene* 25, 4663–4674
11. Ruiz E. (1993) *Manual de patología mamaria*. 1ª ed. Barcelona: Editorial JIMS SA;
12. Lozano R, Knaul F, Gómez-Dantés H, Arreola-Ornelas H, Méndez O. (2008) *Tendencias en la mortalidad por cáncer de mama en México, 1979-2006*. Observatorio de la Salud. Documento de trabajo. Competitividad y Salud, Fundación Mexicana para la Salud
13. Aceves E. (2003) *Boletín trimestral del Johnson & Johnson Medical Mexico*, 3:7
14. Zamora P, Espinosa E, González M (2001). *Cáncer de mama. Clasificación y concepto*. *Medicine* 8(57): 3041-3049.
15. Marinez J (1995). *Biología molecular y cáncer*. En: Herruzo AJ, Rodríguez-Escudero FJ, Comino R, editors. *Avances en Oncología Ginecológica y mamaria*. 1ª ed. Granada. Proyecto Sur de Ediciones; 1995. P. 15-25.
16. Russo J, Russo I (2000). *Desarrollo y aspectos patogenéticos del cáncer de mama humano*. En: Perez Lopez FR, editors. *Cáncer de mama: biología, diagnóstico y tratamiento* 1ª ed. Zaragoza. SEISGE, P. 7-16
17. Zur Hausen H (1991). *Viruses in human cancers*. *Science*. 54(5035):1167-73.
18. S.M. Cohen, in: J. Parsonnet (Ed.), *Microbes and Malignancy: Infection as a Cause of Human Cancers*, Oxford University Press, Oxford, UK 1999, pp. 89–106
19. M. Minami, Y. Daimon, K. Mori, H. Takashima, T. Nakajima, Y. Itoh, T. Okanoue, *Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as*

- an early drastic genetic change leading to hepatocarcinogenesis, *Oncogene* 24 (2005) 4340–4348.
20. Martin D y Gutkind J., Human tumor-associated virus and new insights into the molecular mechanisms of cancer. *Oncogene* 2009 27, S31-S42
  21. Ganem D, Prince A. (2004). Hepatitis B virus infection–natural history and clinical consequences. *N Engl J Med* 350: 1118–1129
  22. Guidotti LG, Chisari FV. (2006). Immunobiology and pathogenesis of viral hepatitis. *Annu Rev Pathol* 1: 23–61
  23. Kao JH, Chen DS. (2002). Global control of hepatitis B virus infection. *Lancet Infect Dis* 2: 395–403.
  24. Colombo M, Kuo G, Choo QL, Donato MF, Del Ninno E, Tommasini MA et al. (1989). Prevalence of antibodies to hepatitis C virus in Italian patients with hepatocellular carcinoma. *Lancet* 2: 1006–1008
  25. Thomas DL, Astemborski J, Rai RM, Anania FA, Schaeffer M, Galai N et al. (2000). The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA* 284:450–456.
  26. Arvanitakis L, Yaseen N, Sharma S. (1995). Latent membrane protein-1 induces cyclin D2 expression, pRb hyperphosphorylation, and loss of TGF-beta 1-mediated growth inhibition in EBV-positive B cells. *J Immunol* 155: 1047–1056.
  27. Brown KD, Hostager BS, Bishop GA. (2001). Differential signaling and tumor necrosis factor receptor-associated factor (TRAF) degradation mediated by CD40 and the Epstein-Barr virus oncoprotein latent membrane protein 1 (LMP1). *J Exp Med* 193:943–954.
  28. Mosialos G, Birkenbach M, Yalamanchili R, VanArsdale T, Ware C, Kieff E. (1995). The Epstein-Barr virus transforming protein LMP1 engages signaling proteins for the tumor necrosis factor receptor family. *Cell* 80: 389–399.
  29. Mant C, Hodgson S, Hobday R, D'Arrigo C, Cason J: A viral aetiology for breast cancer: time to re-examine the postulate. *Intervirology* 2004, 47: 2-13.
  30. Stewart TH, Sage RD, Stewart AF, Cameron DW. Breast cancer incidence highest in the range of one species of house mouse, *Mus domesticus*. *Br J Cancer*. 2000;82:446–451.
  31. Choi Y, Kappler JW, Murrack P: A superantigen encoded in the open reading frame of the 3' long terminal repeat of mouse mammary tumour virus. *Nature* 1991, 350: 203-207.
  32. Brandt-Carlson C, Butel JS, Wheeler D: Phylogenetic and structural analyses of MMTV LTR ORF sequences of exogenous and endogenous origins. *Virology* 1993, 193: 171-185.
  33. Hoshino A, Yee CJ, Campbell M, Woltjer RL, Townsend RL, van der MR et al.: Effects of BRCA1 transgene expression on murine mammary gland development and mutagen-induced mammary neoplasia. *Int J Biol Sci* 2007, 3: 281-291
  34. Theodorou V, Kimm MA, Boer M, Wessels L, Theelen W, Jonkers J et al.: MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nat Genet* 2007, 39: 759-769.

35. Davies BR, Platt-Higgins AM, Schmidt G, Rudland PS: Development of hyperplasias, preneoplasias, and mammary tumors in MMTV-c-erbB-2 and MMTV-TGFalpha transgenic rats. *Am J Pathol* 1999, 155: 303-314.
36. Stewart TH, Sage RD, Stewart AF, Cameron DW: Breast cancer incidence highest in the range of one species of house mouse, *Mus domesticus*. *Br J Cancer* 2000, 82: 446-451
37. Bock M. y Stoye J.P. (2000). Endogenous retroviruses and the human germ line. *Curr. Op. Genet. Develop.* 10, 651-655.
38. Löwer R., Löwer, J. y Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* 93, 5177-5184.
39. Indik S, Gunzburg WH, Salmons B, Rouault F: Mouse mammary tumor virus infects human cells. *Cancer Res* 2005, 65: 6651-6659.
40. Etkind PR, Stewart AF, Dorai T, Purcell DJ, Wiernik PH: Clonal isolation of different strains of mouse mammary tumor virus-like DNA sequences from both the breast tumors and non-Hodgkin's lymphomas of individual patients diagnosed with both malignancies. *Clin Cancer Res* 2004, 10: 5656-5664
41. Ford CE, Faedo M, Crouch R, Lawson JS, Rawlinson WD. Progression from normal breast pathology to breast cancer is associated with increasing prevalence of mouse mammary tumor virus-like sequences in man and women. *Cancer Res* 2004;64:4755-9.
42. Etkind P, Du J, Khan A, Pillitteri J, Wiernik PH: Mouse mammary tumor virus-like ENV gene sequences in human breast tumors and in a lymphoma of a breast cancer patient. *Clin Cancer Res* 2000, 6: 1273-1278.
43. Melana SM, Nepomnaschy I, Sakalian M, Abbott A, Hasa J, Holland JF et al.: Characterization of viral particles isolated from primary cultures of human breast cancer cells. *Cancer Res* 2007, 67: 8960-8965.
44. Liu B, Wang Y, Melana SM, Pelisson I, Najfeld V, Holland JF et al.: Identification of a proviral structure in human breast cancer. *Cancer Res* 2001, 61: 1754-1759.
45. Zapata-Benavides P, Saavedra-Alonso S, Zamora-Avila D, Vargas-Rodarte C, Barrera-Rodriguez R, Salinas-Silva J et al.: Mouse Mammary Tumor Virus-Like Gene Sequences in Breast Cancer Samples of Mexican Women. *Intervirology* 2007, 50: 402-407.
46. Titus-Ernstoff L, Eagen KM, Newcomb PA, Baron JA, Stampfer M, Greenberg ER, Cole BF, Ding WM, Willet WM, Trichopoulos D: Exposure to milk in infancy and breast cancer risk. *J Natl Cancer Inst* 1998;90:921-924.
47. Litvinov SV, Golovkina TV, Kriukova IN, Vasilevskaia LN: A protein related to the main core protein of the mouse mammary cancer virus in a microparticle fraction of human milk (in Russian). *Biull Eksp Biol Med* 1987;103:338-340.
48. Wang Y, Holland JF, Bleiweiss IJ, Melana S, Liu X, Pelisson I, Cantarella A, Stellrecht K, Mani S, Pogo BG: Detection of mammary tumor virus env gene-like sequences in human breast cancer. *Cancer Res* 1995; 55: 5173-5179.
49. Wang Y, Pelisson I, Melana SM, Go V, Holland JF, Pogo BG: MMTV-like env gene sequences in human breast cancer. *Arch Virol* 2001; 146: 171-180.

50. Melana SM, Go V, Holland JF, Pogo BGT: Search for mouse mammary tumor virus-like env sequences in cancer and normal breast from the same individuals. *Clin Cancer Res* 2001; 7: 283–294.
51. Liu B, Wang Y, Melana S, Pelisson I, Najfeld V, Holland J, Pogo B: Identification of a proviral structure in human breast cancer. *Cancer Res* 2001; 61: 1754–1759.
52. Mant C, Gillett CD, Arrigo C, Cason J: Human murine mammary tumor virus-like agents are genetically distinct from endogenous retroviruses and are not detectable in breast cell lines or biopsies. *Virology* 2004; 318: 393–403.
53. Bindra A, Muradrasoli S, Kisekka R, Nordgren H, Wärnberg F, Blomberg J: Search for DNA of exogenous mouse mammary tumor virus-related virus in human breast cancer samples. *J Gen Virol* 2007; 88: 1806–1809.
54. Levine PH, Pogo BG, Klouj A, Coronel S, Woodson K, Melana SM, Murali N, Holland JF: Increasing evidence for a human breast carcinoma virus with geographic patterns. *Cancer (Phila)* 2004; 101: 721–726.
55. Ford CE, Tran D, Deng Y, Ta VT, Rawlinson WD, Lawson JS: Mouse mammary tumor virus-like gene sequences in breast tumors of Australian and Vietnamese women. *Clin Cancer Res* 2003; 9: 1118–1120.
56. Melana SM, Picconi MA, Rossi C, Mural J, Alonio LV, Teyssie A, Holland JF, Pogo BG: Detection of murine mammary tumor virus env gene-like sequences in breast cancer from Argentine patients. *Medicina (B Aires)* 2002; 62: 323–327.
57. Luo T, Wu XT, Zhang MM, Qian K: Study of mouse mammary tumor virus-like gene sequences expressing in breast tumors of Chinese women (in Chinese). *Sichuan Da Xue Xue Bao Yi Xue Ban* 2006; 37: 844–846, 851.
58. Ford CE, Faedo M, Crouch R, Lawson JS, Rawlinson WD: Progression from normal breast pathology to breast cancer is associated with increasing prevalence of mouse mammary tumor virus-like sequences in man and women. *Cancer Res* 2004; 64: 4755–4759.
59. Holland JF, Melana SM, Wang Y, Bleiweiss I, Levine P, Gombe Mbalawa C, Kalengayi M, Ndom P, Ramirez M, Cervantes G, Pogo BGT: Geographic variation in proportion of breast cancers with sequences homologous to MMTV env gene. *Proc Am Assoc Cancer Res* 1998; 39: 55.
60. Etkind P, Du J, Khan A, Pillitteri J, Wiernik P: Mouse mammary tumor virus like env gene sequences in human breast tumors and in a lymphoma of breast cancer patient. *Clin Cancer Res* 2000; 6: 1273–1278
61. Ford CE, Faedo M, Crouch R, Lawson JS, Rawlinson WD. Progression from normal breast pathology to breast cancer is associated with increasing prevalence of mouse mammary tumor virus-like sequences in man and women. *Cancer Res* 2004;64:4755-9.