



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**FACULTAD DE MEDICINA
DIVISIÓN DE ESTUDIOS DE POSGRADO
POSGRADO EN CIENCIAS BIOMÉDICAS
INSTITUTO MEXICANO DEL SEGURO SOCIAL
CENTRO MEDICO NACIONAL SIGLO XXI**

**“ANÁLISIS DE LA EXPRESIÓN DIFERENCIAL EN EL
CARCINOMA DEL CUELLO UTERINO”**

TESIS

**PARA OBTENER EL GRADO DE:
DOCTOR EN CIENCIAS BIOMÉDICAS**

**PRESENTA:
CARLOS GUADALUPE PÉREZ PLASENCIA**

**TUTOR:
DR. MAURICIO SALCEDO VARGAS**

MÉXICO, D.F.

FEBRERO 2007



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

"Lo que he hecho es mostrar que es posible que la forma en que comenzó el universo esté determinada por las leyes de la ciencia. En ese caso, no sería necesario apelar a Dios para decidir cómo comenzó el universo. Esto no prueba que no exista dios, sólo que Dios no es necesario".

Stephen W. Hawking.

"¡Oh insensato hombre, que no puede crear un gusano y sin embargo crea Dioses por docenas!" Michel de Montaigne.

INDICE

1.- Introducción	4
2.- Planteamiento del problema	15
3. Hipótesis	16
4. Objetivos	17
5. Material y métodos	18
6. Resultados	35
7. Discusión	56
8. Conclusiones	61
9. Literatura citada	62
10. Anexos	70

1. INTRODUCCIÓN

I. Cáncer Cérvico Uterino

i) Epidemiología de la enfermedad.

El cáncer cérvico-uterino (CaCU) es un problema de salud pública en los países en vías de desarrollo y constituye una de las principales causas de muerte en las mujeres de 35 a 60 años de edad. En México, el CaCU es la primera causa de muerte por tumores malignos. En 1999 se reportaron 4,590 defunciones con una tasa de 19.9 por 100 000 mujeres, con esto se estima que en ese año, murieron 12.5 mujeres por día (1,2). Actualmente nuestro país ocupa uno de los primeros lugares en morbilidad de esta neoplasia a nivel mundial; así, la incidencia de Carcinoma invasor e *In situ* son de 41.1 y 49.3 por cada 100,000 mujeres, respectivamente (3).

ii) Virus de Papiloma Humano (VPH) como factor etiológico del Cáncer Cérvico-Uterino.

El factor etiológico más importante para el desarrollo del CaCU es la infección con virus de papiloma humano de alto riesgo (VPH) (4-7). Los VPH se han clasificado en función del tipo de tejido que infectan (cutáneos y mucosas) y en asociación con el grado de la lesión (de alto, medio y bajo grado) (8). Un tipo viral se define en base a la diferencia de al menos 10% en la secuencia nucleotídica, cuando la diferencia es del 2-10%, se trata de un subtipo (9). Actualmente, se han identificado más de 100 tipos virales, los cuales han sido clasificados en 16 grupos (10). En las lesiones de bajo grado (Condiloma y NIC I) predominan los tipos 6 y 11 de bajo riesgo en aproximadamente 20-22% de los casos. En las displasias de alto grado (NIC II-NIC III) predomina el tipo 16 entre 30-77%; 37.4-41% de estas lesiones están también relacionadas con el tipo 18 y algunos del grupo de los treinta (31, 33, 35, 39, etc.). Para los carcinomas infiltrantes se ha realizado una

actualización de la clasificación epidemiológica de los tipos virales en estas neoplasias. Así, en el grupo de alto riesgo se encuentran los tipos 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 y 82; se considera igualmente a los tipos 26, 53 y 66 como de probable capacidad carcinogénica. En el grupo de bajo riesgo están los tipos 6, 11, 40, 42, 43, 44, 54, 61, 72, 81. Los tipos 26, 53, 66, 73 y 82 han sido reclasificados de un grupo anteriormente como de riesgo indeterminado a ser incluidos en los de alto riesgo. Los tipos 16, 18, 31, 33, 35, 45, 52 y 58 se encuentran en más del 95% de los carcinomas escamosos (11).

Diferentes reportes indican que la prevalencia de infección con VPH de alto riesgo en CaCU es superior al 99.7% de los casos y que la probabilidad de desarrollar una lesión intraepitelial cervical de alto grado, se incrementa al menos 100 veces en mujeres infectadas con estos tipos virales, comparadas con mujeres no infectadas (revisado en 12).

Para establecer una infección productiva, el VPH debe tener acceso a las células que tienen la capacidad de proliferar es decir, las del estrato basal del epitelio cervical (revisado en 12). Una vez que el VPH ha infectado a las células basales, el genoma viral se replica activamente para lo cual se mantiene en forma circular libre o episomal distribuyéndose a cada célula hija durante la división celular (13).

Los VPH pertenecen a la familia *Papillomaviridae*; su genoma es una molécula de ADN bicatenaria, circular que mide alrededor de ocho kilo bases. El virión consiste de una cápside de 72 unidades o capsómeros que están compuestos de dos proteínas estructurales: L1 la cual constituye el 80% de la partícula viral y L2. El genoma viral se divide en tres regiones: la temprana (E), tardía (L) y una región regulatoria no codificante

(LCR)(14). La región regulatoria presenta diferentes sitios de unión tanto para represores como activadores de la transcripción (15). En la región temprana se encuentran genes involucrados en la promoción de la carcinogénesis (E1-E7). Estos genes se expresan en las células de los estratos basal e intermedio del epitelio escamo-estratificado (16).

E1 y E2 codifican proteínas que son vitales para la replicación extra cromosómica y la terminación del ciclo de vida viral. E1 juega un papel importante en el mantenimiento del genoma viral en forma episomal (17). Por otra parte E2 está involucrado en la regulación negativa y la transcripción de los genes de la región temprana E6 y E7 (18). Así, en los casos de CaCU frecuentemente se ha perdido la expresión de E2 debido a la integración y ruptura en la región E1/E2 del genoma viral (18).

E4 se expresa en las etapas posteriores de la infección, cuando se ensamblan los viriones maduros. No se conoce si tiene propiedades de transformación celular; sin embargo, juega un papel importante en la maduración y replicación del virus (19); además se ha visto que E4 induce el colapso de la red de citoqueratina citoplásmica en queratinocitos humanos, mecanismo que podría permitir la liberación activa de viriones en las células infectadas (20).

E5 puede interactuar con diferentes proteínas celulares transmembranales como los receptores del factor de crecimiento epidérmico (EGF), factor de crecimiento β derivado de plaquetas (PDGF- β) y con el factor estimulador de colonias 1 (21). Se ha visto que en CaCU, E5 se deleta frecuentemente, sugiriendo que este gen no tiene un papel importante en el mantenimiento del fenotipo maligno (revisado en 22).

En cuanto a las oncoproteínas virales E6 y E7, una gran cantidad de artículos evidencian su papel en la promoción y mantenimiento del fenotipo tumoral. Ambas oncoproteínas se expresan en tumores malignos de las regiones ano-genitales (23); en este contexto, se ha observado que tienen capacidad para inmortalizar y transformar diferentes tipos celulares (revisado en 24). Además de las interacciones inicialmente caracterizadas entre E6 con p53 mediada por E6-AP y E7 con pRb que promueve la inactivación de ambas proteínas celulares, se ha descrito una gran cantidad de interacciones con proteínas celulares. Sin embargo, los diferentes procesos inducidos por ambas oncoproteínas virales no son suficientes para convertir a una célula normal en cancerosa (revisado en 22).

Finalmente, la región tardía tiene dos genes (L1 y L2) que codifican para proteínas que conforman la cápside viral. Estos genes se expresan en los queratinocitos diferenciados de las capas superficiales del epitelio. Evidentemente, esto responde a una estrategia que permite al VPH empaquetar su material genético una vez que este ha sido replicado intensamente para finalmente liberar viriones al tracto genital (25).

Uno de los eventos importantes que contribuyen al desarrollo de la neoplasia cervical es la integración viral al genoma celular, con la subsiguiente pérdida en la expresión del represor E2, pues durante este fenómeno frecuentemente ocurre una ruptura en la región E1/E2. Este mecanismo implica que al no tener represor activo, las regiones de E6 y E7 presentan una actividad transcripcional incrementada (26). Por otra parte, la integración viral puede ocurrir en regiones cromosómicas en las cuales se encuentran genes con participación en diferentes tumores; por ejemplo, MYC, TP63, NR4A2, hTERT, etc (revisado en 27).

Existen suficientes evidencias para estimar que los virus de papiloma humano de alto

riesgo tienen la capacidad de inducir cambios celulares en el epitelio cervical normal que conllevan al fenotipo tumoral. Sin embargo, no todas las pacientes infectadas con dichos tipos virales desarrollan carcinoma cervical, pues es frecuente la eliminación espontánea de las infecciones virales. En un estudio realizado por Nubia Muñoz y cols. en Colombia se muestra que la persistencia de la infección por VPH disminuye hasta 23%, con una tasa de persistencia ligeramente mayor en los tipos de alto riesgo (28); esto indica que la presencia de VPHAr es un factor de riesgo necesario pero no suficiente para el desarrollo de la neoplasia cervical, y que es probable la existencia de otros mecanismos moleculares, celulares y ambientales que, en conjunto con el virus, propicien el desarrollo del carcinoma cervical. Para conocer los mecanismos moleculares alterados por el proceso de infección de los virus de papiloma en función de transcritos involucrados, se han empleado en diferentes estudios herramientas de análisis global (microarreglos de expresión, Differential Display, etc). Estas herramientas han generado datos relevantes sobre el proceso de tumorigénesis cervical.

II.- Métodos de análisis del Transcriptoma Celular.

Existen diferentes estrategias diseñadas para analizar la expresión diferencial de genes en un organismo o célula e identificar los genes involucrados en un proceso de desarrollo, en respuestas celulares a estímulos físicos o químicos, o genes involucrados en procesos oncogénicos en un tipo de neoplasia determinado (29). Entre los métodos más empleados para clonar genes expresados diferencialmente, se encuentran el análisis representativo de diferencias (RDA), despliegue diferencial (DD), reacción en cadena de la polimerasa iniciada arbitrariamente (RAP-PCR), e hibridación substractiva (SH). Básicamente, todos estos métodos están diseñados para amplificar y aislar secuencias de ácidos nucleicos presentes en una muestra y ausentes o expresados en un nivel diferente en otra muestra empleada como comparativo. Estas técnicas permiten la

detección de cambios en la expresión de RNAs mensajero por medio de un enriquecimiento selectivo sin ningún conocimiento previo de la secuencia de genes específicos (30). Sin embargo, no pueden detectar transcritos expresados a bajos niveles, además los pasos sucesivos de substracción y amplificación hacen posible la aparición de falsos positivos (revisado en 29)

El análisis en serie de la expresión génica (SAGE) es una compleja metodología que involucra una serie de herramientas de biología molecular y bioinformática, la cual permite el análisis amplio y cuantitativo del perfil de expresión celular. Con esta tecnología de análisis global, los transcritos son convertidos a ADN complementario y después digeridos en fragmentos pequeños o etiquetas (10-14 bp) por enzimas de restricción específicas (31- 33). SAGE está basado en dos principios: representación de los transcritos expresados por secuencias cortas de cDNA o etiquetas y la concatenación de estas etiquetas para su clonación, permitiendo su análisis mediante secuenciación. Se ha calculado que la cantidad de genes presentes en el genoma humano es aproximadamente 30,000-40,000. De esta manera, una secuencia de 10 pares de bases podría identificar a más de 1×10^6 transcritos diferentes (4^{10} , donde cuatro son las diferentes bases que constituyen el código genético); así, el tamaño de estas etiquetas es suficiente para discriminar todos los transcritos presentes en una célula en un momento dado (34).

SAGE puede detectar y cuantificar simultáneamente los niveles de expresión génica en una célula o tejido, tiene la sensibilidad de detectar genes expresados en bajos niveles, además de identificar genes no descritos previamente. En suma, con esta metodología se puede generar un catálogo de genes expresados y los niveles de expresión de estos; es decir, el transcriptoma celular (revisado en 35). Esta valiosa herramienta se ha empleado

en la comprensión del proceso que lleva una célula normal a una célula neoplásica, principalmente en tres aspectos de la biología del cáncer:

- El análisis de diferencias entre los patrones de expresión en distintos tipos de neoplasias y su contraparte normal.

- Identificación de genes aún no descritos que pueden jugar un papel en el desarrollo de alguna neoplasia

- Identificación de genes regulados por oncogenes y genes supresores de tumor.

Este tipo de estudios no sólo nos permite comprender las vías de regulación en los procesos involucrados en la tumorigénesis, sino también identificar nuevos marcadores de pronóstico y diagnóstico. En la mayoría de los casos analizados con esta herramienta, se ha empleado una o dos bibliotecas de tejido normal y unas pocas de tejido neoplásico o líneas celulares; por esto, los genes identificados con SAGE han sido analizados en un mayor número de muestras; encontrando concordancia en los resultados, validando de esta manera la utilidad del gen identificado como probable marcador de pronóstico y/o diagnóstico (revisado en 36). Así, en adenocarcinoma prostático se identificó el antígeno de células tallo de próstata (PSCA), el cual fue expresado en cuatro de seis bibliotecas SAGE de cáncer pancreático, pero no en bibliotecas derivadas de células de los ductos de la próstata normal. En ese mismo estudio se confirmó la sobre expresión de PSCA en 14 de 19 líneas celulares derivadas de esta neoplasia, empleando RT-PCR (32). Por otra parte, en otro estudio realizado en cáncer pancreático, se observó que un inhibidor de metaloproteinasas del tipo I (TIMP-1) podría ser utilizado como marcador de diagnóstico, al medir los niveles en suero se encontraron resultados positivos en sólo 30% de pacientes con esta neoplasia. Sin embargo, en combinación con otros dos marcadores (CA19-9 y antígeno carcinoembrionario), 60% de los pacientes en el estudio fueron diagnosticados eficientemente con alta especificidad y sin falsos positivos (37, 38). Otro

ejemplo del nivel de resolución que se desprende del análisis de librerías SAGE fue la identificación de un nuevo miembro perteneciente a la familia de los antígenos de melanoma (MAGE), denominado MAGE-E1 con tres variantes producidas por splicing alternativo (MAGE-E1a, MAGE-E1b y MAGE-E1c). MAGE-E1a y MAGE-1b fueron expresados específicamente en células de glioma (39). Estos resultados establecen la validez del análisis de bibliotecas SAGE para identificar nuevos marcadores tumorales.

Una de las ventajas que hacen de SAGE una poderosa herramienta de análisis global es la posibilidad de identificar para su posterior caracterización genes aún no descritos; esto se debe a que el análisis se desarrolla utilizando como materia prima la totalidad de los mensajeros expresados por la célula o tejido de interés; es decir, SAGE emplea un sistema de análisis abierto, a diferencia de otras herramientas de análisis global como los microarreglos. De esta manera, en cáncer de mama se observó una etiqueta (cuyo gen luego de ser caracterizado y clonado se le dio el nombre de HIN-1) que no relacionaba con ningún gen o EST (fragmento de secuencia expresado) caracterizado, y que se presentó con un alto número de copias en las librerías de células epiteliales normales; a diferencia de las bibliotecas desarrolladas a partir de carcinoma ductal *in situ* o sus metástasis, en donde no se observó dicha etiqueta (40). En un estudio posterior se emplearon muestras de carcinomas mamarios provenientes de pacientes, en los cuales se observó que HIN-1, se expresó en solo 21% de los tejidos neoplásicos estudiados; en contraste con los tejidos normales, donde la expresión fue en la mayoría. Al clonar y caracterizar el ADN complementario de HIN-1 se observó que es una proteína secretada que regula negativamente el crecimiento de líneas celulares derivadas de cáncer de mama y que se inhibe la expresión por metilación en tumores (41). La identificación de diferencias en tejidos neoplásicos y su contraparte normal a nivel molecular, abre importantes expectativas en cuanto a los mecanismos que propician el desarrollo de un

Cáncer. Identificar estos actores específicos a cada enfermedad no sólo nos permitirá tener las herramientas necesarias para pronosticar y diagnosticar eficientemente, sino que por otra parte podremos comprender mejor el fenómeno analizado a una mayor resolución que hace algunos años.

En cuanto a la identificación de genes regulados por genes que modulan el perfil de expresión como p53, el grupo de Bert Vogelstein induciendo la función de p53 en una línea celular con p53 deletado, encontró que 34 etiquetas difieren en su tasa de expresión, de los cuales sólo tres han sido previamente identificados como genes regulados por p53 (42). Con este tipo de trabajos, es posible conocer la complejidad de genes involucrados en determinada ruta metabólica; por otra parte, es claro que los genes que habitualmente se encuentran mutados en cáncer (por ejemplo p53), pueden regular a nivel transcripcional a un gran número de genes, modificando el patrón de expresión global durante la progresión tumoral.

SAGE ha sido propuesto como una herramienta analítica estratégica en el Proyecto de la Anatomía del Cáncer (CGAP). Se ha creado una base de datos con etiquetas producidos por esta herramienta. A la fecha hay más de tres millones de etiquetas de 88 diferentes librerías derivados de 19 tejidos, tanto normales como tumorales (43). El análisis de estos datos ha arrojado algunos patrones interesantes; en primer lugar, células cancerosas y sus contrapartes normales presentan un patrón de expresión celular muy similar; de hecho, alrededor de 40 genes se expresan en niveles elevados en todos los tejidos cancerosos, no así en los tejidos normales; esto podría sugerir que estos genes estén involucrados en el proceso de tumorigénesis y los señala como marcadores moleculares, o como blancos para estrategias terapéuticas. Por otra parte, se observó de manera obvia que en los tejidos neoplásicos usualmente hay un aumento en la expresión de

genes asociados con la proliferación y supervivencia y una disminución en genes involucrados en diferenciación. Además, se encontró que cerca de 1, 000 transcritos se expresaban en todos los tejidos estudiados, los cuales representan el “transcriptoma mínimo” es decir, aquellos genes cuya expresión es necesaria para mantener la maquinaria celular funcional (37,44).

En el caso del tejido cervical, no existen estudios empleando SAGE; sin embargo, se han realizado diferentes trabajos utilizando la metodología de microarreglos ó microhileras de expresión. Así, el primer estudio realizado sobre Cáncer Cérvico-Uterino reveló la sobreexpresión de los genes de proteínas de unión a guaninas, Factor nuclear NF45 y beta catenina, este trabajo permitió establecer a p16^{INK4} como probable marcador en CaCU; los hallazgos de genes diferencialmente expresados fueron validados mediante la técnica de Northern Blot (45).

Posteriormente, otro grupo reportó las diferencias en expresión génica en queratinocitos infectados con E6 y E7 de VPH16. Por ejemplo, se describió que la oncoproteína viral E6 altera la expresión de genes que afectan la resistencia de la célula hospedera a la infección y la función inmune de la misma (46). Por otro lado, con la misma tecnología, se pudo relacionar directamente a los genes de la NADH deshidrogenasa 4 y de la proteína ribosómica S12 con la progresión y presencia del cáncer cervical (47). En otro trabajo se mostró la expresión diferencial del gen IGFBP-3 causada por la transducción de las proteínas E6/E7 de papilomas de alto y bajo riesgo, que es positiva en lesiones de alto grado y en cáncer invasor; mas no en cérvix normal (48)

De igual forma, se han estudiado los cambios en la expresión global debidos a la integración viral en el genoma del hospedero, demostrando que la integración de VPH 16

en queratinocitos aumenta la expresión de genes de respuesta a IFN (49). En lo que respecta a la búsqueda de genes asociados a radioresistencia, empleando microarreglos de expresión existen a la fecha al menos tres trabajos independientes que por medio de diferentes plataformas de microarreglos empleadas, número y tipo de pacientes y análisis estadístico analizan el perfil de expresión e identifican un grupo de genes capaces de agrupar pacientes que desarrollarán resistencia a la radioterapia (50-52).

2. PLANTEAMIENTO DEL PROBLEMA

En general, los tumores sufren una continua acumulación de cambios genéticos y epigenéticos que les permite escapar de los controles celulares y ambientales. Existe un gran número de trabajos que describen la participación de diferentes genes en la regulación de estos procesos. Sin embargo, técnicas de análisis genómicas sugieren que el número de estos genes puede ser mayor.

Hasta este momento no se ha descrito el transcriptoma completo del carcinoma del cuello uterino, una neoplasia en la que la participación viral es fundamental como factor etiológico. La comparación entre el transcriptoma del carcinoma invasor del cuello del útero con librerías SAGE publicadas para otras neoplasias, permitirá distinguir genes específicos para esta neoplasia involucrados en el proceso oncogénico.

La caracterización de perfiles globales de expresión puede ayudar a elucidar procesos biológicos importantes tanto en células tumorales como normales. Aunque estos estudios sean principalmente descriptivos, pueden ayudar a revelar las bases moleculares de los fenotipos ya caracterizados.

3. HIPÓTESIS:

La comparación del patrón de expresión entre epitelio cervical normal con lesiones invasoras del cuello uterino, permitirá distinguir genes celulares involucrados en el proceso oncogénico.

4. OBJETIVO GENERAL:

Caracterizar el perfil de expresión global del Epitelio del Cuello Uterino.

OBJETIVOS ESPECÍFICOS:

- I- Caracterizar los transcritos presentes en epitelio cervical normal y en lesiones invasoras del cáncer del cuello uterino.
- II.- Comparar los patrones de expresión génica entre los tejidos estudiados.
- III.- Caracterizar las vías metabólicas y de señalización alteradas en el carcinoma cervical del cuello uterino.

5. MATERIAL Y METODOS

I. Tejidos y pacientes.

Se colectaron veinte tejidos de cérvix normal provenientes de mujeres que fueron sujetas a histerectomía por miomatosis uterina, con citología negativa confirmada por análisis histopatológico. Todos los pacientes se encontraban en etapa reproductiva y ninguno de ellos recibió terapia o anticonceptivos hormonales. Así mismo, se colectó 20 tejidos tumorales en etapa clínica IIB, con diagnóstico histopatológico confirmado. Todos los procedimientos fueron aprobados por el comité local de ética e investigación del Hospital de Oncología CMN SXXI-IMSS.

Todos los tejidos obtenidos se dividieron en tres secciones, la parte central se congeló en nitrógeno líquido inmediatamente después de la toma de la muestra y los extremos fueron fijados en etanol 70% durante toda la noche y embebidos en parafina. De estos bloques se cortaron secciones de 5 μm y se realizó la tinción de Hematoxilina/Eosina para confirmar el estadio histopatológico, siempre con la asesoría de un patólogo del departamento de Patología del CMN SXXI. Todos los tejidos fueron clasificados de acuerdo a la FIGO (Federación Internacional de Ginecología y Obstetricia). De esta manera, se seleccionaron las biopsias de Carcinoma invasor que presentaran al menos un 80% de células tumorales. También se verificó que el total de la muestra de los tejidos normales consistiera en tejido microscópicamente normal.

II. Obtención de Acidos nucleicos

El ADN y el ARN fueron obtenidos de los tejidos frescos utilizando homogenizadores de vidrio y 1 ml TRIZOL (Gibco BRL, Life Technologies, Grand Island NY USA). El homogenizado se incubó 15 min en Trizol a temperatura ambiente en agitación constante.

Se agregó 1 ml de cloroformo, se agitó la mezcla hasta homogenizar y posteriormente se centrifugó 15 min a 12,000 g. Para la extracción del ARN se separó la fase superior de la tricapa obtenida del centrifugado y pasándose a otro tubo, haciéndose lo mismo con la extracción de ADN.

Para obtener el ARN, a la solución obtenida mediante el Trizol se agregaron 0.5 ml de Isopropanol, se mezcló y se dejó incubar por 15 min. Posteriormente se centrifugo de la misma manera que la vez anterior. Se eliminó el sobrenadante y la pastilla se lavó con etanol al 75%. Finalmente, el ARN se suspendió en 20-50 μ l de agua MilliQ. El posible ADN contaminante de las muestras se eliminó mediante el tratamiento con DNAsal libre de RNAsa (Ambion Inc. Austin Texas USA).

Para obtener el ADN se procedió de la siguiente manera: la fase obtenida del Trizol se resuspendió en 1 ml de agua MilliQ, y se agregaron 200 μ g/ml de Proteinasa K. Esto se incubó toda la noche a 55° C y al día siguiente se extrajo el ADN con el kit Wizard Genomic ADN Purification kit (Promega), siguiendo las instrucciones del proveedor. Brevemente, al tubo con Proteinasa K se le agregaron 3 μ l de solución de RNAsa, se mezcló y se incubó a 37°C por 20 min, se dejó enfriar a temperatura ambiente y posteriormente se agregaron 200 μ l de solución de precipitación de proteínas. El tubo se centrifugó a 12,000g por 5 min y el sobrenadante se pasó a otro tubo, al cual se le agregaron 600 μ l de isopropanol, se mezcló suavemente por inversión y se volvió a centrifugar bajo las mismas condiciones, se eliminó el sobrenadante y se dejó secar la pastilla. Esta se resuspendió en 100 μ l de agua MilliQ.

III. Cuantificación y Verificación de la calidad del ADN

Tanto el ADN como el ARN fueron cuantificados por espectrofotometría y con el fin de verificar su calidad, las muestras se corrieron en un gel de agarosa al 1% con bromuro de etidio, a 100 Volts durante 90 min. El gel se visualizó con el analizador de imágenes Eagle-eye (Stratagene).

IV. Detección y Tipificación de VPH

En el caso de epitelio cervical normal, la detección viral se realizó por medio de los oligonucleótidos MY11/MY09. La detección de secuencias de VPH en las biopsias de CaCU se realizó mediante PCR empleando oligonucleótidos dirigidos a la región VPH 16/E6. La tipificación de las secuencias virales se confirmó mediante secuenciación directa del producto amplificado.

V. ANÁLISIS EN SERIE DE LA EXPRESIÓN GÉNICA (SAGE)

El procedimiento para la construcción de librerías de expresión SAGE se esquematiza en la figura 1.

i- Preparación de ADN complementario unido a adaptador

En este procedimiento, se sintetizó ADN complementario a partir de mRNA proveniente de una biopsia de cervix sin alteraciones histopatológicas, negativo para secuencias virales de VPH. Una vez sintetizado el ADN complementario, se digirió con la enzima de restricción *NaIII*, la cual reconoce la secuencia 5'-CATG-3' cortando después de G, de esta forma se produce un extremo cohesivo de cuatro bases, esta enzima corta cada 256 pb aproximadamente. Brevemente se describe a continuación la reacción de síntesis y restricción de ADN complementario:

1.-Se resuspendieron las partículas magnéticas acopladas a oligo d(T); y se añadieron a un tubo eppendorf de 2 ml.

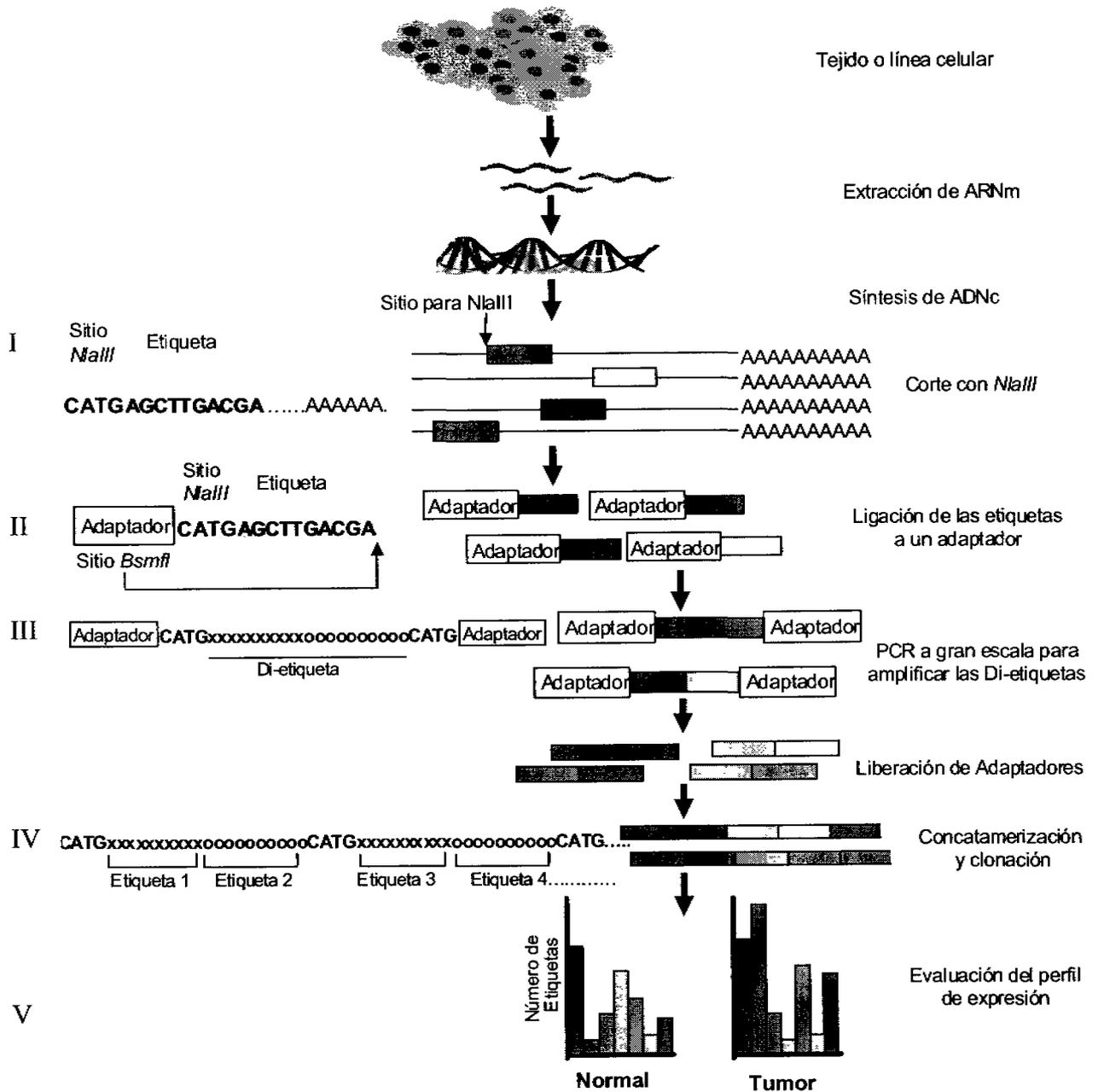


Figura 1.- Representación esquemática de la metodología de Análisis en Serie de Expresión Génica (SAGE).

2.-Se colocó el tubo con las partículas magnéticas en un aditamento especial con imán para lavar con diferentes soluciones amortiguadoras para capacitar a las partículas magnéticas a unirse al extremo poli A de la población de mensajeros presentes en la muestra.

3.- Se emplearon 5 μ g de ARN total, los cuales fueron diluidos en un volumen de 1mL. Se unió a la población de ARN poli adenilado con las partículas magnéticas, se mezclaron ambos durante treinta minutos.

4.- Se colocó el tubo en el aditamento con imán para lavar todo aquello no unido a las partículas magnéticas, es decir, capturar únicamente mRNA.

5.-Al terminar la síntesis de la primera cadena, se realizó la síntesis de la segunda cadena.

Al finalizar la síntesis de ADN complementario es necesario emplear un control de reacción, para esto se amplifica una alícuota de la muestra con oligonucleótidos específicos para GAPDH y EF2 (Tabla 1).

Sí la banda de los controles de amplificación corresponde al tamaño esperado, se continua con la digestión del ADN complementario por *NlaIII*.

ii. Restricción con *NlaIII* (sección I de la figura 1).

Esta enzima de restricción reconoce la secuencia CATG, por lo que la frecuencia teórica de reconocimiento dentro del genoma es cada 256 pb. Un transcrito maduro en promedio tiene 1 Kb, así es posible que cada transcrito presente en el tejido de interés al menos una secuencia de reconocimiento por esta enzima. Por otra parte *NlaIII* produce un extremo 3' cohesivo palindrómico. Esta característica permite que el proceso de ligación a los adaptadores sea favorecido.

La restricción del ADN complementario unido a las partículas magnéticas se llevó a cabo a 37°C durante una hora.

Tabla 1. Oligonucleótidos empleados para este estudio.

Gen	Sentido (5' 3')	Antisentido (5' 3')	TM (°C) ^a	Fragmento (bp)
DTP	GGATTTGCTGGTGCAGTACA	CTGCTCGAATTCAAGCTTCT	55	
EF-1	CATGTGTGTTGAGAGCTTC	GAAAACCAAAGTGGTCCAC	55	350
GAPDH	CATCTCTGCCCCCTCTGCTGA	GGATGACCTTGCCCACAGCCT	60	205
VPH*	GCMCAGGGWCATAAYAATGG	CGTCCMARRGGAWACTGATC	55	450
NICE-3	ACGGCTATGAAACAGCCCGCT	GCACATTGCAACTGACTGGCT	57	330
NICE-4	ACGGAATCCAATGAGGAAGGC	TCAGTATTGGCTGGCTCTGCA	57	294
S100 A8	ATGCCGTCTACAGGGATGAC	ACGCCCATCTTTATCACCAG	58	160
S100 A9	TCAGCTGGAACGCAACATAGA	TCAGCTGCTTGTCTGCATT	56	205
SPRR3	TTCCACAACCTGGAAACACA	TTCAGGGACCTTGGTGTAGC	55	174

iii. Ligación de adaptadores

El ADN complementario digerido por *NlaIII* se liga a adaptadores de ADN sintéticos. Estas secuencias sintéticas presentan algunas características importantes para el presente estudio (figura 2):

Tienen un extremo 3' cohesivo complementario a la secuencia producida por *NlaIII*.

Presentan un sitio de reconocimiento en el extremo 3' para la enzima de restricción *BsmFI* (5'- GGGAC-3'), empleada para producir las etiquetas génicas (sección II de la figura 1).

Tienen unido covalentemente un grupo amino en el carbono 7 del anillo de citosina del extremo 5'. Esta modificación es necesaria, para que la ligación de las etiquetas génicas se lleve a cabo en el extremo terminal 5'.

Finalmente, poseen una secuencia de reconocimiento a los oligonucleótidos empleados para amplificar por PCR la población de etiquetas génicas mismos que no reconocen ninguna secuencia dentro del genoma humano.

La ligación se llevó a cabo a 16°C durante dos horas, empleando la enzima T4 ADN ligasa.

iv.- Producción de etiquetas Génicas (tags).

BsmFI es una enzima de restricción del tipo IIS; así, el sitio de corte se encuentra alejado del dominio de reconocimiento a su secuencia consenso de 10 a 14 pb. De esta manera, la secuencia correspondiente a los 10-14 pb conforma la etiqueta génica, pues es la secuencia del ARN mensajero más cercana al extremo 3', adyacente al sitio de corte (CATG) de *NlaIII*. El uso de la enzima de restricción *BsmFI* permitió la generación de etiquetas génicas de 10 a 14 pb unidas al adaptador (40 pb). La reacción de restricción con *BsmFI* se realizó a 65°C durante una hora.

Adaptador A

BsmFI

5'-TTT GGA TTT GCT GGT GCA GTA CAA CTA GGC TTA ATA GGG A CATG
3'-amino-CCT AAA CGA CCA CGT CAT GTT GAT CCG AAT TAT CCC T
Extremo

cohesivo

Sitio de alineamiento
para el oligo 1

Adaptador B

BsmFI

5'- TTT CTG CTC GAA TTC AAG CTT CTA ACG ATG TAC GGG GA CATG
3'-amino-GAC GAG CTT AAG TTC GAA GAT TGC TAC ATG CCC CT
Extremo

cohesivo

Sitio de alineamiento
para el oligo 2

Figura 2.- Secuencia de los adaptadores.

v.- Producción de dietiquetas Génicas (ditags).

Posterior a la digestión, se precipito el ADN empleando fenol-cloroformo. *BsmFI* genera extremos romos; por lo tanto, se empleó la fracción Klenow de la ADN polimerasa I de *E. coli*, para polimerizar dichos extremos. La reacción de polimerización con el fragmento de Klenow se realizó a 37°C durante una hora.

Se purificó el ADN por medio de la técnica Fenol-Cloroformo.

Finalmente, se realizó una reacción de ligación con las etiquetas unidas a adaptadores, con el objeto de producir dietiquetas flanqueadas por las secuencias de los adaptadores (sección III de la figura 1).

La reacción de ligación se llevó a cabo durante 14 horas a 16°C; empleando la enzima T4 ADN ligasa.

vi.- Amplificación de las dietiquetas.

Con el fin de conocer el número de reacciones de PCR óptimo, para tener las suficientes dietiquetas, pues en los subsecuentes pasos de purificación se pierde una cantidad considerable de material, se realizó una titulación empleando diferentes diluciones del producto de PCR de las dietiquetas, utilizando como control un fragmento de ADN de concentración conocida (1µg/µL), para determinar el número de reacciones de PCR para la obtención de suficientes dietiquetas ligadas a adaptadores. De esta manera, se realizaron diluciones 1/20, 1/40 y 1/80 de las dietiquetas. Se tomó 1µl de estas diluciones y se realizó PCR empleando los iniciadores diseñados para amplificar específicamente la secuencia de los adaptadores (DTP, tabla 1). Las condiciones de PCR fueron las siguientes: 95°C dos minutos; 27 ciclos de 95°C un minuto, 55°C 1 minuto y 72°C 2

minutos; finalmente 72°C durante cinco minutos. El producto de PCR de esta reacción se corrió en un gel de agarosa a una concentración de 4% para identificar dos bandas de 80 y 100 pb. Se cuantificaron por picogreen la concentración de ADN contenida en las PCRs de cada dilución. Esto permitió determinar que la cantidad óptima de reacciones de PCR fueran 200, pues la concentración de ADN obtenida en la dilución 1/40 fue de 20 ng/μl. En el reporte original de este método se recomiendan 200 PCRs cuando la cantidad de la dilución 1/40 este entre 5-250 ng/μl.

vii. Aislamiento y purificación de las dietiquetas de 100 pb.

Los productos de las 200 PCR's fueron recuperados por medio del método Fenol-Cloroformo. El ADN precipitado fue resuspendido en 500μl de LoTE. Se preparó un gel de acrilamida al 12% para separar las bandas que corresponden a las dietiquetas (100 pb) con los adaptadores ligados sin dietiquetas (80 pb). La electroforesis se realizó en una cámara vertical BioRad a 200 V. Se concluyó cuando el colorante xilen-cyanol contenido en el Buffer de carga se encontraba a 2 centímetros del frente de corrida.

El gel se tiñó con bromuro de etidio (0.5 μg/mL), para visualizar y cortar la banda que correspondiera a 100 pb. El ADN contenido en dicha banda, fue eluido y purificado.

viii. Purificación de las dietiquetas.

Como se aprecia en la figura 2, los adaptadores contienen un extremo cohesivo que en realidad es la secuencia de reconocimiento de *NlaII*; de esta forma la digestión con dicha enzima permitió liberar los adaptadores de las dietiquetas. La restricción se llevó a cabo a 37°C durante toda la noche. Posteriormente, el ADN se purificó con Fenol-Cloroformo. Se resuspendió en 50 μl de LoTE y se corrió en un gel preparativo de poliacrilamida al 12%.

Después de teñir el gel con bromuro de etidio, se visualizaron y cortaron las bandas de 26 pb que corresponden a las dietiquetas sin adaptador. Finalmente se eluyó y precipitó el ADN presente en dichas bandas.

ix. Formación de concatámeros y clonación.

Se ligaron las dietiquetas empleando T4 ADN ligasa durante 3 horas (sección IV figura 1). Transcurrido este tiempo se corrieron en un gel de poliacrilamida al 8%. Después de teñir el gel con bromuro de etidio, se observó un barrido de aproximadamente 100pb a 1 Kb. Se dividió dicho barrido en tres regiones: 300-500 pb; 500-800 pb y 800-1Kpb. Finalmente se eluyó y precipitó el ADN presente en dichas bandas.

Se emplearon 2 μ g por cada región descrita arriba del plásmido pZero-1 linearizado con *SphI* (Figura 3). La ligación se llevó a cabo durante 3 horas con T4 ADN ligasa. Se purificó y precipitó el ADN, para finalmente eluirlo en 12 μ l de LoTE.

Para la transformación de bacterias se empleó el método de electroporación, pues es más efectivo cuando se cuenta con cantidades limitadas de ADN. Se añadió 1 μ l de concatámeros clonados por cada vial de bacterias *E. coli* electrocompetentes. Las condiciones de electroporación que se establecieron en el presente estudio fueron las siguientes: 200 Ω de resistencia, 25 μ F de capacitancia a un voltaje de 1.98 KV. Después de la electroporación, se añadió 1 mL de medio SOC, 100 μ l fueron plaqueados en medio selectivo (50 μ g/ml de Zeocina). Finalmente, después de cultivar cada placa a 37°C durante toda la noche, las colonias resultantes fueron tomadas con un palillo estéril y por medio de PCR directa se realizó el análisis de transformantes. Se utilizaron los

oligonucleótidos M13 sentido y antisentido (Tabla 1) para amplificar el inserto correspondiente a los concatámeros. El producto de PCR de cada colonia fue corrido en un gel de agarosa al 1% para determinar el tamaño de dicho inserto. Para el proceso de secuenciación, solo se emplearon aquellos insertos mayores a 500 pb. En el presente estudio se realizaron aproximadamente 1,500 reacciones de secuenciación para obtener el transcriptoma completo del epitelio Cervico-uterino.

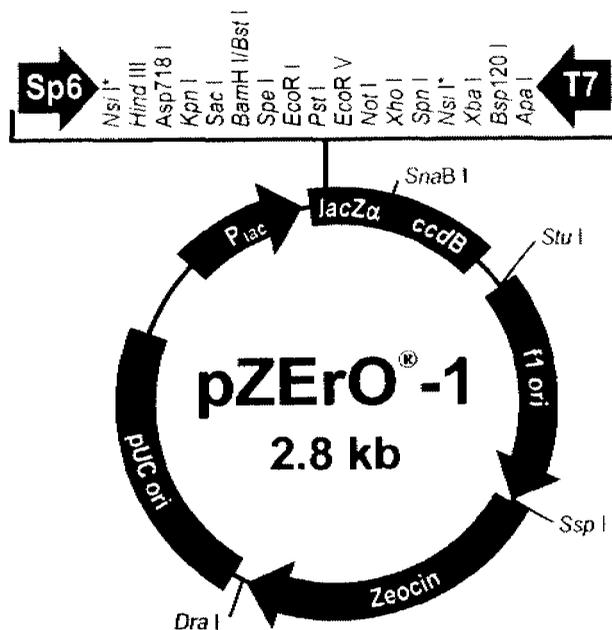


Fig 3. Mapa de pZero. El sitio múltiple de clonación se encuentra en el marco de lectura abierto del gen *ccd*, el cual inhibe la función de la ADN girasa, por lo que las células que adquieren el plásmido no recombinante son incapaces de replicarse en medio selectivo.

x. Análisis de Datos

El resultado final de una biblioteca de expresión tipo SAGE, es una lista en la cual se muestra la frecuencia de cada etiqueta presente. Para obtener esta lista, es necesario extraer los datos generados por el software del secuenciador automático. Así, los datos crudos son obtenidos a partir de los archivos con extensión *.seq. El programa de análisis que recupera las secuencias de las dietiquetas y genera un reporte de cada etiqueta con su nivel de abundancia es SAGE2000.

Para relacionar cada etiqueta con el gen del cual proviene, se empleó la base de datos "most reliable gene to tag"; es decir, se utilizaron las secuencias de etiquetas con una mayor correspondencia al gen del cual provienen, identificadas por RefSeq. Dicha base de datos se encuentra disponible en el sitio <ftp://ftp.ncbi.nlm.nih.gov/pub/sage/map/Hs/NlaIII/>. Finalmente, para poder relacionar ambas tablas, y extraer la información pertinente a cada etiqueta, se empleó el programa Access de Microsoft.

VI. Microarreglos de Expresión

En este trabajo, nuestro principal objetivo fue obtener el transcriptoma completo del epitelio cervical, de esta manera SAGE permitió en parte cumplir dicho objetivo. El principal inconveniente fue el costoso esfuerzo de secuenciar la biblioteca correspondiente al tejido epitelial normal. De esta manera, se decidió emplear otra herramienta que nos permitiera cumplir con los objetivos planteados inicialmente, pero a un costo menor. Así fue como se utilizó la plataforma de microarreglos CodeLink Human Whole Genome Bioarray de General Electric, la cual en una sola laminilla presenta impresas alrededor de 55,000 secuencias diferentes, lo que teóricamente representaría el transcriptoma humano. Estas secuencias fueron obtenidas de las bases de datos de EST's y RefSeq del NCBI y UniGene. Esta plataforma tiene la sensibilidad de detectar una copia por célula y la especificidad de discriminar secuencias altamente homólogas. Esto es posible gracias a que las sondas se encuentran inmersas en una matriz tri-dimensional semisólida que permite un adecuado proceso de hibridación (53-55).

i.- Tejidos, síntesis de cRNA e Hibridación.

Se emplearon tres biopsias de cervix normal y ocho biopsias de carcinoma invasor en estado clínico IIB, positivos para VPH 16.

En esta plataforma de expresión se empleó ARN total como fuente de análisis. La síntesis del ADN complementario se realizó con Superscript II reverso transcriptasa y un oligonucleótido T7. Posteriormente, la segunda cadena del ADN complementario se produjo usando ADN polimerasa I de *E. coli* y RNasa H. El ADN complementario bicatenario resultante, fue purificado en columnas QIAquick (Qiagen, Valencia, CA). Se sintetizó ARN complementario (cRNA) por medio de una reacción de transcripción *in vitro*, usando T7 ARN polimerasa y biotina-11-UTP (Perkin-Elmer, Boston, MA). El cRNA fue

purificado en una columna RNeasy (Qiagen) y cuantificado por espectrofotómetro. Finalmente, se emplearon 10 µg de cRNA que fue fragmentado previamente a 94°C durante 20 minutos en una solución con magnesio. El cRNA fragmentado fue hibridado por 14 horas a 37°C sobre un arreglo de CodeLink Whole Human Genome Bioarray, empleando una incubadora con agitación orbital Innova 4080 (New Brunswick, Edison, NJ) a 300 revoluciones por minuto.

ii.- Post-hibridación, procesamiento y escáneo.

Después de la hibridación, los arreglos fueron lavados en buffer 0.75X TNT filtrado (0.10M Tris-HCl pH 7.6, 0.15M NaCl, 0.05% Teen 20) a 46°C durante 1 hora. Posteriormente, se incubaron a temperatura ambiente, por 30 minutos, en la oscuridad en una solución con estreptavidina-Cy5 (Amersham, Piscataway, NJ USA). Después, los arreglos se lavaron con buffer 1X TNT durante 15 minutos. Como lavado final, se empleó una solución con 0.05% de Tween 20 en agua por 30 segundos, con agitación gentil.

Las laminillas se secaron por centrifugación y se mantuvieron en la oscuridad hasta el procedimiento de escáneo; en ninguno de los casos fue mayor a dos días. Las imágenes fueron capturadas empleando un scanner de la marca GenePix (Arlington, TX).

iii.- Normalización y análisis estadístico.

Las imágenes fueron analizadas empleando el software CodeLink Expression Analysis v4.1. Los valores de expresión fueron globalmente normalizados al valor de la mediana de expresión de todos los puntos en los arreglos. Para evitar problemas al manejar divisores en cero, todos los valores de expresión iguales a ó por debajo de tres veces el valor de la línea del ruido de fondo (background noise) fueron tomados de manera arbitraria como 5. Este procedimiento nos permitió eliminar los eventos con un coeficiente

de variabilidad mayor al 10%, disminuyendo de esta manera la probabilidad de encontrar falsos positivos (55).

Una vez que los valores de expresión estuvieron normalizados, se obtuvieron las tasas de expresión. En este caso utilizamos la mediana de expresión para cada gen de los tres tejidos normales que se emplearon como referencia, para compararla con los valores de expresión con los ocho tejidos tumorales. Se obtuvieron los logaritmos en base 2 de las tasas de expresión para visualizar aumento o disminución en los niveles de expresión de un determinado transcrito.

Los datos fueron analizados inicialmente empleando cluster jerárquico; este procedimiento permite visualizar a los genes de manera agrupada en base a las similitudes en los valores de expresión para cada muestra analizada. De esta manera, la representación gráfica permite observar sobre- expresión (rojo) o sub-expresión (verde) de un gen determinado. Por otra parte, en el eje horizontal se agrupan de manera adyacente a los tejidos o muestras cuyos patrones de expresión son similares. El agrupamiento jerárquico se realizó por medio del programa Cluster; para visualizar el dendrograma, se usó el programa TreeView (56, 57).

El análisis del patrón de expresión génica por agrupamiento jerárquico permite obtener poca información sobre la significancia estadística (*op. cit*). Por esta razón empleamos un método estadístico desarrollado en la Universidad de Stanford, conocido como Análisis Significativo de Microarreglos (SAM, por sus siglas en inglés). Este algoritmo permite identificar genes con cambios estadísticamente significativos en sus niveles de expresión entre dos grupos de muestras diferentes (58). Seleccionamos aquellos genes que tuvieran una tasa de falsos positivos (FDR) menor al 10%. Así, los genes que se

obtuvieron por este algoritmo fueron empleados para los subsecuentes análisis.

iv.- Caracterización de vías metabólicas involucradas

Empleamos el programa Pathway Express (PE) de la universidad de Wayne (<http://vortex.cs.wayne.edu/Projects.html>) para conocer las rutas metabólicas alteradas en Carcinoma Cervical. A partir de una lista de genes y su tasa relativa de expresión, PE busca en la base de datos de la Enciclopedia Kyoto de Genes y Genomas (KEGG) (<http://www.genome.ad.jp/kegg/>) las vías metabólicas asociadas. Después de generar dicha lista, el programa primero calcula un factor de perturbación $PF(g)$ para cada gen; el cual toma en cuenta: (i) la tasa relativa de expresión normalizada para el gen en cuestión y (ii) el número de genes y la tasa de expresión relativa para cada gen que se encuentren dentro de una ruta metabólica particular. El factor de perturbación refleja así la importancia relativa de cada gen diferencialmente regulado en una ruta metabólica particular. Posteriormente se calcula un factor de impacto para la vía metabólica completa, que incluye un valor numérico que toma en consideración: (i) la cantidad de genes diferencialmente regulados en determinada vía y (ii) el global de los factores de perturbación para dicha vía. Los factores de impacto de las vías involucradas se emplean para asignar la posición de cada vía con respecto a las otras (59).

6. RESULTADOS

En este trabajo, nuestro principal objetivo fue obtener un análisis completo de los genes expresados en el epitelio cervical; tanto en el estado normal o histopatológicamente sano como en tejidos invasores en estadio IIB. En este sentido, empleamos dos metodologías que nos permitieron realizar nuestros objetivos: Análisis en Serie de la Expresión Génica (SAGE) (Anexo 1) y microarreglos de expresión (Anexo 2).

I.- Detección de Secuencias de Virus de Papiloma Humano (VPH)

Se purificó ADN y ARN total de las biopsias del cuello uterino. Para determinar la presencia de VPH, se emplearon oligonucleótidos universales MY09/MY11, se tomaron las biopsias de Cérvix Uterino normal que no presentaran secuencia viral alguna.

II. Análisis en Serie de la Expresión Génica (SAGE)

i.- Síntesis de ADN complementario y digestión por *NlaIII*.

En la figura 4 se observa la eficiencia de síntesis de ADN complementario y la eficiencia del proceso de digestión del mismo empleando la enzima de restricción *NlaIII*. Se utilizó una pequeña alícuota de ADN complementario y se buscó la presencia de secuencias de los genes constitutivos GAPDH y EF-1 α (carriles 2 y 4). Para verificar que la reacción de digestión fuera eficiente se empleó una alícuota del producto de digestión de *NlaIII*. Se diseñaron los oligonucleótidos de GAPDH de tal forma que el oligonucleótido antisentido quedará fuera de uno de los sitios de corte de la enzima, por lo que después de la digestión con *NlaIII* se pierde un sitio de alineamiento para GAPDH (caril 3). Por lo tanto, fue posible verificar que los procesos de síntesis de ADN complementario y de digestión del mismo se llevaron a cabo de manera óptima.

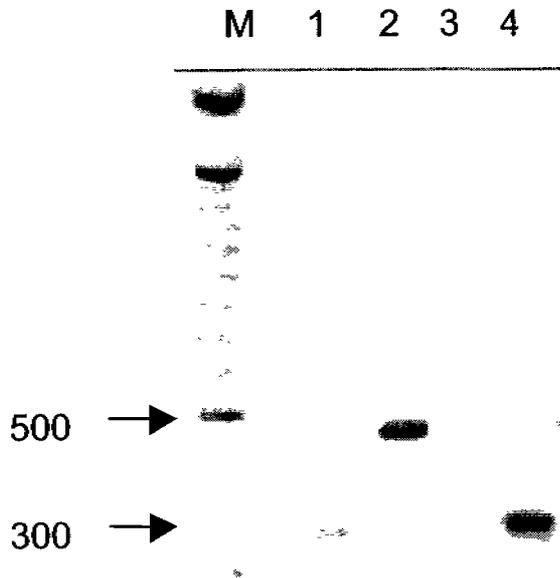


Fig 4. Eficiencia de síntesis de ADN complementario y de restricción con *NlaIII*. Después de sintetizar el ADN complementario se tomó una alícuota para amplificar los genes constitutivos EF-1 α (carril 1) y GAPDH (carril 2). Después de digerir con *NlaIII*, se pierde el sitio de alineación del oligo sentido para GAPDH (carril 3), conservándolo en el caso de EF-1 α (carril 4). M, marcador de peso molecular.

ii.- Amplificación de las dietiquetas por PCR.

Con el objeto de conocer el número de reacciones óptimas para amplificar las dietiquetas ligadas a los adaptadores, se realizó una titulación del contenido inicial de éstas. En la figura 5 se observa una mayor intensidad en la dilución 1:80 del producto de PCR de las dietiquetas en comparación con el fragmento de ADN usado como control. Por lo tanto, la cantidad de dietiquetas que se tiene se encuentra en el rango de 40-200ng/ μ L, de esta manera el número de PCR óptimo según el protocolo inicial de SAGE es de 200 reacciones. Este resultado fue confirmado empleando un fluorómetro, cuya lectura dio un resultado de 250ng// μ L. Sin embargo, es importante la reacción de PCR para confirmar que el producto final tendrá el tamaño esperado. En el caso de SAGE, la identificación de los genes expresados depende de la secuencia, por lo que en todos los casos de amplificación se utilizan ADN polimerasa de alta fidelidad; disminuyendo así la identificación errónea de genes involucrados en el modelo de análisis.

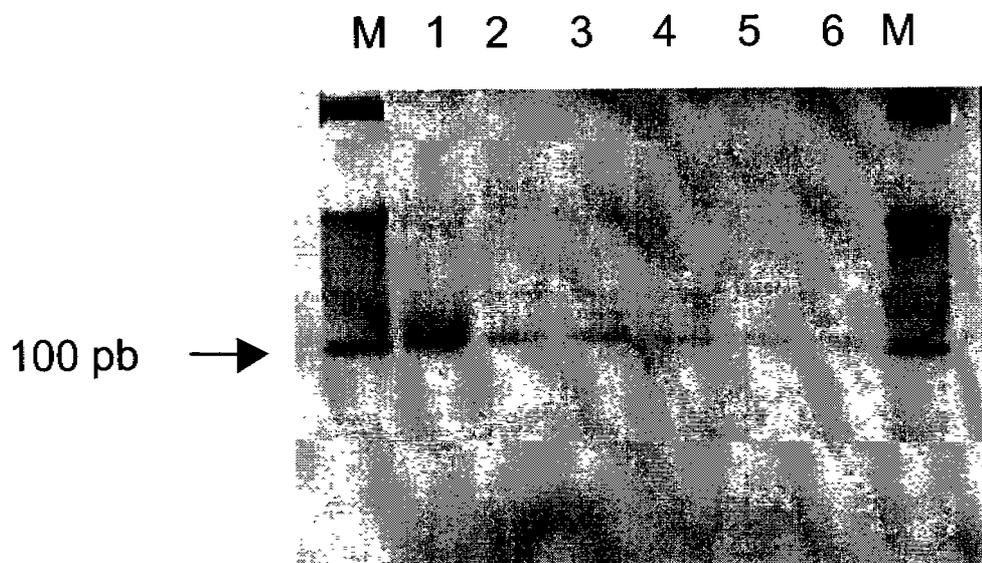


Fig 5. Electroforésis en gel de agarosa al 4% de diferentes diluciones de dietiquetas ligadas a los adaptadores. Carriles 1-3 corresponden a diluciones 1/20, 1/40 y 1/80 respectivamente. Carriles 3-6 corresponden a las mismas diluciones del control positivo para esta reacción con una concentración inicial de 200 ng/ μ l de ADN.

iii.- Purificación de Dietiquetas.

Después de realizar las 200 reacciones de PCR para amplificar la población de dietiquetas ligadas a los adaptadores, se colectaron todas las reacciones en un solo tubo; el ADN presente fue purificado y precipitado. En la figura 6 se observa una electroforésis vertical en gel de poliacrilamida al 8% para separar las bandas de 100 pb que corresponden a las dietiquetas ligadas a adaptadores. Las otras bandas observadas corresponden a secuencias de adaptadores sin etiquetas; evidentemente en la reacción de ligación, los adaptadores se encuentran en exceso, así cuando son amplificados, la población que no contiene dietiquetas ligadas aparece como una banda de 80 pb.

iv.- Digestión con *Nla* III.

Para separar las dietiquetas de los adaptadores, se realizó una digestión con la enzima *Nla* III. El producto de la digestión fue corrido en un gel preparativo de acrilamida al 12% (Fig. 6) con el objeto de recuperar las bandas correspondientes a las dietiquetas de 28 pb para la formación de concatámeros. Las bandas de ADN correspondientes a 100 y 60 pb, son digestiones parciales o incompletas. La banda que corresponde a ~ 40 pb son los adaptadores amplificados.

v.- Formación de Concatámeros y Clonación.

La presencia del extremo cohesivo reconocido por *Nla* III presente en las dietiquetas, hizo posible ligarlos para formar concatámeros de diferente tamaño. Para seleccionar las diferentes poblaciones de concatámeros en función del número de dietiquetas ligadas, se realizó una electroforésis vertical en gel de acrilamida al 8%. En la figura 8 se observa un barrido que corresponde a concatámeros de diferente tamaño. Se seleccionaron tres regiones del gel; una de 300 a 500 pb, otra de 500 a 800 pb y de 800 a 100 pb. Estas tres regiones fueron clonadas cada una en el plásmido pZErO-1. Una alícuota de la mezcla de

reacción fue electroporada a la cepa de *E. Coli* TOP 10; las bacterias recombinantes fueron seleccionadas y analizadas mediante PCR empleando primers específicos para PUC M13.

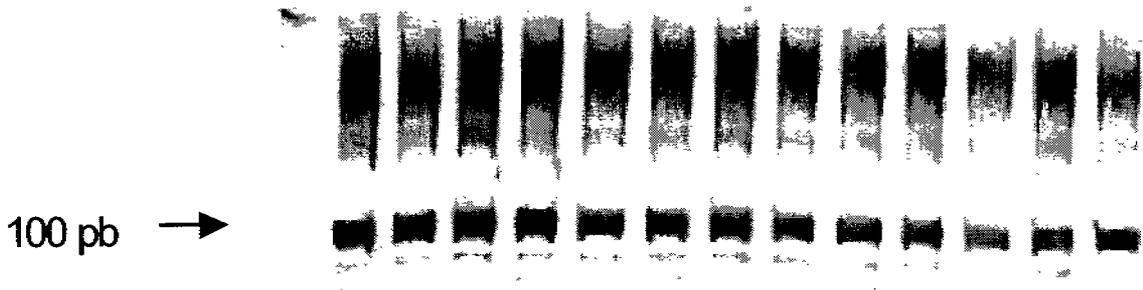


Fig 6. Purificación de dietiquetas de 100 pb, empleando un gel preparativo de acrilamida al 8%. Las bandas de 80 pb corresponden a adaptadores que fueron ligados sin contener dietiquetas

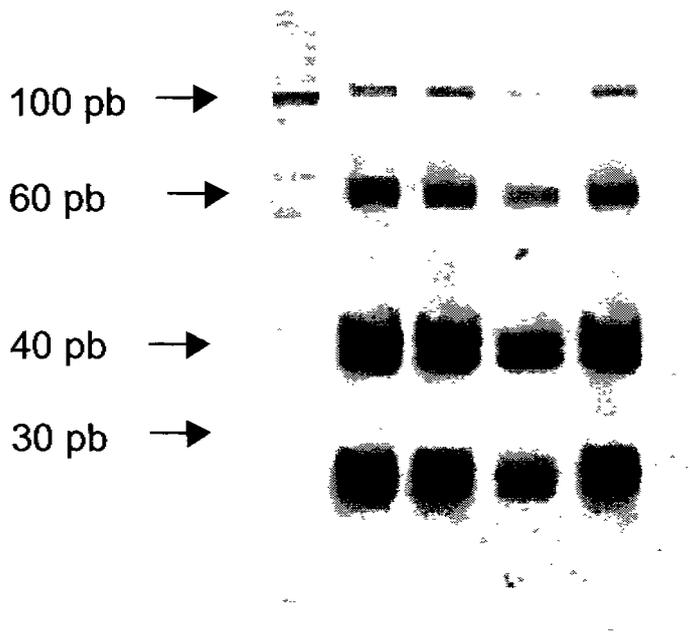


Fig 7. Purificación de dietiquetas de 28 pb en un gel preparativo de acrilamida al 12%. Después de la digestión con NlaIII, se generan diferentes fragmentos; el de 28 pb corresponde a las dietiquetas, los adaptadores de 40 pb y los digeridos parciales de 60 y 100 pb

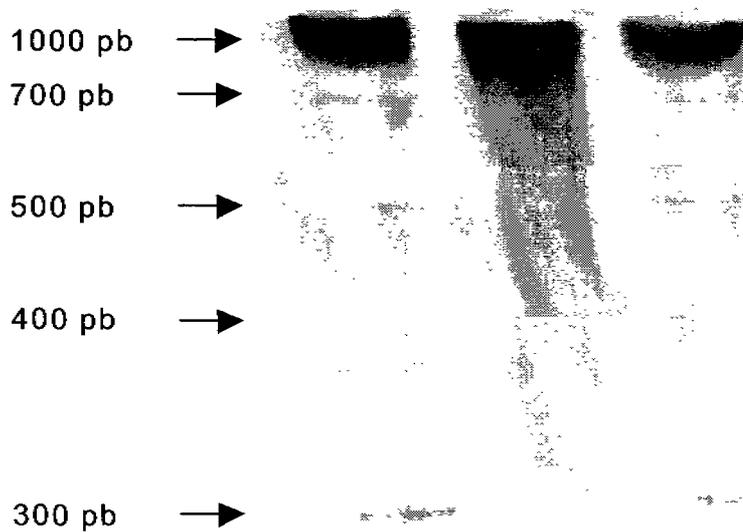


Fig 8. Purificación de concatámeros. Se aislaron tres regiones del gel de acrilamida al 8%; una de 300-500 pb, otra de 500 a 800 pb y de 800 a 1 Kpb

vi.- Secuenciación e identificación de etiquetas.

En la figura 9 se observan los productos de PCR correspondientes a los insertos amplificados de las colonias seleccionadas. Una alícuota de la reacción de PCR, se corrió en un gel de agarosa al 1%; aquellos productos que presentaron un tamaño mayor de 500 pb fueron purificados y cuantificados. Es importante seleccionar los insertos por su tamaño, pues esto permite optimizar las reacciones de secuenciación; un tamaño mayor del inserto corresponde un número mayor de dietiquetas secuenciadas.

Los insertos fueron secuenciados en un secuenciador de Applied Biosystems modelo 310. En la figura 10 se observa el electroferograma típico de una reacción de secuenciación de SAGE, las dietiquetas se encuentran separadas por la secuencia reconocida por la enzima *NlaIII* (CATG), dicha secuencia en esta metodología se le conoce como signo de puntuación. Los archivos con extensión *.seq fueron utilizados para alimentar el software de análisis SAGE2000 (Johns Hopkins University) que identifica los signos de puntuación CTAG, extrayendo la secuencia de cada Etiqueta. De esta manera, se produce una base de datos que muestra las secuencias de 10 pb que corresponde a cada Etiqueta; por medio de una conexión al portal del NCBI, es posible identificar el gen al que corresponde la secuencia de cada etiqueta.

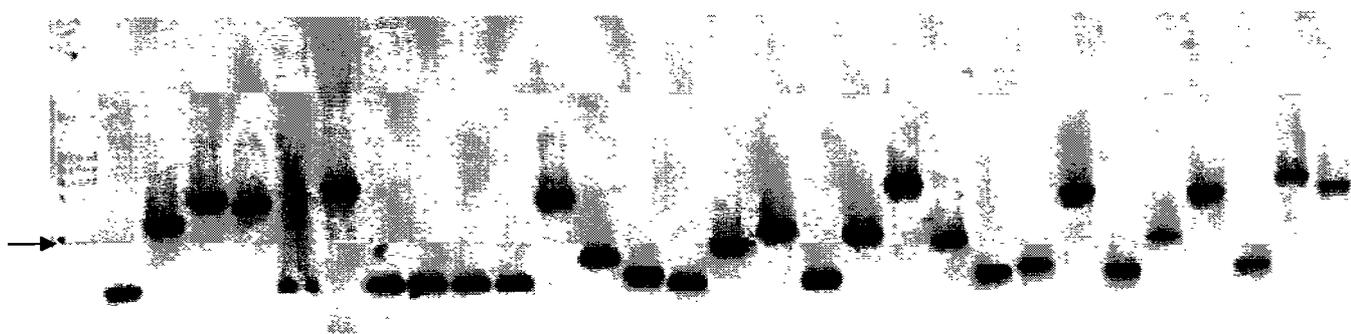


Fig 9. Tamaño de los insertos de las colonias crecidas en placas selectivas con Zeocina. Aquellas clonas que presentaron un inserto mayor a 500 pb fueron secuenciadas. La flecha indica un tamaño de 500 pb.



Fig 10. Fragmento de secuencia en donde se muestran los sitios de puntuación CATG dejados por la enzima de restricción NlaIII. En medio de dos de estos sitios se encuentra una di-etiqueta de 24 pb; el software de análisis extrae la secuencia de la di etiqueta y reporta su frecuencia en una base de datos.

vii.- Caracterización de los genes expresados.

En el presente estudio, se realizó una biblioteca de expresión tipo SAGE que se denominó SAGE_cervix_normal_B_1, dicha biblioteca puede ser consultada en línea en el sitio SAGEmap del NCBI <http://www.ncbi.nih.gov/projects/SAGE/> (figura 11). Esto permitió a nuestro grupo, obtener un reconocimiento como colaboradores del Instituto Nacional de Cáncer (NCI) de los Institutos Nacionales de Salud (NIH) (anexo 3).

La caracterización detallada de la biblioteca se encuentra en el anexo 1. Se secuenciaron un total de 30,418 etiquetas, que correspondieron a 8,062 etiquetas únicas; de las cuales 5,255 estuvieron relacionadas a un gen particular. Las restantes 2,807 etiquetas no tuvieron relación con ningún gen en la base de datos de SAGEmap. Finalmente, 4,340 genes individuales fueron caracterizados, de los cuales 1,215 hasta este momento tienen un nivel de anotación pobre; es decir, aparecen en las bases de datos sin función conocida. En la tabla 2 del anexo 1 aparecen los genes con el mayor nivel de expresión.

viii.- Validación de los resultados.

Con el fin de corroborar los datos de expresión producidos por medio de una metodología de alto rendimiento; empleamos una estrategia de análisis convencional o de búsqueda de un solo gen, como es la Reacción en Cadena de la Polimerasa acoplada a Retro Transcripción (RT-PCR), en un grupo de seis muestras. En la figura 12 se puede apreciar la expresión de cinco genes seleccionados en seis muestras de epitelio cervical normal sin alteraciones histopatológicas ni presencia de ningún tipo de HPV. Este resultado nos permite confirmar que los datos observados en una sola muestra pueden ser representativos del epitelio cervical normal.

GEO help: Mouse over screen elements for information

Scope: Format: Amount: GEO accession:

Sample GSM2455

Status Public on Oct 03, 2002
Title SAGE_cervix_normal_B_1
Sample type SAGE
Anchor NlaIII
Tag Count 30418
Tag Length 10

Source Name normal exocervix
Organism(s) Homo sapiens
Extracted molecule total RNA

Description Uterine cervix normal SAGE, CGAP non-normalized SAGE library, bulk method, HPV negative
Keywords = uterine cervix

Submission date Oct 03, 2002
Contact name Carlos Perez-Plasencia and Mauricio Salcedo
E-mail(s) car_plaz@imss.com
Phone +56.27.69.00 ext 4323
Fax +57.61.09.52
Organization name Mexican Institute for Social Security
Department UIMEO
Lab Oncology Genomics
Street address Cuauhtemoc 300
City Mexico
State/province D.F.
ZIP/Postal code 06720
Country Mexico

Platform ID [\[?\]](#)

Platform ID GPL4

Data table header descriptions

TAG Tag sequence
COUNT Frequency for each TAG
TPM Tags per million

Data table

TAG	COUNT	TPM
TACCTGCAGA	515	16930.76468
TAGGTGTCT	356	11703.59655
TTCCTGCCTC	276	9073.574857
GAGGGAGTTT	201	6607.929515
GTGACCACGG	188	6180.55099
GTGGCCACGG	184	6049.049905
GGGCTGGGGT	173	5687.421921
GCATAATAGG	168	5523.045565
TCAGATCTTT	161	5292.918667
GTTGTGGTTA	155	5095.667039
GGATTTGGCC	151	4964.165954
TTGGGGTTTC	143	4701.163785
TTGGTCTCTT	138	4536.787428
TGCACGTTT	130	4273.785259
ACAAAGCATT	128	4208.034716
AGGGCTTCCA	122	4010.783089
CCACTGCACT	114	3747.780919
GGCAAGCACC	107	3517.654021
CTGGTTAAT	105	3451.903478
TGCATTLAA	103	3386.152936

Total number of rows: **8082**

Table truncated, full table size **197 Kbytes**.

[View full table...](#)

Supplementary data files not provided

Fig 11 (cont). La base de datos que corresponde al transcriptoma del epitelio cervical normal, puede ser consultada en línea en la página web del NCBI <http://www.ncbi.nih.gov/projects/SAGE/>. Cada etiqueta génica presenta un vínculo a la información correspondiente al gen al que pertenece; así como su nivel de expresión en otras bibliotecas SAGE.

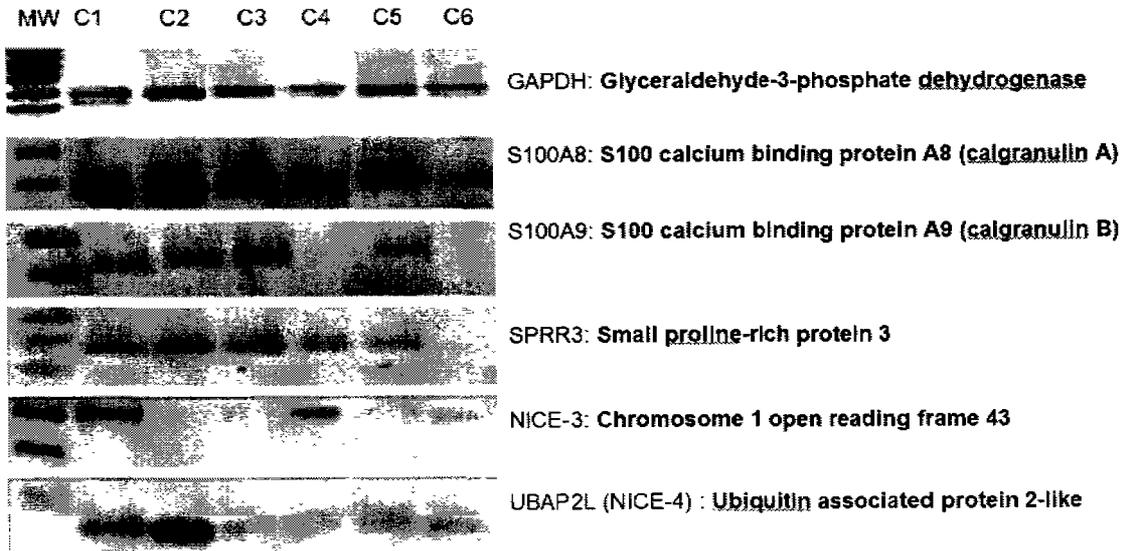


Figura 12. Validación de los resultados de SAGE en diferentes muestras de cervix normal. Se empleó un grupo de seis muestras de epitelio cervical sin alteraciones histopatológicas ni presencia viral, para validar por medio de RT-PCR la expresión de genes caracterizados por medio de SAGE.

II. Microarreglos.

i.-Hibridación y escáneo.

La imagen que se obtuvo del proceso de hibridación, es reflejo del nivel de expresión de cada gen (Figura 13), pues la plataforma de microarreglos que elegimos permite obtener el perfil de expresión global de un tejido particular, debido a que no se requiere una hibridación substractiva (dos colores). Esto significa que el nivel de expresión para cada gen se manifiesta por la intensidad de fluorescencia en el punto correspondiente. Por otra parte, gracias al empleo de controles tanto positivos como negativos (Figura 14), es posible tener un estricto control en la calidad del experimento.

ii.-Análisis por agrupamiento jerárquico.

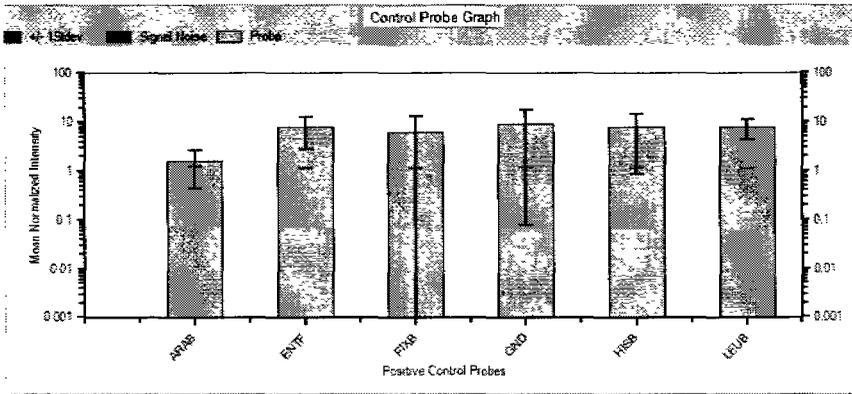
Una vez que los logaritmos de las tasas de expresión (Log_2R) fueron obtenidas (como se describe en la sección de material y métodos), se filtraron los datos eliminando aquellos genes que tuvieran un cambio menor a dos en todos los tejidos. De esta forma obtuvimos 6,000 genes diferencialmente expresados; los cuales corresponden a 3,248 transcritos caracterizados. Estos 6,000 genes fueron empleados para realizar un agrupamiento jerárquico (Fig. 15), empleando para ello el programa Cluster de Eisen (56). Es interesante notar que los tumores presenten entre ellos menores diferencias que los tejidos normales. Este hecho pudiera ser explicado por que todos los tumores empleados en este estudio presentaron infección viral por VPH16, de tal suerte que las alteraciones inducidas por este tipo viral en las siete muestras analizadas son similares.

Por otra parte, los tejidos normales mostraron una gran diferencia entre ellos, esto puede explicarse por el estadio del ciclo menstrual en el que se encontraban; sin embargo, no pudimos tener acceso a esta información. En este sentido la información a la que tuvimos acceso fue la edad, misma que oscilaba entre los 32 y 41 años; ninguno de los pacientes

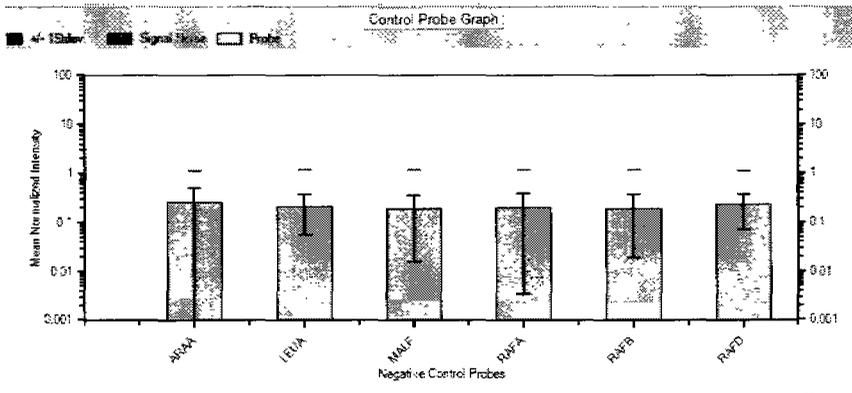
recibía en ese momento terapia anticonceptiva hormonal. Las diferencias globales observadas en los epitelios normales pudieran ser explicadas por el número reducido de muestras.



Figura 13.- Acercamiento de una hibridación escaneada. La imagen fue producida por el Software GenePix Pro v5.0; cada punto representa un gen; la intensidad de fluorescencia esta en función del nivel de expresión de dicho gen.



a)



b)

Figura 14. Controles de calidad en el microarreglo. Dentro del arreglo, existen diferentes controles, tanto positivos (a) como negativos (b); dichos controles son transcritos bacterianos que corresponden a diferentes genes. Cada sonda (representada por una columna en la gráfica) esta distribuida 15 veces dentro del arreglo, de esta forma es posible evaluar la calidad tanto del experimento global, como de zonas específicas en la matriz de puntos del arreglo. La línea roja en las imágenes, representa el nivel de señal de fondo (background).

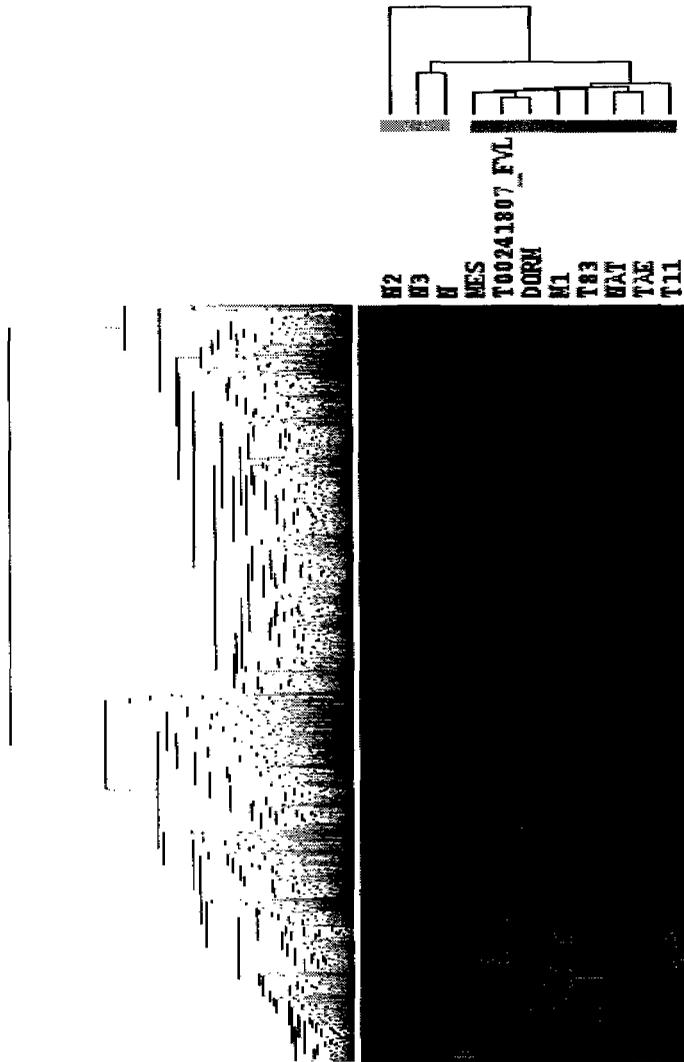


Figura 15. Agrupamiento jerárquico en CaCU VPH16+. Los datos de microarreglos fueron analizados por el programa Cluster de Eisen y visualizados por TreeView. El agrupamiento mostrado, representa alrededor de 6,000 genes. Cada fila es un gen, mientras que cada columna representa un tejido. La línea de color por encima de las columnas indica: verde, tejidos normales; rojo, tumorales. La abundancia relativa del gen en el tejido correlaciona con la intensidad del color; rojo, sobre-expresado; verde, sub-expresado; negro, sin alteración. En el dendrograma, las ocho muestras tumorales se agrupan, mientras que las normales presentan diferencias mayores.

iii.-Análisis Significativo de Microarreglos.

Se reanalizaron los 6,000 genes filtrados por el programa Cluster para encontrar genes que permitieran discriminar de manera significativa los estados normal del tumoral; asimismo, empleando Análisis Significativo de Microarreglos (SAM) (58) encontramos 1,135 y 1,701 genes que fueron sobre y sub expresados respectivamente, con una tasa de descubrimiento falso (FDR) menor al 10%. Así, este análisis supervisado permitió identificar un grupo de genes asociados a carcinoma cervical con infección por VPH 16 (Tablas 1 y 2 del anexo 2).

iv.-Vías metabólicas involucradas.

Para conocer el significado biológico de los genes cuya expresión encontramos alterada en CaCU positivos para VPH16, los genes que obtuvimos por SAM fueron analizados por el programa Pathway Express (PE)(59,60). PE es una herramienta desarrollada recientemente y que se emplea para visualizar datos de expresión en el contexto de vías metabólicas KEGG. La importancia de esta herramienta de análisis consiste en la obtención de un factor de impacto (IF) de la ruta involucrada, el cual puede ayudar a tener una idea más clara del nivel de alteración de dicha ruta. Así, importamos al programa una lista de genes significativamente sobre y sub expresados para convertir los datos de expresión en ilustraciones de rutas metabólicas alteradas. Con el fin de evitar errores en la asignación de IFs en las rutas alteradas, debido al tamaño de la muestra para la lista de genes sobre y sub expresados; se sometieron una cantidad similar de genes en cada caso. De esta forma, encontramos alteraciones importantes en diferentes rutas metabólicas en los carcinomas cervicales (Tabla 3 del anexo 2). En las figuras 2, 3 y 4 se puede observar las vías de señalización de Wnt, TGF-beta y Calcio respectivamente.

7. DISCUSION

I. SAGE

La biblioteca de expresión que obtuvimos fue derivada de tejido ectocervical sin alteraciones histopatológicas; es decir, clínicamente sano. Este tejido provino de una donadora de 38 años de edad, con vida sexual activa. Al momento de la toma de biopsia la paciente no se encontraba bajo terapia anticonceptiva hormonal, o alguna otra que pudiera alterar potencialmente la fisiología cervical. Estos criterios de inclusión de la muestra, nos permitieron corroborar los datos de expresión en un grupo mayor de muestras en cinco genes analizados (figura 12). Este resultado permite validar que los datos de expresión de una sola biblioteca SAGE podrían representar el transcriptoma del epitelio cervical normal.

Los genes que elegimos para validar los datos de expresión de la biblioteca SAGE proveniente de tejido normal cervical, se encuentran en el cromosoma 1q21. En esta región cromosómica se agrupan al menos 27 genes divididos en tres familias, que participan en la diferenciación terminal de tejidos epiteliales (60). Hasta este momento no se ha caracterizado el papel que tienen dichos genes en los procesos de diferenciación del epitelio cervical normal y su contraparte tumoral (artículo en preparación).

La biblioteca SAGE produjo 30,418 etiquetas secuenciadas, las cuales se emplearon para generar una tabla que representa los genes expresados en el epitelio cervical normal. Para tener acceso a la lista completa de genes expresados, por favor visitar el sitio de Internet SAGEmap (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2455>). El catálogo de genes expresados derivado de este estudio representa el primer intento para generar un análisis completo del perfil de expresión del epitelio cervical.

Entre los genes con un mayor nivel de expresión, algunos corresponden a genes ubicuamente expresados (*GRIN2C*, *FTH1*, *GNS*, *RPLP2*, *RPL21*) (Tabla 2 anexo 1); la presencia de este tipo de genes es un resultado común en bibliotecas de expresión SAGE; sin embargo, sus niveles de expresión presentan una importante heterogeneidad. En un reporte de Velculescu *et al.*, empleando análisis de bases de datos de bibliotecas SAGE se encontró que alrededor de mil genes se expresan en todos los tejidos tanto tumorales como normales (32). A esta lista de genes se le ha denominado transcriptoma mínimo (es decir, el grupo de genes expresado en cualquier tipo celular). Para conocer si los datos de expresión del epitelio cervical normal contenían dichos genes, accedimos a la información suplementaria del trabajo de Velculescu; de esta manera, una búsqueda por medio de programas que manejan bases de datos (MySQL), permitió identificar en nuestra biblioteca alrededor de 95% de genes correspondientes al transcriptoma mínimo. El cinco por ciento restante podría ser debido a que el número de etiquetas secuenciadas no fue el óptimo. Se recomienda que para tener un análisis completo deben ser secuenciadas alrededor de 40,000 etiquetas por cada biblioteca, pues en esta cifra es difícil que aparezcan en la tabla de frecuencias nuevas etiquetas (comunicación personal con el creador de la metodología Victor Velculescu).

Uno de los principales problemas que enfrentamos al realizar este trabajo, fue el costo del proceso de secuenciación. En este sentido, para obtener la presente biblioteca de expresión SAGE se hicieron alrededor de 1,500 reacciones de secuenciación. Este importante gasto generó problemas para la construcción de otras bibliotecas; por lo que decidimos emplear otra metodología para definir el transcriptoma completo del carcinoma cervical (anexo 2). De esta manera, elegimos los microarreglos Human Whole Genome CodeLink de General Electric, como una plataforma con la sensibilidad de detectar una copia por célula y la especificidad de distinguir secuencias con alta homología.

b. Microarreglos de Expresión.

El uso de una plataforma de expresión, la cual teóricamente presenta todos los posibles genes en el genoma humano; nos permitió identificar genes diferencialmente expresados en CaCU VPH16+ con respecto a epitelio cervical normal. Algunos de estos genes se han reportado como típicamente involucrados en el proceso de carcinogénesis de manera general ó en CaCU. Ejemplos de estos genes son c-myc, ras, fos, NF-kappaB, EGFR, beta-catenina, etc (61-65)(Tabla 3 anexo 2). Estas evidencias muestran la importancia de emplear herramientas de análisis de alto rendimiento, pues en un solo experimento se analiza un gran número de genes. Los resultados producidos al coincidir con trabajos previos, validan no solo los genes coincidentes con la literatura, sino el grupo de genes identificados por el empleo de herramientas de alto rendimiento, en este caso microarreglos de expresión.

A fin de encontrar un sentido biológico a los genes obtenidos, empleamos una herramienta capaz de producir una imagen del carcinoma cervical basada en las vías metabólicas descritas por KEGG (Enciclopedia Kyoto de Genes y Genomas, por sus siglas en inglés). La base de datos de KEGG presenta redes de proteínas anotadas por evidencias biológicas directas; es decir, todos los datos encontrados empleando esta herramienta tienen sustento en una gran cantidad de artículos científicos en donde se ha estudiado cada vía específica (66). Una evaluación sistemática de las vías alteradas en CaCU se muestra en la Tabla 3 del anexo 2. Algunas de estas rutas tienen un importante papel en el proceso de carcinogénesis en diferentes modelos, además han sido caracterizadas en CaCU. En este contexto, la vía de Wnt se ha propuesto como promotor de la transformación de queratinocitos inducida por VPH (67), además en nuestro grupo hemos observado la activación de esta vía en CCVPH18+ (68).

Otra de las rutas metabólicas que encontramos alteradas en este trabajo, fue la vía de TGF-beta; como se observa en la tabla 3 del anexo 2. Se ha observado que dicha vía es crítica para mantener el control homeostático del crecimiento celular, no solo en células malignas sino además en células que progresan en las etapas tempranas del proceso de los inhibidores de cinasas dependientes de ciclinas p15INK4b y p21CIP1 (revisado en 69). En este sentido, Nees *et al.* han demostrado que las oncoproteínas virales E6 y E7 tienen un efecto negativo en la expresión de genes de respuesta a TGF-beta (46). Las proteínas de la familia Smad juegan un papel clave en la cascada de señalización intracelular de TGF- β ; son las primeras en fosforilarse después que se activa el receptor de TGF- β ; Smad2 y Smad3 se fosforilan y forman complejos heteroméricos con Smad4; estos complejos son translocados al núcleo donde controlan la expresión de genes blanco. Lee y cols. han mostrado que E7 de VPH 16 bloquea la señalización de TGF- β bloqueando la unión del heterómero de proteínas Smad a sus secuencias blanco en la región promotora de los genes que regulan (70). Además, se ha observado la disminución en la expresión de Smad2 y 4 en carcinomas cervicales (71). En este trabajo, hemos mostrado la subexpresión de los genes Smad (Figura 3 del anexo 2). Estas observaciones podrían sustentar la función de la vía metabólica de TGF- β como supresor de tumores en epitelio cervical. De esta manera, la disminución en la expresión de genes involucrados en la vía de señalización de TGF- β promovida por la infección de VPH, puede representar un mecanismo importante que contribuye a la estimulación del crecimiento, immortalización y carcinogénesis del epitelio cervical

La importancia del presente trabajo radica en la caracterización de genes desregulados en CaCU infectado por VPH16. Este tipo viral se detecta en aproximadamente 50% de los

casos de CaCU en el mundo (72), mientras que en México puede detectarse hasta en 70% de la población (73); de esta manera, el presente trabajo contribuye al conocimiento de las alteraciones en cuanto al perfil de expresión en CaCU producidas por la infección del tipo viral más común. Esta caracterización fue posible gracias a la combinación de dos metodologías que permitieron realizar un análisis completo de los genes expresados en epitelio cervical tanto en tejidos sin alteraciones histopatológicas como tumorales.

8. CONCLUSIONES

Este trabajo es el primer esfuerzo para obtener el transcriptoma del epitelio cervical. Esto fue posible gracias a una metodología que permitió analizar de manera completa los genes expresados en este tejido. Aunque nuestra biblioteca SAGE fue producida de un solo donador, la mayoría de las muestras analizadas expresaron los genes seleccionados; lo que muestra la reproducibilidad de este estudio, inclusive con diferentes metodologías.

Esta metodología es compleja y demasiado costosa, principalmente por el gran esfuerzo de secuenciación que se requiere para lograr un catálogo completo de genes expresados. Sin embargo, la gran riqueza de información derivada de este estudio justifica el esfuerzo y permite un mayor conocimiento de la biología y fisiología en términos de los genes expresados en este tejido. En un futuro cercano, estos datos podrían aportar conocimiento en los procesos relacionados a la infección viral u otras patologías que afectan el epitelio cervical.

Estudios preliminares en CaCU empleando herramientas de alto rendimiento para obtener el perfil de expresión en este tejido, han sido limitados principalmente por la cantidad de genes impresos en el arreglo. En este trabajo empleamos herramientas de análisis de expresión que nos permitieron tener un catálogo completo de genes expresados en CaCU; gracias a esto fue posible caracterizar vías celulares en las cuales una proporción importante de genes mostró desregulación en su nivel de expresión.

La importancia de este trabajo es mostrar las vías metabólicas alteradas y los genes que participan en dichas vías, con el objeto de incrementar nuestro conocimiento sobre las alteraciones moleculares inducidas por la infección por VPH16.

9. LITERATURA CITADA.

1. Dirección General de Estadística e Informática/Instituto Nacional de Estadística, Geografía e Informática (INEGI). Estadísticas en Salud, mortalidad. México, D.F.: Secretaría de Salud, 1999.
2. Trujillo-Gutiérrez JJ, Méndez-Gaona JA, Velazco-Contreras ME. **El Cáncer Cérvico Uterino problema viejo, reto actual de la Salud Comunitaria.** *Revista Salud Comunitaria.* IMSS 1998;2(2):14-19.
3. Programa de prevención y control del cáncer cervico-uterino. *Secretaría de Salud, México* 1992
4. Schiffman M, Bauer H, Hoover R. **Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia.** *J Natl Cancer Inst* 1993; **85**: 958-964.
5. Munoz N. **Human papillomavirus and cancer: the epidemiological evidence.** *J Clin Virol* 2000; (1-2):1-5.
6. Walboomers J, Jacobs M, Manos M, Bosch F, Kummer K, Shah K, Snijders P, Peto J, Meijer C, Muñoz N. **Human Papillomavirus is a necessary cause of invasive cervical cancer worldwide.** *J Pathol* 1999; **189**: 12-19.
7. Bosch F, Lorincz A, Muñoz N, Meijer C, Shah K. **The causal relation between human papillomavirus and cervical cancer.** *J Clin Pathol* 2002; **55**:244-265.
8. de Villiers E. **Human pathogenic papillomavirus types: an update.** *Curr Top Microbiol Immunol* 1994;**186**:1-12.
9. zur Hausen. **Papillomaviruses in human cancers.** *Proc Assoc Am Physicians* 1999; **111**:581-587.
10. Bernard HU, Chan SY, Manos MM, Ong CK, Villa LL, Delius H, Peyton CL, Bauer HM,

Wheeler CM. **Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms.** *J Infect Dis* 1994; **170**:1077-85.

11. Munoz N. **Human papillomavirus and cancer: the epidemiological evidence.** *J Clin Virol*; 2000; **1-2**:1-5. Review.

12.- zur Hausen, H. **Papillomavirus infections: a major cause of human cancers.** *Biochem. Biophys. Acta*, 1996; **1288**: F55–F78.

13. Matsukura T, Koi Sumiko, Sugase M. **Both episomal and integrated forms of Human Papillomavirus Type 16 are involved in Invasive Cervical cancers.** *Virology*, 1989; **172**:63-72.

14. Park T-W, Fujiwara H, and Wright TC. **Molecular biology of cervical cancer and its precursors.** *Cancer*, 1995; **76**:1902-1913.

15. Turek LP. **The structure, function and regulation of papillomavirus gene in infection and cervical cancer.** *Adv Viral Res*, 1994; **44**:305-356.

16. Palefsky JM, Holly EA. **Molecular virology and epidemiology of human papillomavirus and cervical cancer.** *Cancer Epidemiol Biomarkers Prev*, 1995; **4**:415-428.

17. Ward P, Coleman DV, and Malcolm DB. **Regulatory mechanisms of the papillomaviruses.** *Trends Genet*, 1989; **5**:97-98.

18. Stevenson M, Hudson LC, Burns JE, Stewart RL, Wells M, Maitland NJ. **Inverse relationship between the expression of the human papillomavirus type 16 transcription factor E2 and virus ADN copy number during the progression of cervical intraepithelial neoplasia.** *J Gen Viro*, 2000; **81**:1825-32.

19. Brown DR, Fan L, Jones J, and Bryan. **Colocalization of human papillomavirus type 11 E1^AE4 and L1 proteins in human foreskin implants grown in athymic mice.**

J. Virology, 1994; **201**:46-54.

20. Doorbar J, Ely S, Sterling J, McLean C, and Crawford L. **Specific interaction between HPV 16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network.** *Nature*, 1999; **1352**:824-827.

21. Hwang ES, Nottoli T, and DiMaio D. **The HPV 16 E5 protein: expression, detection, and stable complex formation with transmembrane proteins in COS cells.** *Virology*, 1995; **211**:227-233.

22. zur Hausen. **Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis.** *J Natl Cancer Inst*, 2000; **92**:690-698.

23. Schwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, et al. **Structure and transcription of human papillomavirus sequences in cervical carcinoma cells.** *Nature*; 1985; **314**:111-114.

24. McDougall JK. **Immortalization and transformation of human cells by human papillomavirus.** *Curr Top Microbiol Immunol*, 1994; **186**:101-19.

25. Munger K, Howley P. **Human Papillomavirus immortalization and transformation functions.** *Virus Res*, 2002;**89**:213-228.

26. Ueda Y, Enomoto T, Miyatake T, Ozaki K, Yoshizaki T, Kanao H, Ueno Y, Nakashima R, Shroyer KR, Murata Y. **Monoclonal expansion with integration of high-risk type human papillomavirus is an essential step for cervical carcinogenesis: association of clonal status and human papillomavirus infection with clinical outcome in cervical intraepithelial neoplasia.** *Lab Invest*, 2003; **83**:1517-1527.

27. Nicolas Wentzensen, Svetlana Vinokurova and Magnus von Knebel Doeberitz. **Systematic Review of Genomic Integration Sites of Human Papillomavirus Genomes in Epithelial Dysplasia and Invasive Cancer of the Female Lower Genital Tract.** *Cancer Research*, 2004; **64**: 3878-3884.

28. Molano M, Van den Brule A, Plummer M, Weiderpass E, Posso H, Arslan A, Meijer

- CJ, Munoz N, Franceschi S; HPV Study Group. **Determinants of clearance of human papillomavirus infections in Colombian women with normal cytology: a population-based, 5-year follow-up study.** *Am J Epidemiol*, 2003; **158**:486-94.
29. Carulli J, Artniger M, Swain , Root C, Chee L, Tulig C, Guerin J, Osborne M, Stein G, Lian J, Lomedico P. **High throughput of differential gene expression.** *J Cell Biochem Supp*, 1998; **30/31**: 286-296.
30. Pastorian K, Hawel L, Byus C. **Optimization of cDNA representational different analysis for the identification of differentially expressed mRNAs.** *Analytical Biochemistry*, 2000; **283**: 89-98
31. Velculescu V and Vogelstein B. **Analysing uncharted transcriptomes with SAGE.** *Trends in Genetics*, 2000; **16**:423-425
32. Argani P, Rosty C, Reiter R, Wilentz R, Murugesan S, Leach S, Ryu B, Skinner H, Goggins M, Jaffee E, Yeo C, Cameron J, Kern S, Hruban R. **Discovery of new markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma.** *Cancer Research*, 2001; **61**: 4320-4324
- 33 Velculescu V. E. *et al.* **Serial Analysis of gene Expression.** *Science*, 1995; **270**: 4484-487
34. Yamamoto M, Wahatsuki T, Hada A, Ryo A. **Use of serial analysis of gene expression (SAGE) technology.** *J Immunol Methods*, 2001; **250**:45-66
35. Margulies EH, Kardia SL, Innis JW. **Identification and prevention of a GC content bias in SAGE libraries.** *Nucleic Acids Res*, 2000; **15**:60-70
36. Polyak K, *et al.* **Gene discovery using the serial analysis of gene expression technique: implications for cancer research.** *Journal of Clinical Oncology*, 2001; **19**:2948-2958.
37. Zhang L, Zhou W, Velculescu V, Kern S, Hruban R, Hamilton S, Vogelstein B, Kinzler K. **Gene normal expression profiles in normal and cancer cells.** *Science*, 1997; **276**:

1268-1272

38. Zhou W, Sokoll LJ, Bruzek DJ, et al. **Identifying markers for pancreatic cancer by gene expression analysis.** *Cancer Epidemiol Biomarkers Prev*, 1998; 7:109-112,
39. Sasaki M, Nakahira K, Kawano Y, Katakura H, Yoshimine T, Shimizu K, Kim S, Ikenaka K. **MAGE-E1, a new member of the melanoma-associated antigen gene family and its expression in human glioma.** *Cancer Res*, 2001; 61:4809-4814
40. Polyak K, et al. **HIN-1, a putative cytokine highly expressed in normal but not cancerous mammary epithelial cells.** *Proc Natl Acad Sci*, 2001; 98:9796-9801.
41. Polyak K, et al. **A SAGE (serial analysis of gene expression) view of breast cancer tumor progression.** *Cancer Research*, 2001; 61:5697-5702.
42. Yu, J et al. **Identification and classification of p53 regulated genes.** *Proc. Natl. Acad. Sci*, 1999; 96:14517-14522
43. www.ncbi.nlm.nih.gov/SAGE.
44. Velculescu V. E. et al. **Analysis of human transcriptomes.** *Nature Genetics*, 1999; 23: 387-388
45. Shim C, Zhang W, Rhee CH, Lee JH,. **Profiling of differentially expressed genes in human primary cervical cancer by complementary ADN expression array.** *Clin Can Res*, 1998; 4: 3045-3050.
46. Nees M, Geoghegan JM, Munson P, Prabhu V, Liu Y, Androphy E, Woodworth CD. **Human papillomavirus type 16 E6 and E7 proteins inhibit differentiation-dependent expression of transforming growth factor-beta2 in cervical keratinocytes.** *Cancer Res*, 2000; 60:4289-98.
47. Cheng Q, Lau WM, Tay SK, Chew SH, Ho TH, Hui KM. **Identification and characterization of genes involved in the carcinogenesis of human squamous cell cervical carcinoma.** *Int J Cancer*, 2002; 98:419-26.
48. Berger AJ, Baege A, Guillemette T, Deeds J, Meyer R, Disbrow G, Schlegel R,

Schlegel R. **Insulin-like growth factor-binding protein 3 expression increases during immortalization of cervical keratinocytes by human papillomavirus type 16 E6 and E7 proteins.** *Am J Pathol*, 2002;161:603-10.

49. Alazawi W, Pett M, Arch B, Scott L, Freeman T, Stanley MA, Coleman N. **Changes in cervical keratinocyte gene expression associated with integration of human papillomavirus 16.** *Cancer Res*, 2002;62:6959-65.

50. Wong YF, Sahota DS, Cheung TH et al. **Gene expression pattern associated with radiotherapy sensitivity in cervical cancer.** *CancerJ*, 2006;12:189-193.

51. Kitahara O, Katagiri T, Tsunoda T et al. **Classification of sensitivity or resistance of cervical cancers to ionizing radiation according to expression profiles of 62 genes selected by cDNA microarray analysis.** *Neoplasia*, 2002;4:295-303.

52. Wong YF, Selvanayagam ZE, Wei N et al. **Expression genomics of cervical cancer: molecular classification and prediction of radiotherapy response by ADN microarray.** *Clin Cancer Res*, 2003; 9:5486-5492.

53. Ramakrishnan R, Dorris D, Lublinsky A, Nguyen A, Domanus M, Prokhorova A, Gieser L, Touma E, Lockner R, Tata M, Zhu X, Patterson M, Shippy R, Sendera TJ, Mazumder A: **An assessment of Motorola CodeLink microarray performance for gene expression profiling applications.** *Nucleic Acids Res* 2002, 30:e30.

54. Shippy R, Sendera T, Lockner R, Palaniappan C, Kaysser-Kranich T, Watts G, Alsobrook J. **Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations.** *BMC Genomics* 2004, 5:61.

55. Dorris D, Nguyen A, Gieser L, Lockner R, Lublinsky A, Patterson M, Touma E, Sendera T, Elghanian R, Mazumder A. **Oligodeoxyribonucleotide probe accessibility on a three-dimensional ADN microarray surface and the effect of hybridization time on the accuracy of expression ratios.** *BMC Biotechnology* 2003, 3:6.

56. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci U S A* 1998, **95**:14863-14868.
57. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A*, 2001, **98**:5116-5121
58. Khatri P, Sellamuthu S, Malhotra P, Amin K, Done A, Draghici S: **Recent additions and improvements to the Onto-Tools.Nucleic.** *Nucleic Acids Res* 2005, **33**:W762-W765.
59. <http://vortex.cs.wayne.edu/Projects.html>
60. Marenholz I, Volz A, Ziegler A, Davies A, Ragoussis I, Korge BP, Mischke D. **Genetic analysis of the epidermal differentiation complex (EDC) on human chromosome 1q21: chromosomal orientation, new markers, and a 6-Mb YAC contig.** *Genomics* 1996, **37**:295-302.
61. Ocadiz R, Saucedo R, Cruz M, Graef AM, Gariglio P: **High correlation between molecular alterations of the c-myc oncogene and carcinoma of the uterine cervix.** *Cancer Res*, 1987; **47**:4173-7
62. Agnantis NJ, Spandidos DA, Mahera H, Parissi P, Kakkanas A, Pintzas A, Papacharalampous NX: **Immunohistochemical study of ras oncogene expression in endometrial and cervical human lesions.** *Eur J Gynaecol Oncol* 1988, **9**:360-365.
63. Cheung TH, Leung JO, Chung TK, Lam SK, To KF, Wong YF: **c-fos overexpression is associated with the pathoneogenesis of invasive cervical cancer.** *Gynecol Obstet Invest* 1997, **43**:200-3
64. Nair A, Venkatraman M, Maliekal TT, Nair B, Karunakaran D: **NF-kappaB is constitutively activated in high-grade squamous intraepithelial lesions and squamous cell carcinomas of the human uterine cervix.** *Oncogene* 2003, **22**:50-58
65. Xu YH, Richert N, Ito S, Merlino GT, Pastan I: **Characterization of epidermal growth factor receptor gene expression in malignant and normal human cell lines.** *Proc Natl Acad Sci U S A*, 1984; **81**:7308-7312.

66. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucleic Acids Res* 2004, **32**:D277-D280.
67. Uren A, Fallen S, Yuan H, Usubutun A, Kucukali T, Schlegel R, Toretzky JA: **Activation of the canonical Wnt pathway during genital keratinocyte transformation: a model for cervical cancer progression.** *Cancer Res* 2005, **65**:6199-6206.
68. Vazquez-Ortiz G: **Patrón de expresión de células no neoplásicas y células invasoras del cervix humano.** *Tesis de Doctorado.* Universidad nacional Autónoma de México, Facultad de Medicina; 2005.
69. Roberts AB, Wakefield LM: **The two faces of transforming growth factor β in carcinogenesis.** *Proc Natl Acad Sci U S A* 2003, **100**: 8621-8623
70. Lee DK, Kim BC, Kim IY, Cho EA, Satterwhite DJ, Kim SJ: **The human papilloma virus E7 oncoprotein inhibits transforming growth factor-beta signaling by blocking binding of the Smad complex to its target sequence.** *J Biol Chem* 2002, **277**:38557-38564
71. Maliekal TT, Antony ML, Nair A, Paulmurugan R, Karunagaran D: **Loss of expression, and mutations of Smad 2 and Smad 4 in human cervical cancer.** *Oncogene* 2003, **22**:4889-4897.
72. Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, Schiffman MH, Moreno V, Kurman R, Shah KV: **Prevalence of human papillomavirus in cervical cancer: a worldwide perspective.** International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst* 1995, **87**:796-802.
73. Pina-Sanchez P, Hernandez-Hernandez DM, Lopez-Romero R, Vazquez-Ortiz G, Perez-Plasencia C, Lizano-Soberon M, Gonzalez-Sanchez JL, Cruz-Talonia F, Salcedo M. **Human papillomavirus-specific viral types are common in Mexican women affected by cervical lesions.** *Int J Gynecol Cancer* 2006;**16**:1041-7.