

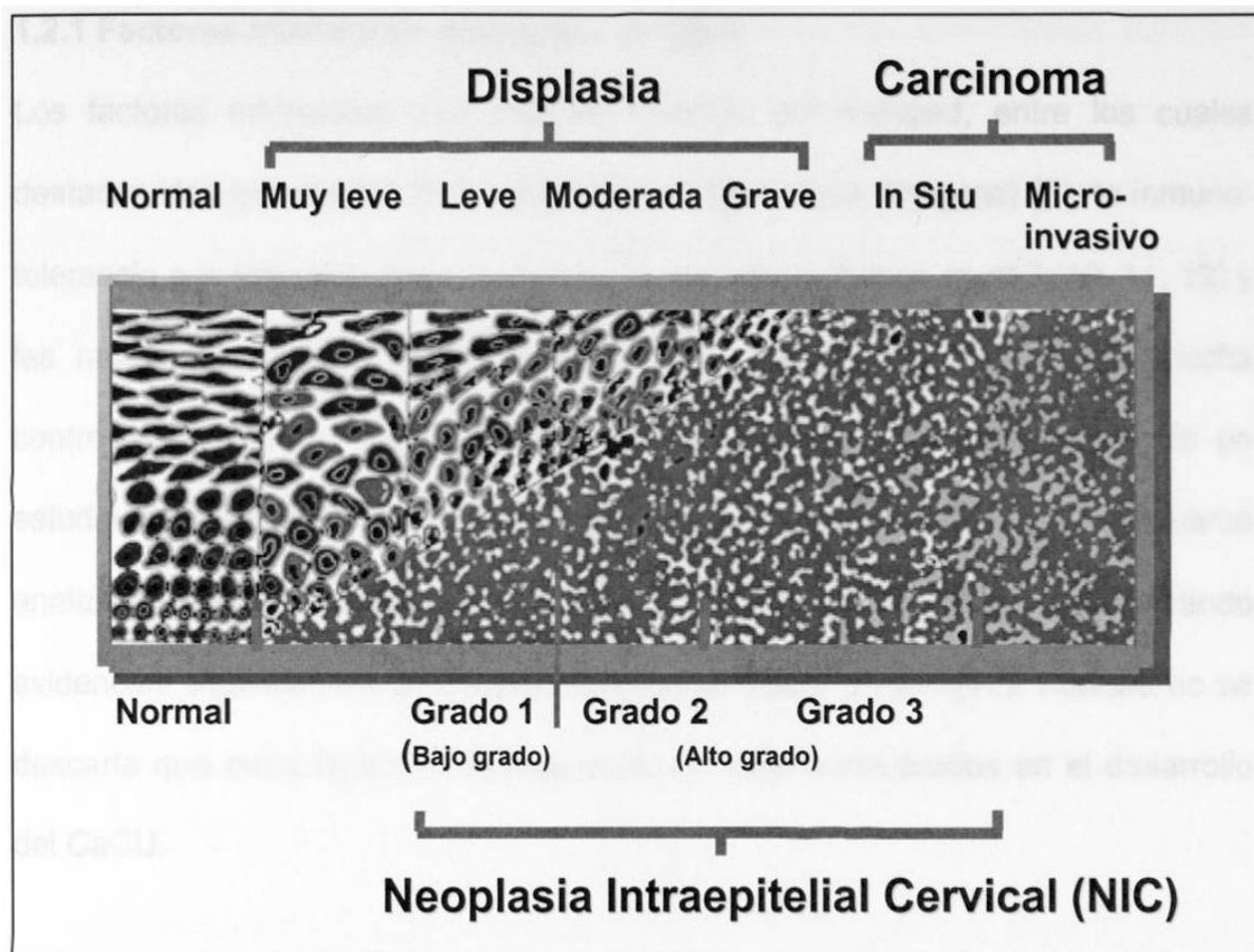
## **CAPÍTULO I**

### **I. INTRODUCCIÓN**

El cáncer cervicouterino (CaCU) es un importante problema de salud, ya que en la población mundial ocupa el segundo lugar como causa de muerte por cáncer, el cuál solo es superado por el cáncer de seno (1,2). Aproximadamente el 2 ó 3% de todas las mujeres mayores de 40 años desarrollarán CaCU, con una incidencia global de 400,000 casos nuevos diagnosticados cada año (3). En México la incidencia de este cáncer es muy alta, constituye el tercer motivo de defunción por tumores malignos en la población general y el primero en el sexo femenino (4, 5).

#### **1.1. El Cáncer de Cérvix.**

El CaCU es un proceso que se desarrolla en etapas y la evolución de la carcinogénesis se puede ilustrar con claridad (Figura 1). El cáncer invasivo es precedido por distintas lesiones intra-epiteliales histológica y citológicamente distinguibles. Las neoplasias intraepiteliales cervicales (NICs) forman un espectro de atipias del epitelio caracterizadas por alteraciones de la diferenciación de severidad creciente (5) .



**Figura 1. Evolución del cáncer cervicouterino.** Se muestra la relación entre tejido normal, displasia cervical (NIC), carcinoma y lesión intraepitelial escamosa. También en la parte de abajo se muestra el equivalente citológico de acuerdo a la clasificación del Papanicolau.

## 1.2 Factores relacionados al CaCU.

Se ha tratado de identificar a los factores epidemiológicos asociados a CaCU, los cuales se pueden clasificar en dos tipos: factores intrínsecos, los cuales están determinados por el componente genético de cada una de las mujeres y los factores extrínsecos, encabezados por los hábitos y el medio ambiente (5, 6, 7).

### **1.2.1 Factores Intrínsecos Asociados al CaCU.**

Los factores intrínsecos son aquellos propios del huésped, entre los cuales destacan los genes HLA (del inglés, Human Leukocyte Antigens) (8), la inmunotolerancia a la infección crónica viral (6, 9), los polimorfismos en p53 (10, 11, 12) y las mutaciones en los genes H-ras y K-ras (13, 14, 15, 16). Existe mucha controversia sobre si estos factores están asociados a la enfermedad. En un estudio reciente realizado en nuestro laboratorio por Martínez Garza y cols. fueron analizados estos factores en pacientes del Noreste de México, encontrando evidencias significativas de asociación entre el CaCU y HLA (17). Aunque no se descarta que otros factores internos pudieran estar involucrados en el desarrollo del CaCU.

### **1.2.2 Factores Extrínsecos Asociados al CaCU.**

Se ha demostrado en estudios epidemiológicos que las mujeres con un inicio de su actividad sexual a edad temprana, así como las que tienen múltiples compañeros sexuales, están en un riesgo mayor de padecer CaCU, aunque no se conoce con exactitud como es que estos factores modifican el riesgo, pero se cree que en edades tempranas existe un incremento de la susceptibilidad del cérvix a infecciones, debido a la metaplasia pavimentosa que está ocurriendo (6, 7, 18).

Existen otros factores de riesgo asociados al CaCU como: los patrones poblacionales (los rangos más elevados de CaCU se encuentran en

Latinoamérica) y el mayor riesgo en mujeres de clase socioeconómica baja; así como también el menor riesgo en mujeres judías de Israel, deficiencias nutricionales, principalmente de Vitamina A, C, D y  $\beta$ -carotenos, folatos, hábito de fumar (se han encontrado sustancias carcinogénicas derivadas del cigarro en biopsias de cervix), uso de anticonceptivos orales, paridad, etc (4, 5, 6, 7, 19).

También se ha determinado el papel del comportamiento sexual masculino como un factor determinante en el riesgo para CaCU: 1) Existe un incremento de CaCU y cáncer de pene en ciertas áreas geográficas; 2) Las esposas de hombres con cáncer de pene tienen de 3 a 6 veces mayor riesgo de padecer CaCU que las mujeres control; 3) Las mujeres casadas con hombres quienes sus primeras esposas presentaron CaCU, tienen un riesgo doble de padecer CaCU; 4) Mujeres quienes sus parejas tienen de 10 ó más compañeras sexuales, tienen un riesgo 5 a 7 veces mayor que las mujeres control. Todos estos factores han sido relacionados, sin embargo no se les ha demostrado asociaciones causativas con el CaCU (18, 19).

De todos estos factores extrínsecos, el único factor que se le a encontrado asociación causativa con el CaCU es la infección por virus de papiloma humano (VPH) de alto riesgo (tipos 16, 18, 33, 45, 51, etc.). Esta asociación se ha correlacionado por la presencia del ADN viral en un 80-90% de los cánceres, porque se ha encontrado integración del genoma viral al genoma celular del

huésped, se ha demostrado la presencia de oncoproteínas virales que desregulan el ciclo celular y se ha demostrado su capacidad transformante de cultivos primarios de queratinocitos humanos (19, 20, 21,22 A pesar de todas estas evidencias, el VPH por sí solo no resulta ser un elemento suficiente para generar cáncer, pues en la actualidad se sabe que del 30-60% de las infecciones iniciales con VPHs son resueltas por el huésped (6, 7, 18, 19, 20). De hecho, 1 de cada 3 casos experimentan regresión de los cambios provocados por la infección y sólo una cuarta parte de las infecciones evoluciona hacia cáncer. Del total de lesiones que progresan, 10% lo hacen a cáncer in situ y menos del 1% llegan a tumores invasivos (3, 21, 22, 23).

### 1.3 El genoma del VPH.

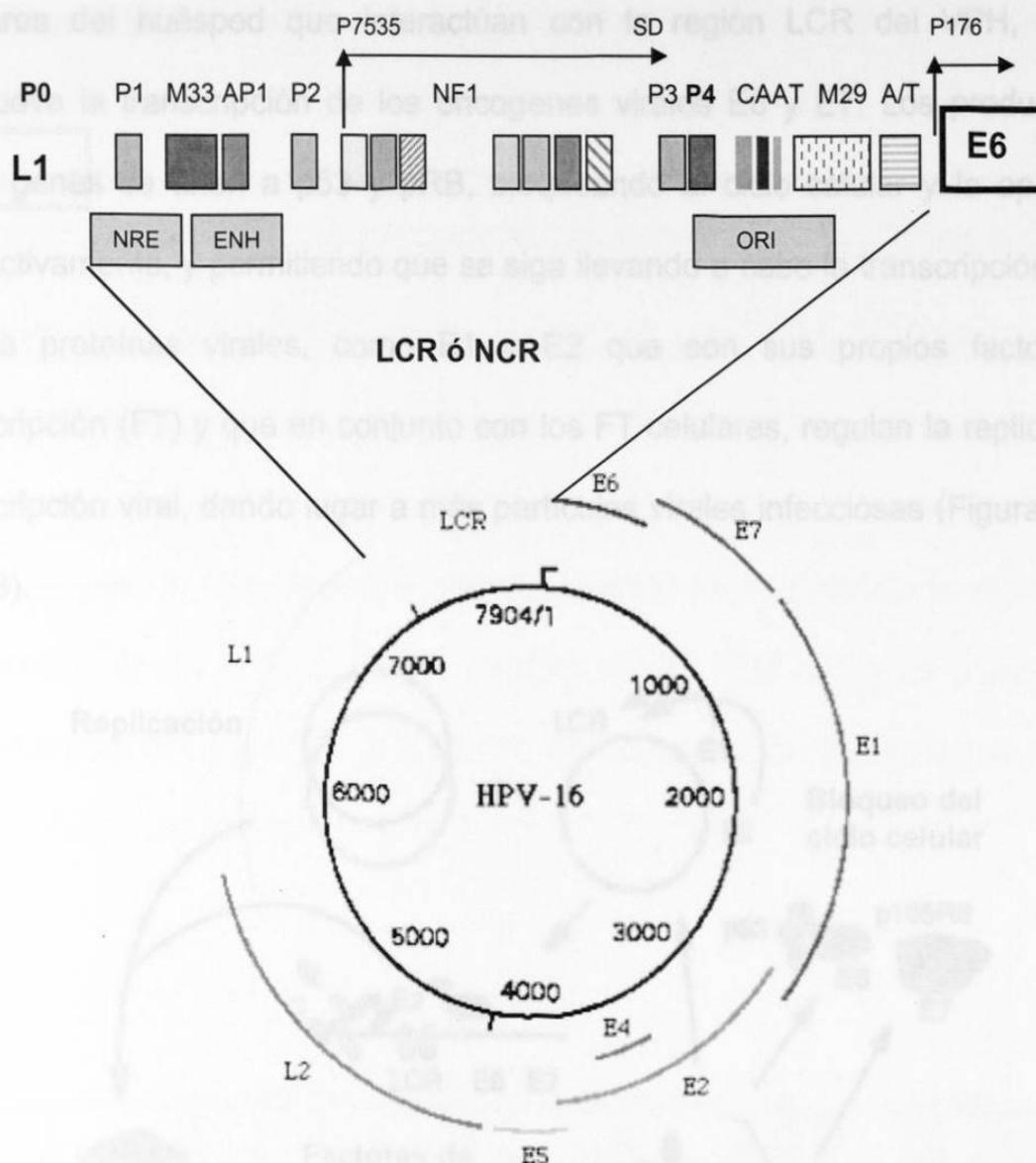
Este consiste en un ADN circular de aproximadamente 7200 a 8000 pb (8 Kb), asociado a histonas celulares que forman un compuesto parecido a la cromatina. A todas las secuencias codificantes dentro del genoma de estos virus se les denominan ORF (del inglés: *Open Reading Frame* ó marcos de lectura abiertos). Esto significa que una región ORF es un segmento de ADN que tiene la información suficiente para codificar para una proteína (24, 25).

A la fecha se han descrito más de 10 ORFs, lo que conduce a la expresión de diferentes proteínas a partir de un solo fragmento de ADN. Estas secuencias codificantes han sido clasificadas como "tempranas ó secuencias E" (E del inglés

*Early*) y "tardías ó secuencias L" (L, del inglés *Late*) para indicar el momento de la expresión durante el ciclo de vida del virus.

Todos los genomas de VPHs contienen hasta ocho genes E y la mayoría contienen los genes E2, E4, E5, E6 y E7. El gen E3 solo se encuentra en el virus del papiloma bovino (VPB). Los genes E codifican para la replicación y transformación celular y su expresión es detectable en áreas proliferativas de lesiones inducidas por VPHs.

Los genes L, específicamente los L1 y L2, codifican para proteínas estructurales y su expresión está restringida a una pequeña parte del epitelio en diferenciación, donde también ocurre la replicación del ADN viral. Los genes L1 y E6 están separados por una región que no codifica para proteínas, que se ha denominado "Región No Codificante" ó NCR (del inglés: *Non Coding Región*), ó "Región Larga de Control" ó LCR (del inglés: *Long Control Region*), ó también "Región Reguladora río Arriba" ó URR (del inglés: *Upstream Regulatory Region*). Esta región LCR contiene secuencias críticas de ADN, con diferentes actividades de promotor y enhancer para regular la transcripción y la replicación de genes tanto virales como celulares. La organización de estas secuencias codificantes se muestra en la figura 2 (24, 25).

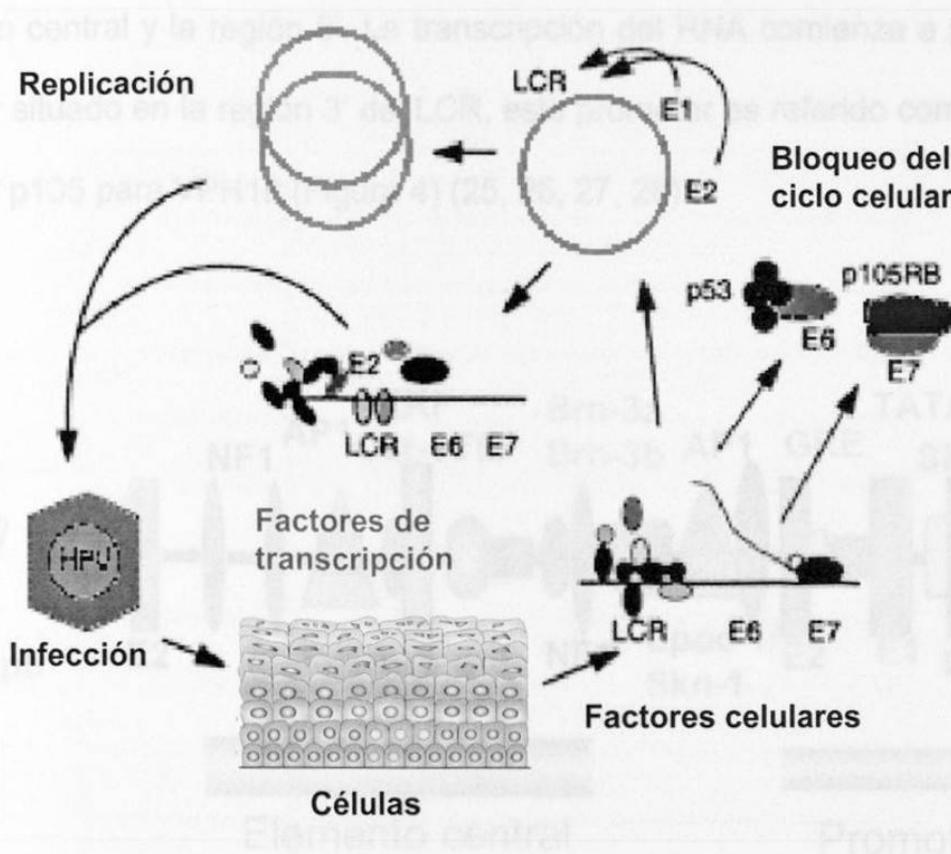


**Figura 2. Modelo del genoma del VPH.** En la parte inferior se ilustran algunas de los ORFs para los genes que conforman el genoma de este virus (E6, E7, E1, E2, E4, E5, L2 y L1), y la parte superior muestra con más detalle la organización y distribución de los elementos reguladores de la región NCR: **P7535**, **P175** son los sitios "cap" de los promotores en NCR; **SD**: sitio donador de splicing; **M33**, **M29**, **CAAT**, **A/T**: secuencias altamente conservadas entre los VPs; **P0-P4**, **AP1**, **NF1**: Sitios de unión de la proteína E2 y de factores de transcripción celular; **NRE**: elemento regulador negativo; **ENH**: elemento potenciador; secuencia mínima de origen de replicación viral.

Figura 3.- Ciclo del VPH.- Cuando el VPH infecta las células del epitelio estratificado, el primer contacto del genoma viral se establece con los FT, los cuales se unen a la región LCR para transcribir las proteínas virales E5 y E7. Estas proteínas se unen a p53 y

Cuando la infección de VPHs inicia en las células basales de epitelios estratificados, ocurre una combinación particular de factores transcripcionales

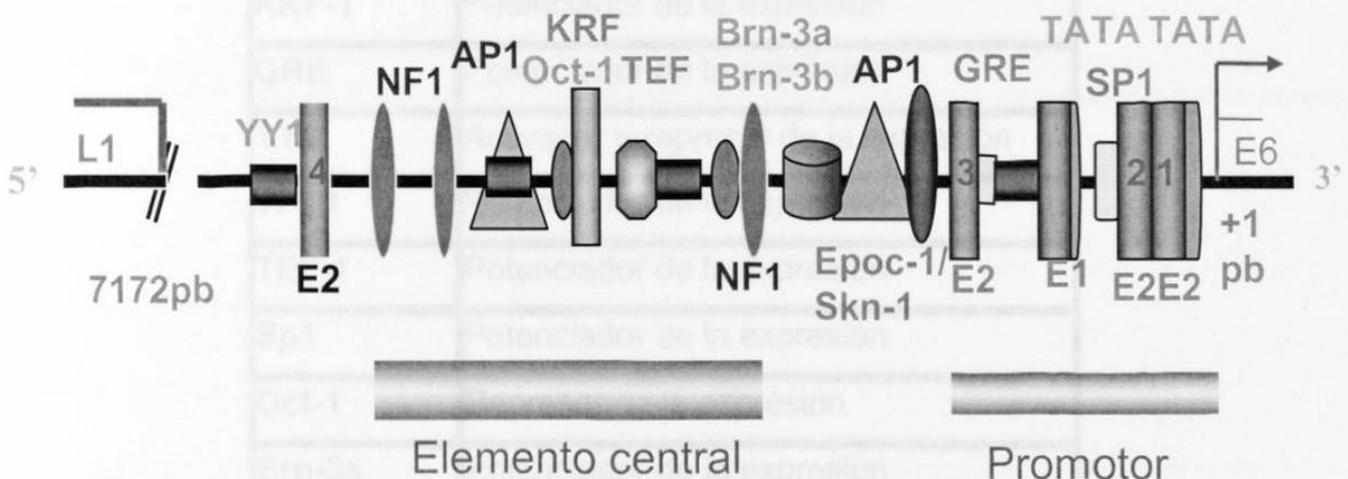
celulares del huésped que interactúan con la región LCR del VPH, lo cual promueve la transcripción de los oncogenes virales E6 y E7. Los productos de estos genes se unen a p53 y pRB, bloqueando el ciclo celular y la apoptosis, respectivamente, y permitiendo que se siga llevando a cabo la transcripción de las demás proteínas virales, como E1 y E2 que son sus propios factores de transcripción (FT) y que en conjunto con los FT celulares, regulan la replicación y transcripción viral, dando lugar a más partículas virales infecciosas (Figura 3) (21, 22, 23).



**Figura 3.- Ciclo del VPH.-** Cuando el VPH infecta las células del epitelio estratificado, el primer contacto del genoma viral se establece con los FT, los cuales se unen a la región LCR para transcribir las proteínas virales E6 y E7. Estas proteínas se unen a p53 y pRB bloqueando la apoptosis y el ciclo celular, respectivamente, permitiendo que se expresen las proteínas, E1 y E2 que son los propios FT del virus y que también se unen a la región LCR. Los FTs celulares y virales estimulan la producción de nuevas partículas virales infecciosas.

### 1.3.1 Región LCR del VPH

La región LCR abarca desde la región de la terminación del gen L1 hasta el codón para metionina del gen E6 (23, 24, 25). La región LCR es la menos conservada entre los papilomavirus. Esta contiene el promotor temprano (P97), sitios de unión para FT y el origen de replicación localizado entre dos sitios de unión de E2. La longitud de la región LCR varía entre 800 a 1000 pb para los diferentes tipos de VPH. Esta región se encuentra localizada directamente río arriba del gen E6 y río abajo del gen L1. Esta región puede ser dividida en tres partes: la región 5', el segmento central y la región 3'. La transcripción del RNA comienza a partir de un promotor situado en la región 3' del LCR, este promotor es referido como p97 para VPH16 y p105 para VPH18 (Figura 4) (25, 26, 27, 28).



**Figura 4.- Representación esquemática del LCR del VPH.** La región LCR se divide en tres partes: el promotor cercano (región 3'), el potenciador (elemento central) y la región distal (región 5'). La región LCR del VPH posee distintos elementos de acción en cis sobre los cuales se unen distintos FT básicos que se encuentran en todos los tipos celulares, además de los FT específicos de tejido, que le dan la especificidad de expresión. Así mismo tiene sitios de unión para los FT viral E2 y E1, los cuales regulan la transcripción de los genes virales.

El promotor consiste de una caja TATA y un sitio de inicio de la transcripción que es regulado por diferentes potenciadores localizados en la región 3' y central del LCR (26, 27, 28). A este promotor se le unen FT del huésped (Tabla 1) que tienen sitios de unión en la región LCR de varios VPHs de alto riesgo, los cuales se han visto que regulan la expresión de los genes virales (29, 30, 31, 32, 33, 34, 35, 36, 37). La región 5' del LCR contienen señales de terminación y poliadenilación de los transcritos virales tardíos (26,29).

**Tabla 1.- Factores de transcripción celulares que regulan la expresión de genes del VPH.**

<b>FACTOR</b>	<b>FUNCIÓN EN LA REGIÓN LCR</b>
AP1	Potenciador de la expresión
NFI	Activador o represor de la expresión
KRF-1	Potenciador de la expresión
GRE	Potenciador de la expresión
YY1	Activador o represor de la expresión
TFIID	Potenciador de la expresión
TEF-1	Potenciador de la expresión
Sp1	Potenciador de la expresión
Oct-1	Represor de la expresión
Brn-3a	Potenciador de la expresión
Brn-3b	Represor de la expresión

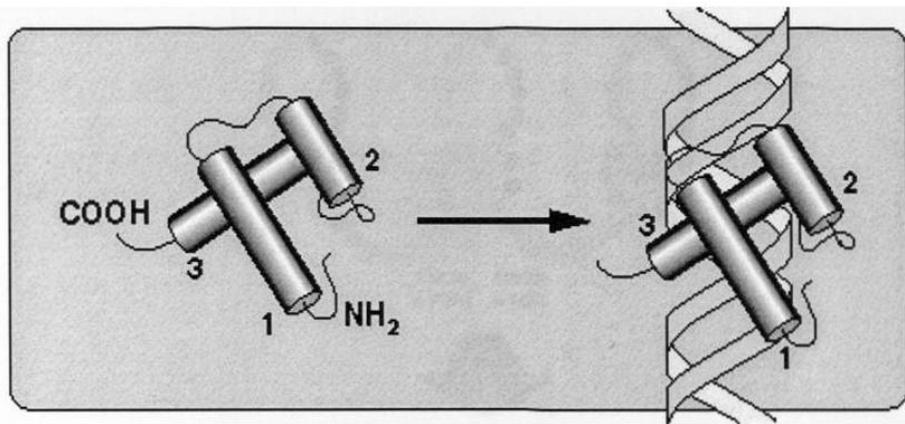
### **1.3 Factores de Transcripción.**

Las proteínas que no forman parte de la molécula de la RNA polimerasa, pero que son necesarias para la iniciación de la transcripción, se denominan FT. Estas proteínas controlan dónde, cuándo y cómo se expresan los genes. Los FT son estructuras modulares con al menos dos dominios funcionales: uno que se une a secuencias de DNA de los promotores y de los potenciadores (dominio de unión a DNA), y otra que activa la transcripción mediante interacciones proteína-proteína (dominio de activación trans). Esta región se une a la RNA polimerasa o a otros factores de transcripción. Los dominios de unión a DNA tienen distintos patrones estructurales tridimensionales, o motivos. La comparación entre las secuencias de numerosos FT sugieren que existen tipos comunes de dominios que se clasifican en: 1) receptores de esteroides, 2) dedos de zinc, 3) hélice-giro-hélice (HTH, del inglés helix-turn-helix), 4) hélice-brazo-hélice y 5) cremalleras de leucina (38, 39). A continuación se describirán los tres dominios mas comunes e importantes (39).

#### **1.3.1 Hélice-Giro-Hélice.**

Este fue el primer dominio de unión que se descubrió. Se caracteriza por su conformación geométrica, más que por una secuencia aminoacídica característica. Este motivo tiene dos hélices  $\alpha$  adyacentes separadas por un giro de varios aminoácidos, que permite que la proteína se una al DNA, pero a diferencia de otros motivos de unión a DNA, el patrón HTH no puede plegarse o funcionar solo,

sino que siempre forma parte de un dominio de unión mayor. Los residuos aminoacídicos que están fuera del motivo HTH son importantes para la regulación del reconocimiento y de la unión al DNA (Figura 5) (38, 39).



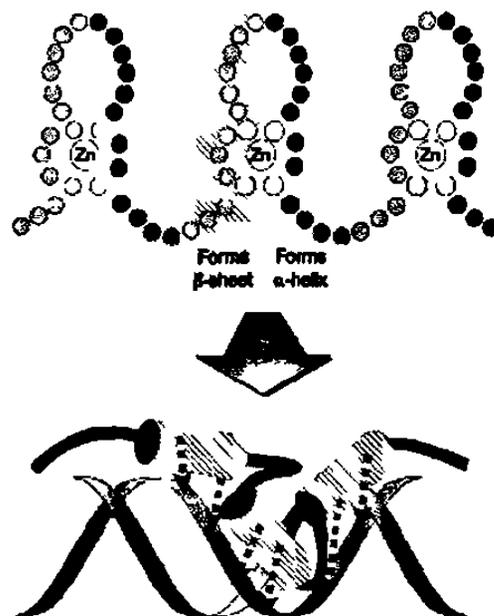
**Figura 5.- Esquema de un FT con estructura de Hélice-Giro-Hélice.-Tres hélices del homeodominio unido al surco mayor del DNA (40).**

### 1.3.2 Los Dedos de Zinc.

Descubiertos en el factor de transcripción TFIIA de *Xenopus*, estos dedos de zinc contienen grupos de cisteínas y dos residuos de histidinas en intervalos repetidos.

La repetición consenso es Cys-N<sub>2-4</sub>-Cys-N<sub>12-14</sub>-His-N<sub>3</sub>-His. Los residuos de cisteína e histidina unen átomos de zinc covalentemente, plegando los aminoácidos en lazos conocidos como dedos de zinc. Cada dedo está formado por 23 aminoácidos aproximadamente, con un lazo de 12 a 14 a.a. entre los residuos de cys e his, y con una unión entre lazos formada por 7 u 8 a.a. Los a.a. del lazo

interaccionan con secuencias específicas del DNA y se unen al surco mayor del DNA, envolviéndose alrededor de él. Dentro del surco mayor, los dedos de zinc establecen contactos con un grupo de bases del DNA especialmente con cadenas ricas en G, con las cuales forman puentes de hidrógeno (Figura 6) (38, 39).

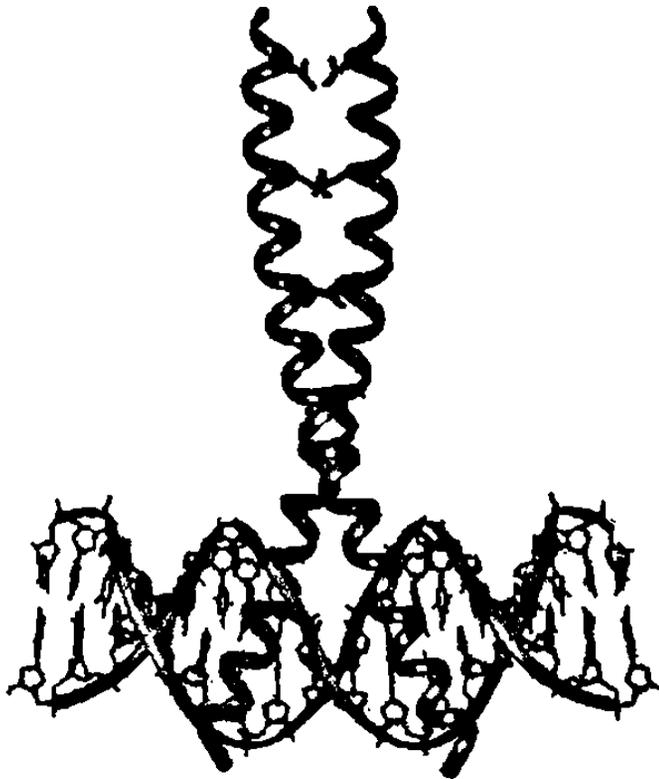


**Figura 6.- Esquema de un FT con estructura de Dedos de Zinc.-** La  $\alpha$ -hélice se inserta en el surco mayor del DNA y la  $\beta$  hélice en el surco menor (38).

### 1.3.3 Las Cremalleras de Leucina.

Observadas inicialmente como un trecho de 35 a.a. en una proteína de hígado de rata, estas tienen cuatro residuos de leucina separados por siete a.a., y está flanqueado por a.a. básicos. Las regiones ricas en leucina forman una hélice con los residuos de leucina sobresaliendo en cada vuelta. Cuando se dimerizan dos de estas moléculas, los residuos de leucina se cierran como una cremallera.

El dímero contiene dos regiones de hélice alfa adyacentes a la cremallera que se unen a los residuos fosfatos y a bases específicas del DNA (ver Figura 7) (38, 39).



**Figura 7- Esquema de un FT con estructura de Cremallera de Leucina.** Las regiones ricas en leucina forman una hélice con los residuos de leucinas sobresaliendo en cada vuelta. Cuando dimerizan dos de estas moléculas, los residuos de leucina se cierran como una cremallera (41).

#### 1.3.4 FT involucrados en CaCU

De los once FTs que se unen a la región del LCR (Tabla1), se eligieron cuatro: Brn-3a, Brn-3b, Oct1 y YY1, de acuerdo a sus antecedentes en el CaCU y la disponibilidad de sus secuencias en el GenBank.

#### **1.3.4.1 Brn-3a y Brn-3b.**

Se identificaron por primera vez en células neuronales y sus dominios de unión al DNA son del tipo POU (hélice-giro-hélice) (42). Brn-3b es antagonista de Brn-3a, debido a un cambio de una isoleucina por valina. Este cambio a.a. convierte a Brn-3b de un represor en un potenciador de la transcripción (43). Brn-3a es un factor de transcripción que activa la expresión de E6 Y E7 a partir del promotor de HPV16. Un estudio reciente de D. Ndisang y cols. demuestra que Brn-3a se encuentra sobre expresado hasta 300 veces en biopsias cervicales de mujeres con NIC III, comparadas con biopsias de cérvix normal (42).

Otros estudios han demostrado que Brn-3a es un potente activador del protooncogen Bcl-2, pero esta actividad es fuertemente inhibida por la proteína p53 (44). Por las evidencias presentadas en estos estudios se sugiere un papel crítico de Brn-3a y Brn-3b en la regulación de la expresión génica de HPV y en el control del crecimiento y transformación de las células cervicales.

#### **1.3.4.2 Oct1.**

Su dominio de unión al DNA es de tipo POU. Este es un factor de transcripción de células B, inmunoglobulinas, H2b, snRNAs y en algunas células T (45), por lo que es un factor muy importante en la respuesta inmune que monta el huésped a una infección viral. En estudios realizados con la región LCR del VPH 16 se ha

observado que mutaciones en el sitio de unión al factor Oct-1, bloquean la unión de este factor a la región LCR, incrementando significativamente la actividad enhancer/promotor de esta región y aumentando la expresión de las oncoproteínas E6 y E7 del virus (46).

#### **1.3.4.3 YY1.**

Este es un FT ubicuo. Su dominio de unión forma dedos de zinc (47) y puede actuar como un represor o un activador dependiendo del contexto del promotor. YY1 es un potente activador de la región reguladora río arriba (URR) del VPH 18, pero su capacidad de activación es dependiente de una región adyacente al sitio de unión al DNA. La delección de esta región adyacente causa que YY1 actúe como represor de la actividad del URR del VPH 18 (48). Además, mutaciones en el sitio de unión de YY1 incrementan la actividad del promotor P97 del VPH 16 de 3 a 6.5 veces (49).

Otros autores han analizado la secuencia promotora del interferón gamma, encontrando un sitio de unión para YY1 y han determinado que YY1 actúa como inhibidor de la transcripción del IFN-gamma, y que su sobre expresión podría generar una inmunosupresión del huésped, facilitando la infección y/o proliferación del VPH (50). Adicionalmente se ha visto que YY1 activa a c-Myc, el cual es un oncogen asociado a la proliferación en CaCU (51).

#### **1.4 Polimorfismos de un solo nucleótido (SNP en inglés).**

El genoma humano no es una entidad estática. Antes al contrario, el DNA está sometido a multitud de alteraciones que pueden provocar numerosas enfermedades. Estas alteraciones genéticas pueden implicar grandes reorganizaciones cromosómicas, incluyendo duplicaciones o deleciones de fragmentos y hasta de cromosomas enteros. Asimismo, son muy frecuentes modificaciones más sutiles que implican cambios en uno o en unos pocos nucleótidos. Estos cambios que sufre la estructura del DNA se denominan mutaciones.

Las mutaciones se originan por múltiples causas, incluyendo errores en los mecanismos de replicación y reparación del DNA y factores ambientales u ocupacionales. Además, las reacciones químicas espontáneas de despurinación y desaminación son una fuente extraordinariamente frecuente de alteraciones genéticas. Las consecuencias de la acumulación de mutaciones en un organismo pueden ser también diversas. Así, las mutaciones pueden tener efectos deletéreos y causar enfermedades, incluyendo el cáncer (52). Sin embargo, en ocasiones las mutaciones pueden ser neutras o incluso beneficiosas. Este tipo de mutaciones nos sirve para introducir el concepto de polimorfismo, que surge de la variación alélica entre individuos de la misma especie. Un polimorfismo es considerado como tal cuando su frecuencia en la población es superior al 1%. Hay varios tipos de polimorfismos (inserciones, deleciones, cambios en el número de secuencias

repetidas), pero los más frecuentes son los denominados SNPs (single nucleotide polymorphisms del inglés) (53).

Los SNPs son secuencias alternativas (alelos) que existen en los genomas de individuos de determinadas poblaciones (54). Por ejemplo un individuo podría tener una guanina en una posición particular de su genoma (locus), donde otro individuo podría tener una citosina exactamente en la misma localización.

Los SNPs son las formas más comunes de polimorfismos genéticos en el genoma humano, que ocurren en promedio una vez por cada 1,000 pb. (55). El genoma humano es dividido en genes y regiones intergénicas. Las primeras se dividen a su vez en regiones codificantes y no codificantes de las proteínas, conocidos como exones e intrones respectivamente. Por lo tanto un SNP que ocurre dentro de un exon podría resultar en un cambio aminoacídico entre individuos, esto es conocido como un SNP codificante.

Los a.a. pueden variar en su tamaño y carga, un cambio de a.a. puede resultar en un cambio conformacional en la proteína, afectando la interacción proteína-proteína o la unión ligando-receptor. Además un SNP podría resultar en la formación de un codón de terminación, lo cual resultaría en una proteína truncada no funcional. Los SNPs en intrones no afectarían directamente en el cambio de una proteína, pero podrían servir como marcadores de genes cercanos (56, 57).

### **1.4.1 SNPs en FT involucrados en enfermedades.**

La asociación de SNPs en FT a patologías es muy escasa. Esto se debe principalmente a que se pensaba que es casi imposible encontrar cambios en sus secuencias y mucho menos que estos cambios estén asociados a enfermedades. Con la secuenciación completa del genoma humano se han encontrado 1.42 millones de SNPs en todo el genoma (56), por lo que es muy probable encontrar SNPs en las secuencias de los FT.

En los últimos años se han hecho algunos estudios sobre SNPs en FT los cuales no solo demuestran la presencia de polimorfismos, si no también la relación de éstos a algunas patologías. Por ejemplo los factores de transcripción de tiroides 1 y 2 los cuales se han asociados a hipotiroidismo congénito y disgenesia tiroidea (58, 59, 60, 61), SNPs en el FT AP-2 se les han asociado a trastornos de personalidad (62) y por ultimo SNPs en el factor de transcripción de migración de macrófagos (MIF) se han asociado a artritis idiopática juvenil (63).

## CAPÍTULO II

### JUSTIFICACIÓN Y OBJETIVOS

#### 2.1 Justificación.

A la fecha no existen antecedentes sobre SNPs en factores de transcripción que puedan asociarse al desarrollo de neoplasias. Estos cambios podrían provocar un cambio en la afinidad entre los factores y la región LCR del VPH, potenciando la expresión de E6 y E7, produciendo una desregulación del ciclo celular y desencadenando la transformación maligna. Por otro lado, estos cambios en la afinidad podrían alterar la unión de éstos a la región LCR, pudiendo provocar que la expresión de E6 y E7 también se altere, explicando porqué algunos huéspedes infectados con VPH nunca desencadenarán CaCU.

Para esta tesis se propuso el análisis de la secuencia del dominio de unión al DNA de algunos de los FT que participan en la infección por VPH, regulando la expresión de las oncoproteínas virales. Este estudio determinaría por primera vez la presencia de polimorfismos en los dominios de los FT, los cuales podrían alterar la afinidad de éstos al sitio de unión al DNA, provocando una alteración en la expresión de los genes celulares y virales.

## **2.2 Objetivo General.**

Determinar polimorfismos en los factores de transcripción celulares involucrados en la expresión de genes del VPH, en un grupo de mujeres sin CaCU.

## **2.3 Objetivos Específicos:**

- Seleccionar los factores de transcripción celulares que serían analizados en este trabajo.
- Determinar si existe variabilidad en la secuencia de los factores de transcripción y determinar los SNPs presentes en las regiones codificantes de los FT celulares.
- Determinar si los cambios encontrados pudieran tener implicaciones en la secuencia aminoacídica del factor de transcripción.
- Determinar si estos SNPs pueden estudiarse por metodologías sencillas como PCR-RFLP.

### **CAPÍTULO III**

#### **ESTRATEGIA GENERAL**

Para cumplir los objetivos planteados, se siguió la estrategia detallada e ilustrada a continuación (Figura 8).

1.-Se recolectaron un total de 100 muestras de sangre periférica con EDTA como anticoagulante. De cada una se extrajo el RNA con el método descrito por Chomczynski y Sacchi, 1987 y se estableció un banco de muestras.

2.-Se almacenaron las muestras de RNA a  $-70^{\circ}\text{C}$  hasta el momento de su utilización.

3.-Se realizó la búsqueda de las secuencias de los FT, en el GenBank.

4.-Se diseñaron pares de oligonucleótidos para amplificar los dominios de unión de los FT Oct1, YY1, Brn-3a y Brn-3b.

5.-Se estandarizaron las reacciones de RT-PCR para los oligonucleótidos diseñados y se eligieron los FT a estudiar en este trabajo.

6.-Se amplificaron los factores de transcripción para los RNAs extraídos.

7.-Los productos amplificados (PA) se purificaron usando fosfatasa alcalina y exonucleasa ó extracción fenol-cloroformo a partir de agarosa.

8.-Los PA purificados se secuenciaron y todas las secuencias obtenidas se analizaron y compararon con la secuencia reportada en el GenBank en busca de SNPs.

9.-Se determinó el efecto de la mutación en la secuencia de a.a del FT y se determinó la naturaleza del cambio bioquímico en la secuencia de la proteína.

10-Se determino si el SNP o la contraparte normal se asociaban a sitios de reconocimiento para enzimas de restricción por el método de PCR-RFLP.



**Figura 8.- Esquema de la Estrategia General.** Se ilustra los pasos que comprenden la estrategia experimental realizada en este trabajo, los números corresponden al número del paso de la estrategia general descrita en la página anterior.

## **CAPÍTULO IV**

### **MATERIAL Y MÉTODOS**

#### **4.1 Área de trabajo, reactivos y equipo**

##### **4.1.1.-Área de trabajo.**

El trabajo se realizó en el laboratorio de Genética Molecular de la Unidad de Laboratorios de Ingeniería y Expresión Genéticas (ULIEG) del Departamento de Bioquímica, de la Facultad de Medicina de la U.A.N.L.

##### **4.1.2.-Material Biológico.**

El material biológico que se utilizó fueron 100 muestras de sangre periférica tratadas con EDTA como anticoagulante, de mujeres sin CaCU que normalmente se hacen la citología vaginal y que acuden a la consulta del Hospital Universitario. A las mujeres que aceptaron participar se les aplicó un cuestionario (Anexo 1) y se les solicitó la firma de una carta de consentimiento informado (Anexo 2). Además, se les tomó una muestra de 5 ml de sangre en un tubo con EDTA que se colocó inmediatamente en hielo y se trasladó al laboratorio del Departamento de Bioquímica de la Facultad de Medicina, UANL., donde se almacenó hasta su análisis.

#### **4.1.3.-Reactivos químicos.**

Los reactivos que se utilizaron para la extracción de RNA fueron: Acetato de sodio, EDTA, NH<sub>4</sub>Cl, NaHCO<sub>3</sub>, DEPC, citrato de sodio, sarcocil, β-mercapto-etanol y fenol, adquiridos de Sigma Chemical Company (Saint Louis, MO, EUA), isotiocianato de guanidina de Life technologies (Rockville, MD, EUA), alcohol isoamílico y cloroformo de Merck (México, D.F.).

Para realizar la retro transcripción ligada a la Reacción en Cadena de la Polimerasa (RT-PCR), los reactivos fueron obtenidos de las siguientes casas comerciales: Los oligonucleótidos, iniciadores al azar, DTT, la enzima M-MLV y los dNTPs se adquirieron de la compañía Gibco BRL (Gaithersburg, MD, EUA), la amplitaq gol con gene Amp twelve paq 10X PCR bufferII roche kit fue adquirido a Applied Biosystems (México, D.F.), el BSA fue comprado a New England Biolabs (NEB, Beverly, MA, EUA).

Para la electroforesis en geles de agarosa y poliacrilamida se utilizó agarosa, trizma base, EDTA, ácido bórico, azul de bromofenol, xilencianol, acrilamida, N,N-metilenbisaacrilamida, TEMED, γ-metacriloxi propil tri-metoxisilano, urea y bromuro de etidio todos de Sigma Chemical Company (Saint Louis, MO, EUA), mientras que el persulfato de amonio se adquirió de BioARad (CA, EUA).

Para la purificación de los PA fueron utilizadas dos enzimas: la fosfatasa alcalina (1.2U) y la exonucleasa I (6U) (PCR Product Pre-Sequencing Kit, Gibco-BRL, Life Tech, Gaithersburg MD, EUA).

En la secuenciación se usó: PCR product presequencing kit, Thermo Sequenase cycle sequencing kit de USB corporation (Cleveland, Ohio, USA), IRDye<sup>™</sup> 800 Terminator Mixes, KB<sup>plus</sup> Gel Matrix 3.7%, IR2 STOP Solution RED de Li-COR Biosciences (Lincoln, USA) y Vivid violet DNA/RNA Co-precipitante de CHIMERx (Milwaukee, Wisconsin, USA).

#### **4.1.4.- Material consumible.**

Los tubos de microcentrífuga (de 0.2, 0.5, 1.5 y 2.0 ml), las puntillas (de 0.01, 0.2, 1.0 ml) para las micropipetas de precisión, los tubos cónicos de polipropileno (de 15 y 50 ml) y los guantes de látex fueron comprados a Cell Associates (Houston, TX, EUA). Las micropipetas de precisión de volumen variable de 2, 10, 20 200 y 1000  $\mu$ l fueron obtenidas de Rainin Instruments (Woburn, MA, EUA). Los cepillos citológicos desechables utilizados fueron comprados a las Industrias DM, S.A. de C.V. (México, D.F.).

#### **4.1.5.- Equipo.**

Se utilizó una centrífuga clínica de mesa Beckman TJ6 y un vórtex modelo 37600 de Thermolyne (Dubuque, Iowa, USA), una balanza granataria Sartorius modelo 1206 MP (Cambh, Göttingen, Alemania), una centrífuga Beckman Allegra™ 21R, modelo 367570, un agitador de tubos BRNAstead Thermoline modelo Maxi Mix (Bubuque, IOWA, EUA), una placa agitadora con control de temperatura marca Corning modelo PC320 (N.Y., EUA) y una microcentrífuga Eppendorf modelo 5412 (Hamburg, Alemania).

Los termocicladores utilizados fueron: un Termociclador PTC-100 de 60 pozos de 0.5 ml, MJ-Research (Watertown, MA, EUA), un PCR-Express de 96 pozos de 0.2 ml, Hibayd (Reino Unido), un secuenciador automático DNA sequenecer Long Reader 4200 (LI-COR, California, EUA) y un termomixer 5436 (Eppendorf).

Para analizar los productos amplificados por PCR, se utilizaron cámaras de electroforesis horizontal Fotodyne (Hartland, WI, EUA), cámaras verticales de 20 x 20 cms y una fuente poder de Gibco-BRL modelo 250 (Gaithersburg, MD, EUA). Para analizar los geles de agarosa, se utilizó el equipo de fotodocumentación Gel Doc 1000 y el programa molecular Analyst de Bio Rad (Hercules, Ca, EUA).

**4.1.5.1.- Apoyo computacional.** El procesamiento de datos fue realizado en una computadora personal Toshiba modelo 2115 y un sistema de análisis dimensional de geles compatible con PC, constituido de una cámara de video, una fuente de luz UV y una computadora (Gel Doc System, BIO RAD).

El procesador de texto utilizado fue *Microsoft Word XP* (Microsoft Corporation), Procesadores gráficos *Microsoft Power Point XP* (Microsoft Corporation), Adobe Photoshop *Limited Edition 5.0* (Adobe System Incorporated) y HP Scan-Jet 5100 C (Hewlett Packard, Inc).

Los programas utilizados por vía INTERNET fueron: *Entrez* (National Center for Biotechnology Information (NCBI); BLAST Network Service (Blaster); *Gen Bank* (ICEBEG, Trieste, Italia). El programa utilizado para comunicación en la red fue el *Microsoft Internet Explorer* versión 5.0

## **4.2 Métodos**

### **4.2.1.-Características de la población de estudio.**

Se tomaron 100 muestras de sangre periférica de mujeres (en el periodo de 1 de febrero a 30 de abril del 2002) que tenían una edad entre los 30 y 60 años, una residencia de 5 años o más en el estado de Nuevo León, que tuvieran una

citología vaginal negativa para CaCU o NIC de al menos seis meses, que no hubieran tenido cáncer de cervix o displasia, que no estuvieran, ni hubieran recibido tratamiento de quimioterapia y radioterapia, y por último, que consultarán en el Hospital Universitario.

#### **4.2.2.-Extracción de RNA.**

Las muestras de sangre periférica de mujeres se procesaron utilizando el método de Chomczynski y Sacchi (64). Se colocaron 5 ml de sangre en un tubo falcon de 50ml y se le agregaron 35 ml de solución de lisis ( $\text{NH}_4\text{Cl}$  155 mM,  $\text{NaHCO}_3$  10 mM y EDTA 0.1 mM, pH 7.4), se incubó la muestra en hielo durante 15 min y se centrifugó durante 10 min a 2000 rpm. Al terminar se descartó el sobrenadante y se resuspendió la pastilla de leucocitos en 20 ml de solución de lisis, se incubó en hielo 5 min centrifugándose 10 min a 2000 rpm, se volvió a descartar el sobrenadante y se lavó nuevamente la pastilla con solución de lisis. Posteriormente la pastilla se resuspendió en 1 ml de  $\text{H}_2\text{O-DEPC}$  y se añadió 1.8 ml de solución D (Iso-tiocianato de guanidina 4M, Citrato de sodio 25mM, Sarcosil 0.5% y  $\beta$ -mercapto-etanol 0.1M) y se mezcló por inversión, se le agregaron 180  $\mu\text{l}$  de acetato de sodio 2M pH=4.0, 1.8 ml de fenol equilibrado con  $\text{H}_2\text{O-DEPC}$  a pH 7 y 360  $\mu\text{l}$  de Sevag (alcohol isoamílico-cloroformo 1:49). La suspensión formada se agitó en el vortex por 10 seg y se incubó en hielo durante 15 min, se centrifugó a 10,000 rpm, por 20 min a 4°C, la fase acuosa se transfirió a otro tubo

falcon y se le añadieron 2.5 vol de etanol absoluto, se mezcló por inversión y se colocó a  $-20^{\circ}\text{C}$  durante 1 h, se centrifugó a 10,000 rpm por 20 min a  $4^{\circ}\text{C}$ , se decantó el sobrenadante y la pastilla se resuspendió en 500  $\mu\text{l}$  de solución D, se transfirió a un tubo de 2 ml y se añadieron 2 vol de etanol absoluto, y se almacenó a  $-20^{\circ}\text{C}$  hasta su utilización, donde se tomó una alícuota de 200  $\mu\text{l}$  y se centrifugó a 14,000 rpm a  $4^{\circ}\text{C}$ , se decantó el sobrenadante y la pastilla se secó al medio ambiente y se resuspendió en 10  $\mu\text{l}$  de  $\text{H}_2\text{O-DEPC}$ , almacenándose a  $-70^{\circ}\text{C}$ .

#### **4.2.3.- Análisis electroforético del RNA.**

Para observar la calidad de los RNAs extraídos, estos se visualizaron en un gel de agarosa en condiciones desnaturizantes al 1% (250  $\mu\text{l}$  de isotiocianato de guanidina 1M y 5  $\mu\text{l}$  de bromuro de etidio 100X).

#### **4.2.4.- Diseño de los oligonucleótidos.**

El primer paso fue elegir 4 de los 11 FT que se unen a la región LCR del VPH. Esto fue hecho en base a que estuvieran sus secuencias reportadas en el genbank o algún otro banco de DNA y la relevancia de estos factores en las neoplasias. Los oligonucleótidos se diseñaron a partir de secuencias reportadas de los DNAs de los 4 FT. Las secuencias se obtuvieron del GenBank vía Internet (<http://www.ncbi.nlm.nih.gov/>).

Los oligonucleótidos se diseñaron sobre las regiones de los dominios de unión al DNA para amplificar las secuencias de los FT. Para el diseño se utilizaron los programas Oligo 4.0 y Amplify 5.1. Estos programas permitieron controlar los factores como: las temperaturas de alineamiento, los malos apareamientos, los dímeros entre los oligonucleótidos, etc. La especificidad de los oligonucleótidos se comprobó utilizando el programa Blast.

#### **4.2.5 Reacción de RT.**

Los DNAC se sintetizaron a partir de RNA total (RNAt) utilizando 1µl de hexámeros al azar más 1µl de RNA, desnaturalizando a 65°C 5 min e incubación por 2 min en hielo. Posteriormente se le agrego la mezcla de buffer, dNTPs, DTT, RNAsin, BSA, H<sub>2</sub>O y la enzima MMLV-RT. Las condiciones utilizadas al principio de la RT fueron 25°C 10 min, 42°C 60 min y 70°C 15 min, las concentraciones de la reacción se muestra en la tabla 2.

#### **4.2.6 PCR.**

Una vez que se tenían los DNAC, se realizaron las PCRs utilizando los oligonucleótidos diseñados y utilizando al principio las condiciones descritas en la tabla 3 para todos los FT.

Tabla 2- Condiciones Iniciales para la RT de los FT.

Reactivos	Volumen	Concentracion final
RNA de sangre periférica	1 $\mu$ l	50 ng/ $\mu$ l aprox.
random primers (0.5 $\mu$ g/ $\mu$ l)	1 $\mu$ l	0.05 $\mu$ g/ $\mu$ l
Buffer 5X	2 $\mu$ l	1x
dNTPs (10mM)	1 $\mu$ l	1mM
DTT (100 mM)	1 $\mu$ l	10mM
RNAsin (20U/ $\mu$ l )	0.5 $\mu$ l	1 U/ $\mu$ l
BSA (10U/ $\mu$ l )	0.2 $\mu$ l	0.2 U/ $\mu$ l
MMLV-RT (2 U/ $\mu$ l)	0.2 $\mu$ l	0.04 U/ $\mu$ l
H2O	3.2 $\mu$ l	-----
Volumen total	10 $\mu$ l	-----

El programa de amplificación utilizado al inicio para la PCR consistió un paso inicial de desnaturalización (a 95°C por 5 min, una sola vez), un paso de alineamiento (95°C por 30 seg, 55°C por 30seg y 75°C por 30 seg todo esto 30 veces más) y por ultimo un paso de extensión (a 75°C por 15 min, un sola vez). Tanto las condiciones descritas en la tabla 3 como el programa de PCR son utilizados como estándares en la ULIEG para iniciar la estandarización de las amplificaciones por PCR.

Una vez terminadas las PCR todos los PAs eran almacenados a -20 °C hasta el momento de la electroforesis para verificar la amplificación y la secuenciación de estos.

**Tabla 3.- Condiciones iniciales de la PCR para los FT**

<b>Reactivos</b>	<b>Volumen</b>	<b>Concentración</b>
Buffer 10X	2 $\mu$ l	1X
dNTPs (10mM)	0.5 $\mu$ l	0.25 mM
MgCl <sub>2</sub> 25 mM	2 $\mu$ l	2.5 mM
OligoF 5 $\mu$ M	2.5 $\mu$ l	0.62 $\mu$ M
OligoR 5 $\mu$ M	2.5 $\mu$ l	0.62 $\mu$ M
Taq gold 5U/ $\mu$ l	0.5 $\mu$ l	0.12U/ $\mu$ l
H <sub>2</sub> O	9 $\mu$ l	10.9
DNAc	1 $\mu$ l	---
Volumen total	20 $\mu$ l l	---

#### 4.2.7 Electroforesis en gel de agarosa.

La amplificación de los dominios de unión de los FT se verificó en geles de agarosa al 2%. El gel de agarosa al 2% se preparó en solución amortiguadora TBE 1X (0.089 M Tris-Borato y 0.089 M ácido bórico; pH=7.9), disolviendo en éste la agarosa con calor. La agarosa líquida se vació a un molde de electroforesis y una vez que ésta gelificó a temperatura ambiente, el gel se depositó en una cámara de electroforesis y se cubrió con buffer TBE 1X. Se mezcló 1 $\mu$ l de las muestras y 1  $\mu$ l de jugo azul 6X (0.3% azul de bromofenol, 0.3% xilencianol y 30% glicerol). Se realizó la electroforesis, inicialmente a 50 V mientras se internaban las muestras y el colorante en el gel y posteriormente a 110 V por 1 hora (dejando correr el azul de bromofenol tres cuartas partes del gel). Concluída la electroforesis, el gel se colocó en una solución de bromuro de etidio a 2  $\mu$ g/ml por

5 ó 10 min y luego se lavó varias veces con agua corriente para quitar el exceso de bromuro y así proceder a su observación en un trans iluminador de luz UV.

#### **4.2.8 Purificación de los PA.**

Después de confirmar que las PCRs habían funcionado, se tenían que purificar los PAs. Se siguieron dos métodos: El primero cuando los geles presentaban inespecificidades, se cortaban las bandas de interés y se ponían en tubos de 0.5 ml con un hueco en la parte de abajo y empaquetados con fibra de vidrio dentro de un tubo de 1.5 ml , se centrifugaban a 14,000 rpm por 5 min y a la solución obtenida en el tubo de 1.5 ml se le hacía extracción fenol-cloroformo. El segundo método se utilizaba cuando en el gel se advertía una sola banda, entonces a 10  $\mu$ l del PA se le agregaba 1.5  $\mu$ l de exonucleasa y 1.5 $\mu$ l de fosfatasa alcalina en un tubo eppendorf de 0.5 ml, se mezclaban y se ponían en un termociclador 30 min a 37°C, para que actúen las enzimas y 15 min a 80°C, para inactivarlas (Figura 9), este ultimo método fue el que se utilizó en la mayoría de las muestras.

#### **4.2.9 Secuenciación de los PA.**

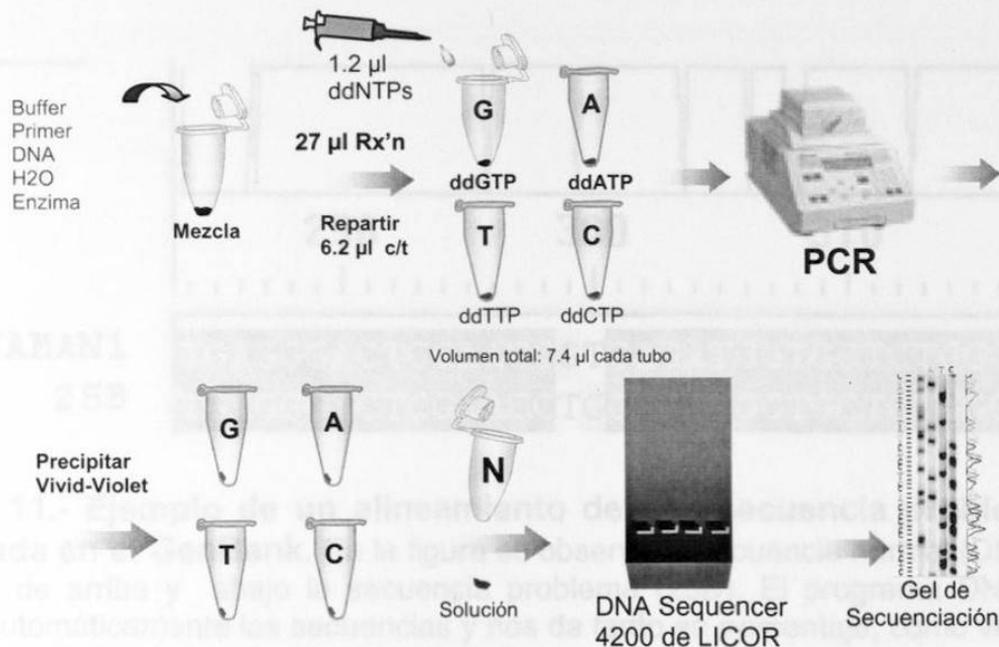
La secuenciación se llevó a cabo en un secuenciador automatizado DNA Sequencer 4200 de LICOR (Anexo 3)



**Figura 9.- Esquema para la purificación de los PA.-** El método de purificación depende de la calidad del PA, si son PAs con una sola banda se purifican utilizando exonucleasa y fosfatasa alcalina, si son PAs con muchas inespecificidades se corren en geles de agarosa, se corta la banda, se pasa la agarosa por fibra de vidrio y se realiza una extracción fenol-cloroformo.

El sistema LI-COR emplea un detector fluorescente. En la reacción de secuenciación la DNA polimerasa incorpora terminadores nucleotídicos marcados con un colorante que absorbe en el infrarrojo (IRDye) en los fragmentos de DNA sintetizados, los cuales migran a diferente longitud dependiendo del tamaño de estos en el gel de acrilamida, un láser de diodos excita el colorante infrarrojo y la señal emitida por el colorante incorporado en los fragmentos es registrada por el detector fluorescente. Los datos son enviados a un software, el cual traduce esas señales a una imagen de una serie de bandas, estas bandas representan la secuencia de los PA, las cuales debe ser corroboradas en forma manual con la secuencia blanco.

El protocolo que se utilizó en la secuenciación de las muestras se describe y observa en la Figura 10.



**Figura 10.- Esquema del protocolo de secuenciación.** En un tubo eppendorf de 0.5 µl se le agrega el buffer de reacción 10X, el oligonucleótido (5µM), el DNA (ng) y la enzima secuenasa; todo esto lleva a un volumen total de 27 µl (Anexo 3), el cual se reparte en cuatro tubos con 6 µl cada uno, después se le agrega 1.2 µl de mezcla de los terminadores nucleotidicos (ddGTP, ddATP, ddTTP, ddCTP), se mezcla y se realiza la PCR, una vez terminada la PCR se procede a precipitar los PAs utilizando vivid violet y se resuspenden en 3µl de solución Stop, desnaturalizando por 10 min (Anexo 4), para posteriormente cargar todas las muestras en un gel desnaturalizante de acrilamida al 3.7% corriendo la electroforesis por seis horas a 2000 volts, al termino de ese tiempo se observa y analiza la imagen registrada en el secuenciador, registrándose todos los datos obtenidos en un diario de condiciones de reacción (Anexo 5) y otro diario de resultados (Anexo 6).

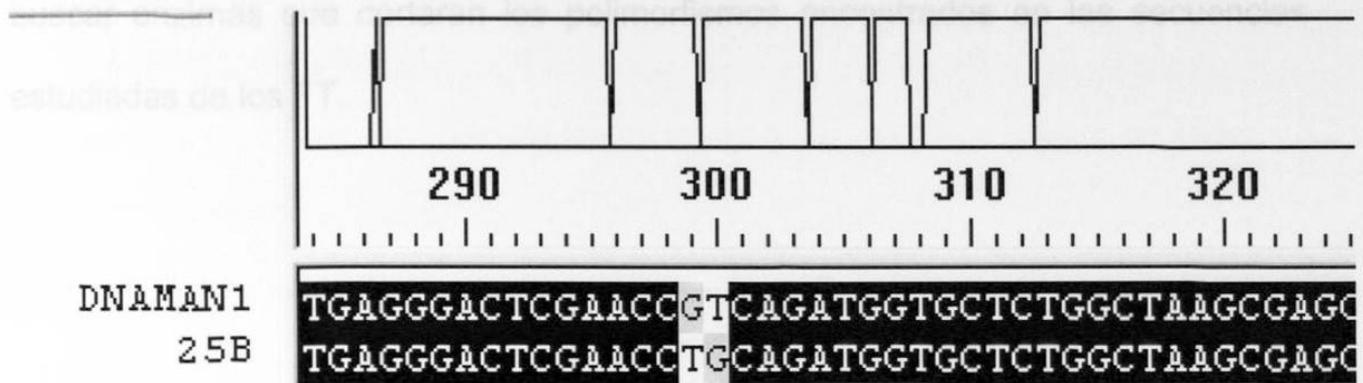
#### 4.2.10 Análisis de las secuencias nucleotídicas.

Una vez obtenidas las secuencias de los dominios de unión de los FT, se compararon mediante el programa BLAST con las secuencias reportadas en el

GenBank y se realizaron alineamientos utilizando el programa DNAMAN 5.2.2.

(Figura 11).

computacional utilizando el programa DNAMAN 5.2.2 y el DNASinder 1.3 para



**Figura 11.- Ejemplo de un alineamiento de una secuencia problema con la reportada en el GenBank.** En la figura se observa la secuencia normal (DNAMAN1) en la parte de arriba y abajo la secuencia problema (25B). El programa DNAMAN 5.2.2 alinea automáticamente las secuencias y nos da tanto en porcentaje, como visualmente la similitud y diferencias de las dos secuencias comparadas.

#### 4.2.11. Cambios producidos por SNP en las proteínas.

Al realizar los alineamientos de las secuencias problemas con la secuencia normal reportada, se observaron los cambios nucleotídicos (SNPs) de manera visual y se localizó el codón en donde se encuentra el cambio, esto se observa en el programa DNAMAN 5.2.2 como un hueco ó un sombreado de manera que si existe una diferencia en las secuencias es fácilmente detectada. Posteriormente se observó el marco de lectura del DNAC reportado en el GenBank, agrupando todos los codones y se verifica si el SNP encontrado en el codón genera un cambio de aminoácido utilizando el código genético.

#### **4.2.12. Búsqueda de RFLPs para los SNPs encontrados.**

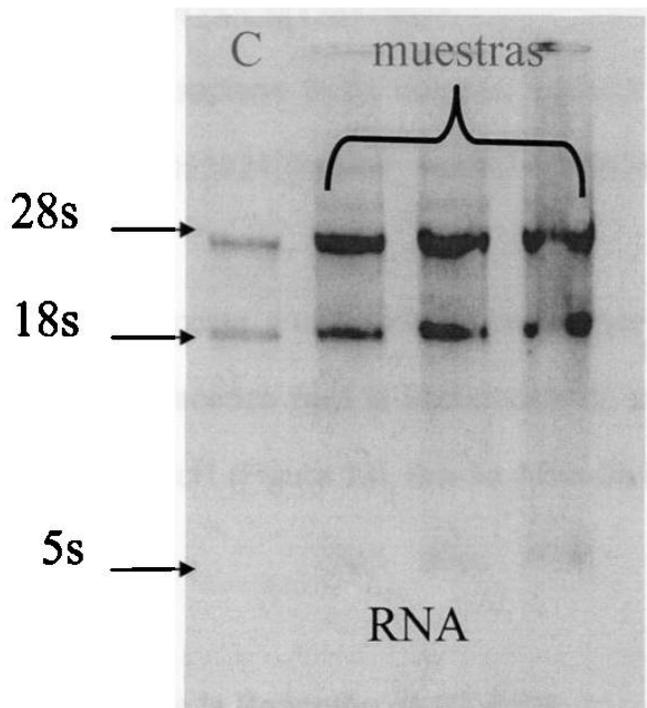
Para analizar los SNPs encontrados en las muestras se hizo un análisis computacional utilizando el programa DNAMAN 5.2.2 y el DNASTrider 1.3 para buscar enzimas que cortaran los polimorfismos encontrados en las secuencias estudiadas de los FT.

## CAPÍTULO V

### RESULTADOS

#### 5.1 Extracción de RNA

Se creó un banco de RNAs de las muestras y se verificó la calidad de estos geles de agarosa (Figura 12).



**Figura 12.-RNA en gel de Agarosa al 1%. El carril C es un RNA de sangre que se usó como marcador en donde se observa las subunidades ribosomales 28s y 18s, en los siguientes carriles son tres muestras en las cuales se observan tanto las subunidades 28s, 18s y la 5s.**

#### 5.2 Diseño de Oligonucleótidos

Las secuencias de los FT estudiados se encontraron en GenBank como sigue a continuación:

1: NM\_006237 Homo sapiens POU domain, class 4, transcription factor 1 (POU4F1), mRNA

gi|20127494|ref|NM\_006237.2|[20127494]

2:NM\_004575 Homo sapiens POU domain, class 4, transcription factor 2 (POU4F2), mRNA

gi|4758947|ref|NM\_004575.1|[4758947]

3:XM\_007245 Homo sapiens YY1 transcription factor (YY1), mRNA

gi|13647907|ref|XM\_007245.3|[13647907]

4:XM\_043760 Homo sapiens POU domain, class 2, transcription factor 1 (POU2F1), mRNA gi|20538219|ref|XM\_043760.5|[20538219]

En base a estas secuencias y utilizando los programas Oligo 4.0 y Amplify 1.2, se diseñaron oligonucleótidos para la secuencias de los dominios de unión al DNA de los FT YY1, Oct1 (Figura 13), Bm-3a y Bm-3b (Figura 14).

### **5.3 Estandarización de la Reacción de RT-PCR**

Una vez teniendo los RNAs extraídos y los oligonucleótidos, se procedió a la estandarización de la reacción de RT-PCR. Para la RT se utilizaron las condiciones descritas en la tabla 4. Al DNAc obtenido se le hizo la PCR, siendo las condiciones estandarizadas para la amplificación de YY1 y Oct1 las que se observan en la tabla 5 y para Bm-3b y Bm-3a de la tabla 6.