UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

"ANÁLISIS *IN SILICO* DEL TRANSCRIPTOMAS DEL EPITELIO CERVICAL NORMAL"

T E S I S

QUE PARA OBTENER EL TITULO DE:
LICENCIADO EN BIOLOGÍA

P R E S E N T A :

HUGO ARREOLA DE LA CRUZ

ASESOR:
Dr. MAURICIO SALCEDO VARGA

MÉXICO, D. F. 2004

AGRADECIMIENTOS

A mis Padres, por su amor tan incondicional, y por brindarme su apoyo y las herramientas necesarias para la realización de esta meta en mi vida, GRACIAS.

A mis hermanos, Servando y Eloisa; por su amor y apoyo, ya que en los momentos más difíciles siempre han estado conmigo.

A ti Irina, por ser la mujer que eres, ya que eres un pilar importante en mi vida, ya que siempre en tenido tu apoyo para seguir adelante, esperando toda mi vida contar con el.

Al Dr. Mauricio Salcedo, por su gran apoyo y comprensión, no me queda mas que agradecerle por la gran persona que es en todo sentido, GRACIAS.

A ti Carlos, por esas jornadas en el laboratorio, y por ser una gran persona.

A mis grandes amigos de la Universidad (Alberto, Alya, Elena, Eri ca, Gerardo, Miran, Mata y Sandra), por esos grandes momento que vivimos, y por soportarme. Espero seguir contando con ustedes.

A los chavos del laboratorio, por formar el grupo que somos, y por ser tan ameno el trabajo.

A la Familia Reyes González, por ser quienes son conmigo, y por soportarme.

ÍNDICE

RESUMEN

El Cáncer Cérvico -uterino (CaCu) se ha visto fuertemente asociado a la infección con Virus de Papiloma Humano (VPH), un factor sexualmente transmisible. Datos de la Organización Mundial de la Salud (OMS) muestran que la frecuencia anual de mortali dad por CaCu, es de 237,000 casos. En México, de acuerdo con la Secretaria de Salud, el CaCu es la neoplasia maligna con mayor incidencia (32.5%) entre la población femenina. Para entender los cambios moleculares que sufre un tejido normal, con respecto a su contraparte transformada, primeramente se necesita conocer cómo está organizado el tejido normal; es en este sentido que por medio del Análisis en Serie de Expresión Génica (SAGE), nos permite evaluar transcriptomas de una manera cuantitativa. Esta pod erosa herramienta ha sido usada básicamente para comparar tejidos tumorales, así como de su contraparte normal. En este trabajo se utilizo la biblioteca de expresión de Cérvix Normal, para su caracterización de manera preliminar por medio del análisis *ins ilico*, ya que no hay nada descrito en Cérvix en términos de expresión diferencial.

INTRODUCCIÓN

Características Generales del Cuello del Útero Normal.

El Cérvix o **Cuellouterino** , es la parte inferior del útero que ocupa aproximadamente la tercera parte de este órgano. Constituye el canal de comunicación del cuerpo uterino con la vagina de la mujer. El conducto cervical mide aproximadamente de 2.5 a 3.0 cm y mantiene la secreción de moco cervical de manera continua [1] (Fig.1).

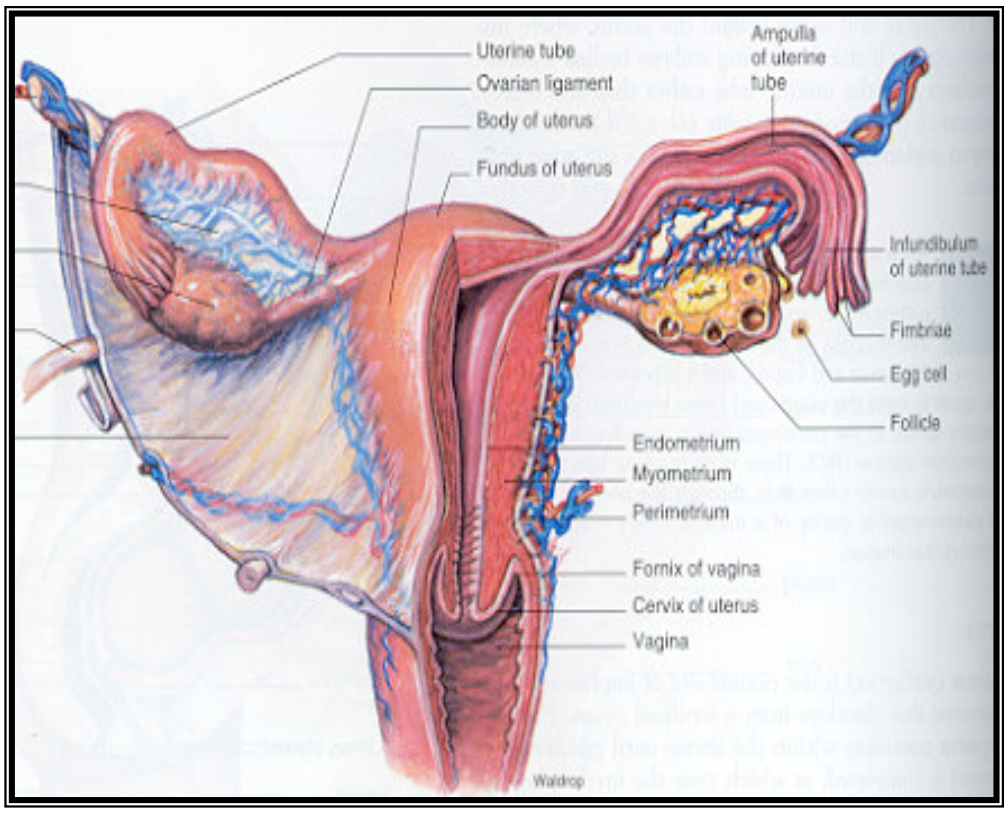


Fig. 1. Localización anatómica del Cérvix.

La superficie más expuesta del cérvix y que queda a la luz de la vagina es conocida como exocérvix, y la porción del cérvix en relación al canal endocervical es conocida como el endocérvix. El exocérvix está

revestido de un epitelio escamoso estratificado que tiene una doble función; en primera, segrega y acumula cantidades variables de glicógeno en su estrato superior en respuesta a la circulación de estrógeno, especialmente estradiol; y, en segundo lugar, es un tejido protector debido a su superficie cornificada.

El epitelio escamoso cervical está dividido según Dierks en 5 estratos:

- Estrato de células basales : son células pequeñas, cilíndricas, con escaso citoplasma y núcleo relativamente grande de forma oval con cromatina densa; estas células son usualmente inactivas mitóticamente.

- Estrato de células parabasales o espinosas: son células poliédricas con núcleos bastante grandes y puentes intercelulares evidentes, son mas largas que las basales por su incremento de citoplasma, su núcleo tiene ligeramente menos cromatina densa y la actividad mitótica está presente. En ocasiones estas células pueden ser vacuoladas.

- Estrato de la zona media : las células son más grandes, planas, con citoplasma abundante, muy vacuoladas con abundante glucógeno y cantidades variables de inclusiones granulares, estas células son conocidas como células intermedias.

- Estrato granuloso: es inconstante e irreconocible, consta de una banda estrecha de células aplanadas que se tiñen intensamente y contienen gránulos de queratohialina.

- Estrato superficial: consta de células alargadas y aplanadas con núcleos pequeños redondos y picnóticos y un citoplasma abundante[2] (Fig.2).

El epitelio endocervical es mono estratificado cilíndrico y mucho más alto que el epitelio del endometrio . El endocérvix esta compuesto por un solo estrato de epitelio secretor de mucina ; sus núcleos son compactos,

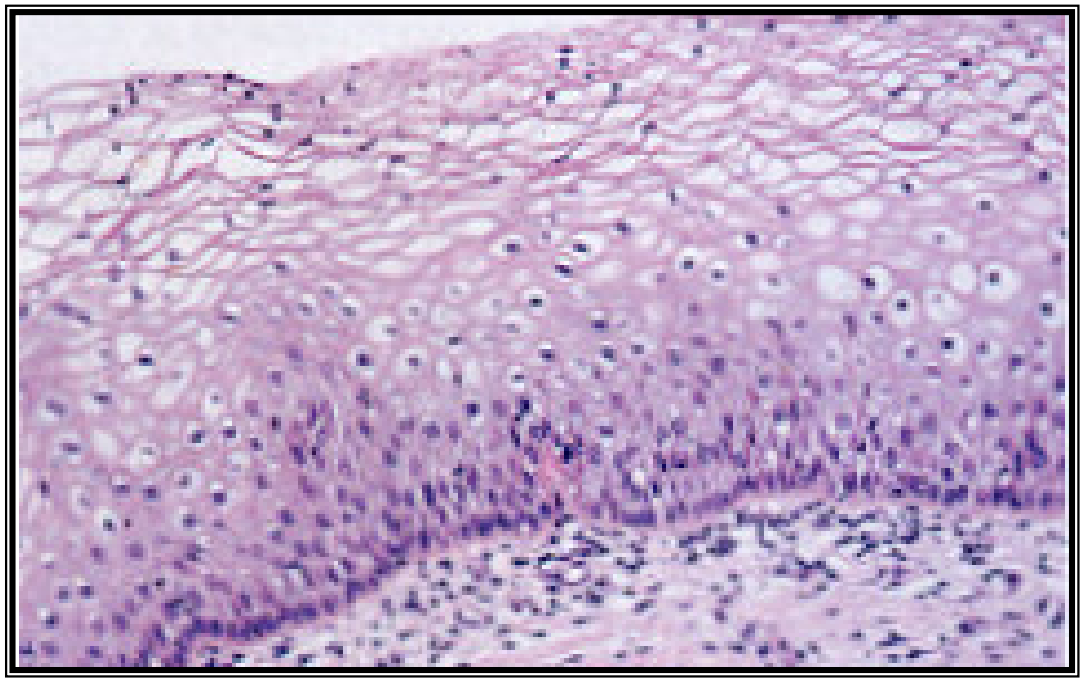


Fig. 2. Epitelio escamoso cervical.

pequeños y bien teñidos y se sitúan en el polo basal. Las células miden de 20 a 35 micras de altura y de 5 a 9 micras de anchura. Estas células segregan mucina. Por su forma cilíndrica y alargada se les ha llamado también células en empalizada [1].

La unión del epitelio cilíndrico con el epitelio escamoso del exocérvix es brusca, encontrándose los dos tipos de epitelio totalmente diferenciados en un punto definido, que se conoce con el nombre de unión

escamocilíndrica. Aunqueestonose observa, entodosloscasos, enla mayoría de ellos existe una zona transicional, que separa los dos epitelios. Estos dos tipos no son constantes en la totalidad de la circunferencia del conducto cervical [2]. Esta unión escamocilíndrica es donde se presentan la mayoría de las enfermedades epiteliales del cérvixenlamujer,comoeselcasodelCáncerCérvico -uterino(CaCu)o cáncer del cuellodelamatriz, el cual representaunode losmayores problemas de salud en la mujer en el ámbito mundial. Datos de la Organización Mundial de la Salud (OMS) muestran que la frecuencia anual de mortalidad por CaCu, es de 237,000 casos [3]. En México, de acuerdoconlaSecretariadeSalud,elCaCues laneoplasiamalignacon mayorincidencia (32.5%)entrela poblaciónfemenina [4].

Asimismo, uno de los principales agentes etiológicos asociados al desarrollo del CaCu es la infección por Virus de Papiloma Humano (VPH),unfactorsexualmentetransmisible. Estudiosrecientes,utilizando metodologías muy sensibles de detección, han confirmado una positividadaVPHenel99.7%delosdecasosdeCaCu [5]. ElVPHinducela transformacióncelularpormediodelainhibicióndelosmecanismosde regulación del ciclo celular, modificando el perfil global de expresión genética,quetraecomoconsecuencia alteracionesdelfenotipocelular.

LosVirusdelPapilomaHumanopertenecenalafamiliaPapillomaviridae, y comprenden alrededor de 120 tipos distintos [6]. Son epiteli otrópicos, conunadoblecadenadeDNAcircularenvueltoenunacubiertaprotéica llamadacápside. Sugenomasecomponededosgruposfuncionalesde marcosdelecturaabiertos,llamadostempranos(E),ytardíos(L);están separados por una Región Larga de Control (LCR), que es necesaria paralareplicaciónviral normal yel control delareplicacióndegenes [7]. En la región temprana se encuentran genes que participan en la

replicación viral y la inmortalización celular, por ejemplo E6 y E7. La principal función biológica de E6 hasta la fecha parece ser la inactivación de la proteína supresora de tumor p53. Por su parte, E7 es capaz de formar complejos con varias proteínas celulares, incluyendo a la proteína supresora de tumor retinoblastoma (pRb). Finalmente, los genes de la región tardía codifican para proteínas de los capsómeros [8] (Fig. 3). En general, los tumores sufren una continua acumulación de cambios genéticos y epigenéticos que les permiten escapar de los controles de regulación celular y ambiental.

Se han desarrollado diferentes metodologías para caracterizar el perfil de expresión global en una célula o tejido 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 [10].

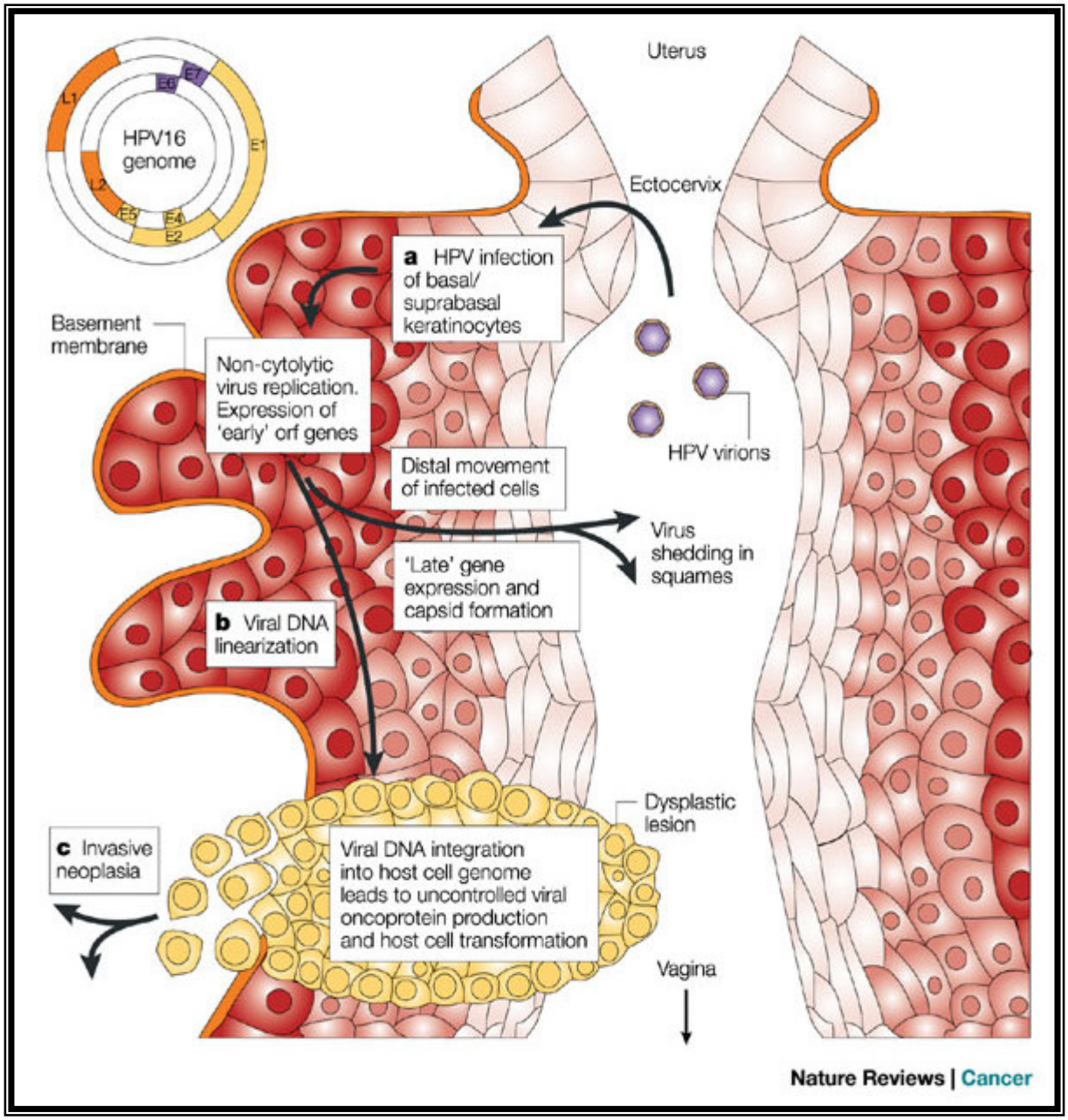


Fig. 3. Esquema que representa la biología de la infección de VPH[9].

ANTECEDENTES

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 nucléicos 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[11]. 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 [12]. Se han empleado otros métodos como las hileras de cDNA u oligonucleótidos (Microarrays); aún cuando puedan comparar miles de genes, están limitados por el uso de transcritos previamente caracterizados y/o de secuencias conocidas.

En este sentido, el análisis en serie de la expresión génica (SAGE), que es una plataforma metodológica que involucra herramientas de biología molecular como de bio informática, nos permite el análisis amplio y cuantitativo del perfil de expresión celular. Gracias a esta herramienta se conoce que un grupo de genes (alrededor de 40) se sobreexpresan en diferentes neoplasias (por ejemplo, carcinoma colorectal, mama, melanoma, etc.), es decir, el transcriptoma mínimo.

Con esta plataforma de análisis global, los transcritos (RNAm) son convertidos a cDNA, después digeridos en fragmentos pequeños o

etiquetas (tags de 10 -14 pb) por enzimas de restricción específicas [13]. SAGE está basado en dos principios: la representación de los transcritos expresados por secuencias cortas de cDNA o etiquetas y la concatenación de estas etiquetas para su clonación y análisis mediante la secuenciación. A partir del borrador del genoma humano , se sabe que éste consta de 30,000 a 40,000 genes; una secuencia de 10 pares de bases podría identificar a más de 1X10 [6] transcritos diferentes (4 [10]). Por lo que el tamaño de estos tags es suficiente para discriminar todos los transcritos presentes de un a célula en un momento dado [14]. Con esta plataforma metodológica se pueden detectar y cuantificar simultáneamente los niveles de expresión génica de una célula en un momento dado; posee la sensibilidad de detectar genes expresados en bajos niveles, así como identificar genes nuevos, debido a que el análisis se desarrolla utilizando la totalidad de los mRNA de la muestra a estudiar, es decir SAGE es un sistema de análisis abierto, a diferencia de otras herramientas como son las microhileras de tejidos en los cuales se debe tener un previo conocimiento de los genes objeto de estudio. En suma, SAGE puede definir la historia celular en función de los niveles de expresión génica de un linaje específico o de una entidad patológica dada [15]. Esta herramienta se ha em pleado en la comprensión del proceso que lleva una célula normal a su contraparte neoplásica, principalmente en tres aspectos:

- 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 estudio nos permite comprender cuáles son las vías de regulación en los procesos involucrados en la tumorigénesis; asimismo, nos permite la identificación de nuevos marcadores de pronóstico y diagnóstico. En los estudios realizados por medio de esta plataforma, se ha encontrado concordancia en los resultados, validando de esta forma la utilidad del gen identificado como posible marcador de pronóstico y/o diagnóstico[16]. Un ejemplo de esto, empleando SAGE, es el estudio que valida o reafirma el gran potencial del análisis de Bibliotecas SAGE que 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[17]. Con SAGE se pueden identificar las diferencias entre un tejido neoplásico y su contraparte normal; en cáncer de mama se identificó una etiqueta que no se relacionaba con ningún gen o EST (fragmento de secuencia expresado), 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 dicho tag (que se le dio el nombre de HIN-1 a este nuevo gen)[18]. Al aumentar el número de casos analizados específicamente para la expresión de HIN-1, se observó en sólo 21% de los tejidos neoplásicos estudiados. Al clonar y expresar el cDNA de HIN-1 se observó que es una proteína secretada que regula negativamente el crecimiento celular de líneas celulares derivadas de cáncer de mama[19]. 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 en 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 diferencia de 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 indujeron la función de p53 en una línea celular con p53 deletado, encontraron que 34 tags diferían en sus tasas de expresión, de los cuales sólo tres han sido previamente identificados como genes regulados por p53 [20].

SAGE ha sido propuesto como una herramienta analítica estratégica en el Proyecto de la Anatomía del Cáncer (CGAP), y se ha creado una base de datos con tags producidos por esta estrategia. A la fecha hay más de tres millones de tags de 88 diferentes librerías derivados de 19 tejidos, tanto normales como tumorales (www.ncbi.nlm.nih.gov/SAGE). 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 [21,22].

DESARROLLO DEL ANÁLISIS EN SERIE DE EXPRESIÓN GÉNICA (SAGE)

- Obtención y purificación de RNA total (5 µg) a partir de material biológico.

- Por medio de partículas paramagnéticas acopladas a oligo (dT) biotinilado, se aisló el mRNA.

- Para la síntesis de la primera cadena de cDNA se utilizó la enzima transcriptasa Reversa SuperScript; para la síntesis de la segunda cadena de cDNA se realiza una digestión parcial con la enzima RNAsa H, y la síntesis se lleva a cabo con las enzimas DNA ligasa y la DNA polimerasa.

- Se efectúa una digestión con la enzima Nla III, que reconoce la secuencia palíndrome CATG, por lo que genera extremos cohesivos, y corta aproximadamente cada 250 pares de bases.

- Una vez realizada la digestión con Nla III, se divide la población de cDNA en dos fracciones (A y B), esto con la finalidad de ligarlos a dos adaptadores sintéticos (aproximadamente de 40 pb cada uno), que tienen extremos cohesivos complementarios a la secuencia CATG. Posteriormente para generar los tags; al final de la secue ncia del adaptador se encuentra un sitio de reconocimiento para la enzima BsmF I, que es unas enzima de tipo II, por lo que su sitio de reconocimiento se encuentra alejado de su sitio de corte; para esta enzima hasta 14 pb, y de esta forma se forman los t ags.

- Una vez que fueron ligados los tags a los adaptadores, lo siguiente es unir los ditags, es decir, tag -tag, por lo que los adaptadores tienen una modificación en su extremo 3', un grupo amino unido por medio de un enlace covalente, por lo que no se pu ede unir adaptador con adaptador, de esta forma nos asegura que nuestra ligación es tag - tag.

- Se hace la amplificación de los ditags por medio de reacciones de PCR, que para conocer el número óptimo de reacciones a fin de tener el suficiente número de di tags, se hace una titulación con diferentes diluciones del producto de PCR, utilizando como control DNA de concentración conocida, ya que en los subsecuentes pasos de purificación se pierde una cantidad considerable de material.

- Después de realizadas la s reacciones de PCR, las reacciones se concentran en un solo tubo, el DNA presente se purifica empleando Fenol/Cloroformo y se precipitan por medio de Acetato de Amonio. Posteriormente las muestras son sujetas a un gel preparativo de poliacrilamida al 8%, y se cortan las bandas correspondientes a los ditags, que son las de 100 pb aproximadamente, las otras bandas presentes corresponden a los adaptadores y a digestiones incompletas. Finalmente, el DNA correspondiente a los ditags es cortado y eluído.

- Nuevamente se hace una digestión con la enzima NlaIII, para liberar los ditags de los adaptadores, ya que los adaptadores tienen un extremo cohesivo complementario a NlaIII. Los ditags liberados se corren en un gel preparativo de poliacrilamida al 12% para co rtar la banda correspondiente a los ditags, para su posterior concatamerización.

- Por medio de la enzima T4 DNA ligasa fue posible concatenar los ditags de 26 pb. Cada concatámero contiene de 30 – 50 etiquetas. Para tener diferentes poblaciones de los conc atámeros, se corrió un gel preparativo de poliacrilamida al 8%, de donde se seleccionaron tres regiones; de 300 a 500 pb, de 500 a 800 pb y de 800 a 1000 pb. Cada región se clona en un vector de expresión (pZero).

- Posteriormente se electropora na una cep a E. coli TOP10.

  - Las bacterias recombinantes se crecen en placas con 50 μg de Zeocina. Las colonias crecidas se pican y se analizan por medio de PCR empleando iniciadores específicos para PUC M13, aquellas colonias con insertos mayores a 500 pb, (lo que equivale a un inserto mayor a 25 tags) se purifican y se someten a reacción de marcaje con Big Dye (Perkin Elmer) y analizadas en un secuenciador de Applied biosistems modelo 310. Los resultados obtenidos servirán para alimentar el software de análisis SAGE 2000 (Johns Hopkins University); que reconoce los signos de puntuación CATG generados por NlaIII, extrayendo la secuencia de cada tag. De esta manera, se produce una base de datos en Excel que muestra las secuencias de 10 pb que corresponde a cada tag (Fig.4).

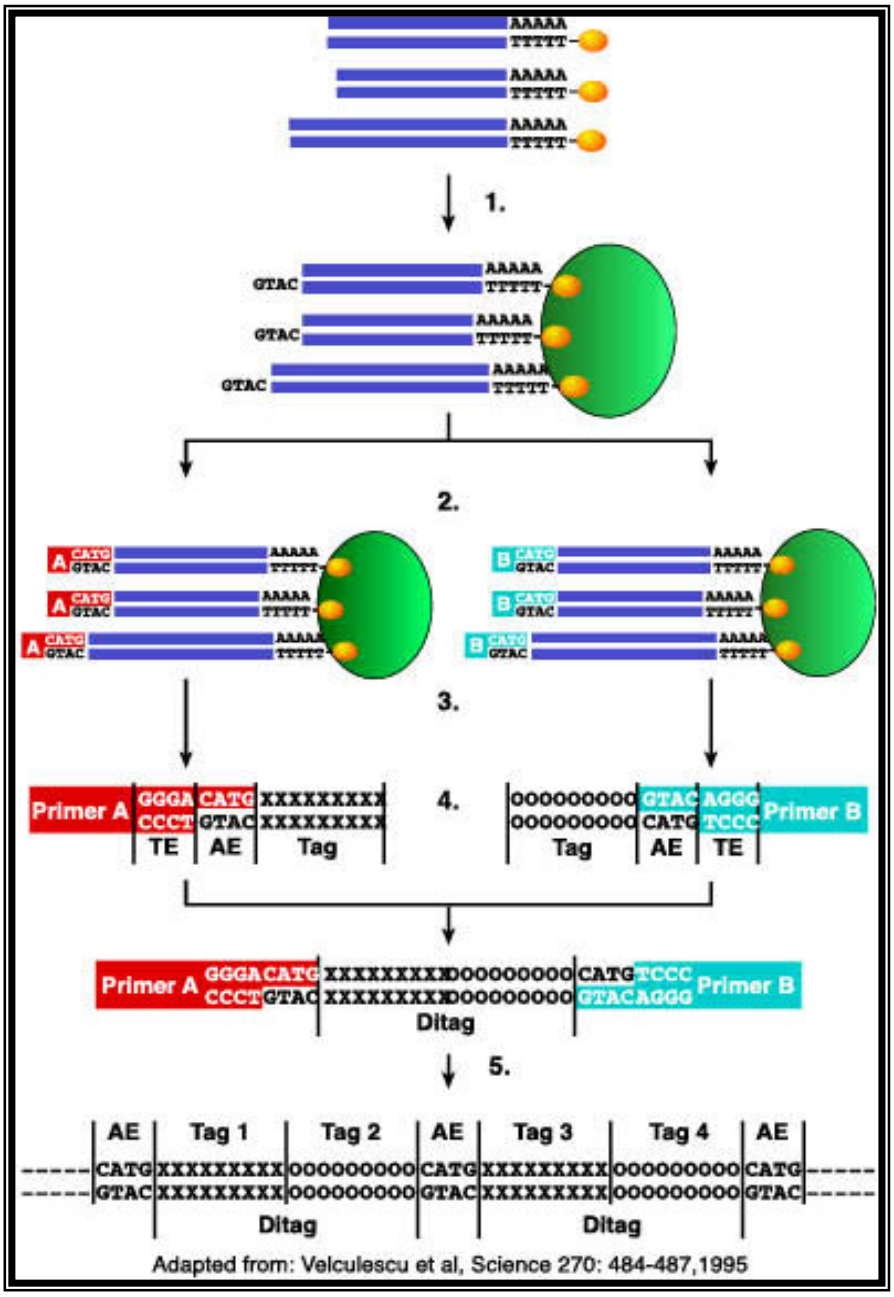


Fig. 4. Proceso experimental del Análisis en Serie de Expresión Génica (SAGE).

La gran cantidad de datos de secuencias disponibles en las bases de datos públicas plantean nuevos retos para los investigadores. En la era post-Genómica, la interpretación de los datos y el desarrollo de modelos que informen del fenómeno biológico específico constituye una prioridad. Hoy en día, a raíz de la gran cantidad de datos generados ya a partir del proyecto del Genoma Humano, la interpretación de esos datos en términos biológicos es de gran impo rtancia; es aquí donde el papel de la bioinformática juega un papel fundamental en la elaboración de modelos matemáticos que puedan dar una interpretación biológica de esa gran cantidad de datos. En este contexto, es de gran importancia la caracterización de los transcriptomas de un tejido normal y de su contraparte transformada, así como identificar los genes que se están expresando de manera diferencial [23].

A la fecha hay una cantidad limitada de análisis *in silico* de bibliotecas de expresión SAGE que h an arrojado resultados muy interesantes en cuanto a la identificación de genes específicos; existe, por ejemplo, un estudio en el cual se describen tres nuevos genes expresados exclusivamente en próstata por medio de la substracción *in silico* de bibliotecas de EST´s [24]. Diez candidatos a genes específicos de granulocitos han sido identificados por el análisis extenso de secuencias de bibliotecas de cDNA derivadas de granulocitos y de otros once tejidos, es decir, una línea celular de hepatocitos, hígado fe tal, hígado de adulto, grasa visceral, pulmón, mucosa colónica, keratinocitos, cornea y retina [25].

JUSTIFICACIÓN

Hasta el momento no hay nada descrito en Cérvix en términos de expresión diferencial, la comparación de esta biblioteca con respecto de otras bibliotecas de expresión SAGE publicadas en la red, permitirá distinguir genes específicos para este tejido.

Para entender los cambios moleculares que sufre un tejido normal, con respecto a su contraparte transformada, primeramente se necesita conocer cómo está organizado el tejido normal, para poder entender los cambios que originan el fenotipo transformado y la caracterización de perfiles globales de expresión que pueden ayudar a explicar procesos biológicos importantes en las células normales. Este tipo de estudios ayuda a comprender las bases moleculares de los fenotipos ya caracterizados, y así poder prevenir el desarrollo del CaCu.

OBJETIVO GENERAL

Caracterizar por análisis *in silico* el perfil de expresión global de la biblioteca de expresión SAGE del epitelio cervical normal.

OBJETIVOS PARTICULARES:

- Caracterizar de manera preliminar *in silico* a nivel funcional, la biblioteca de expresión del epitelio cervical normal.

- Mapear físicamente en su posición cromosómica los genes expresados.

MATERIAL Y METODOS

La biblioteca de expresión de Cérvix uterino normal, fue elaborada en el Laboratorio de Oncología Genómica, a cargo del Dr. Mauricio Salcedo Vargas, de la Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, Centro Médico Nacional Siglo XXI, IMSS, en colaboración con el Dr. Gregory J. Riggins, Director del proyecto de la Anatomía de la Célula Cancerosa del Centro Nacional de Biotecnología (NCBI) de los Estados Unidos. La biblioteca de expresión se encuentra disponible en un sitio público en la red del NIH (Instituto Nacional de la Salud); SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE/), en el cual se encuentra una gran cantidad de bibliotecas de expresión tanto tumorales como normales. De cada biblioteca se da una breve descripción del origen del tejido, título de la biblioteca de expresión, la cuenta total de tags, el nombre de los autores, fecha en que se publicó, entre otros; de una serie de herramientas para el análisis de las bibliotecas (Fig. 5).

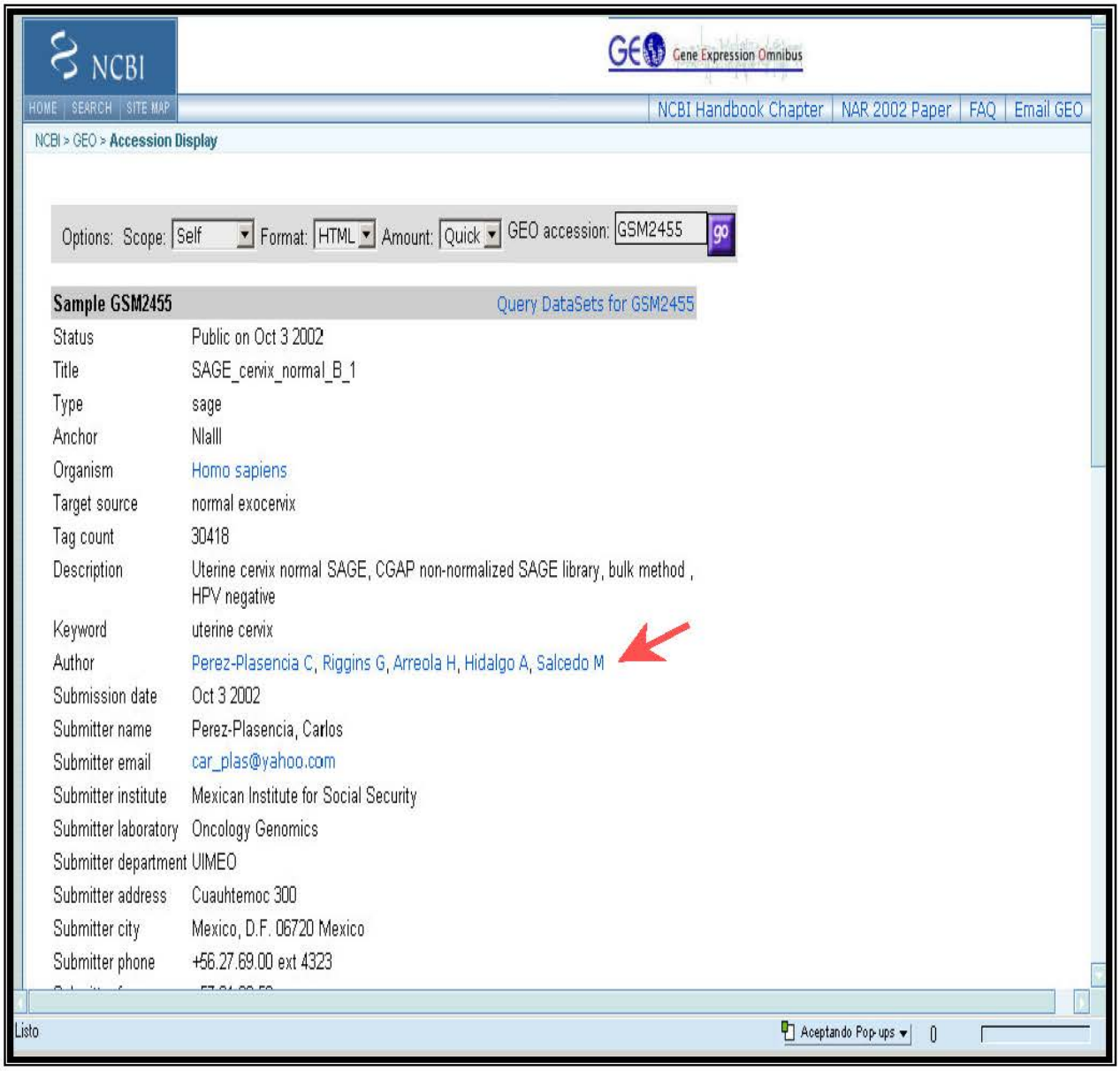


Fig. 5. Descripción de la biblioteca de expresión.

Una vez que la biblioteca de expresión se descargó de la red, los datos se cargaron al programa de MS Access con la lista de actualizaciones de unigene para que de esta manera se pudiera tener la lista de los genes totales presentes en la librería así como el nombre del gen, función y su número de identificación para sus análisis en las bases de datos en internet. Una vez realizada la consulta en el programa de MS Access, la tabla resultante se exportó al progr amade Excel, que contiene los datos de los tags, nombre de los genes, así como su función de toda la biblioteca de expresión, etc.

FUNCIÓN BIOLÓGICA

Para tener una relación de la biblioteca de expresión en cuanto a su función biológica, se utilizó o tra herramienta bioinformática (FATIGO) en la red (http://fatigo.bioinfo.cnio.es) (Fig. 6). Inicialmente esta herramienta fue desarrollada para permitir la explotación de datos resultantes de microarreglos, co n el objeto de identificar las rutas metabólicas que se alteran entre dos o más estadios diferentes. Además, esta herramienta nos permite el análisis de una gran cantidad de datos (genes), en cuanto a su proceso biológico, definiendo el organismo o la base de datos a analizar. Asimismo, FATIGO se puede emplear para comparar listas de genes y encontrar cambios diferencialmente representadas entre ellas.

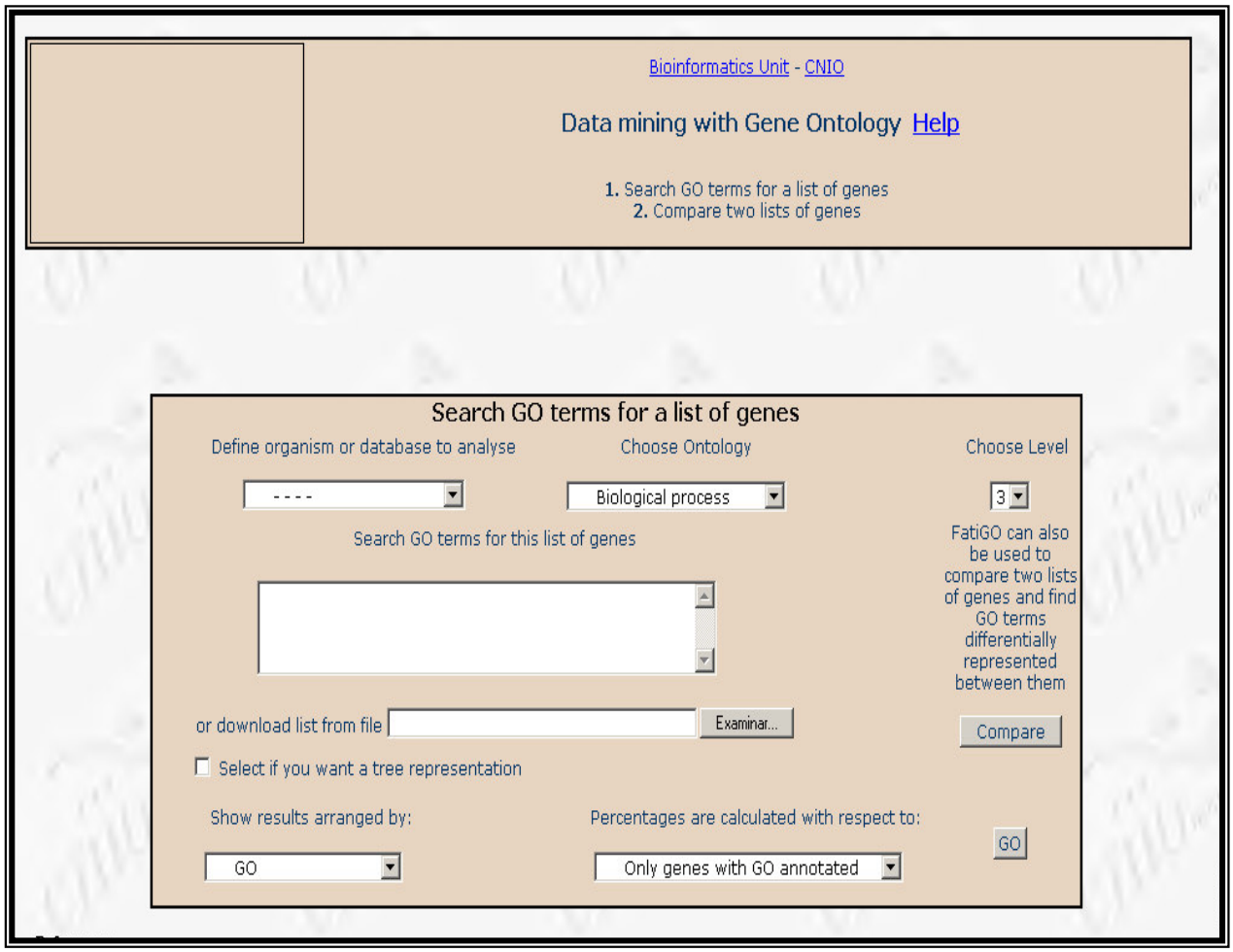


Fig. 6. Página web (Gene Ontology), herramienta utilizada para el análisis de una gran cantidad de datos (genes), en términos de su proceso biológico.

Para poder comparar los resultados con bases de datos de EST´s, se realizó la búsqueda en, la fuente de SOURCE; la cual fue desarrollada por la Universidad de Stanford (http://source.stanford.edu/cgi-bin/sourceSearch)[26], y así corroborar los resultados anteriores de Gene Ontology (Fig. 7). Esta herramienta unificada, compila y colecta datos procedentes de una gran variedad de bases de datos científicas, encapsulando la biología molecular de miles de genes derivados de Homo Sapiens, Ratus Norvertus, entre otros organismos. Asimismo, permite generar tablas en donde el usuario programa la obtención de los datos requeridos de algún gen ó genes; por ejemplo, datos de expresión, papel biológico, etc.

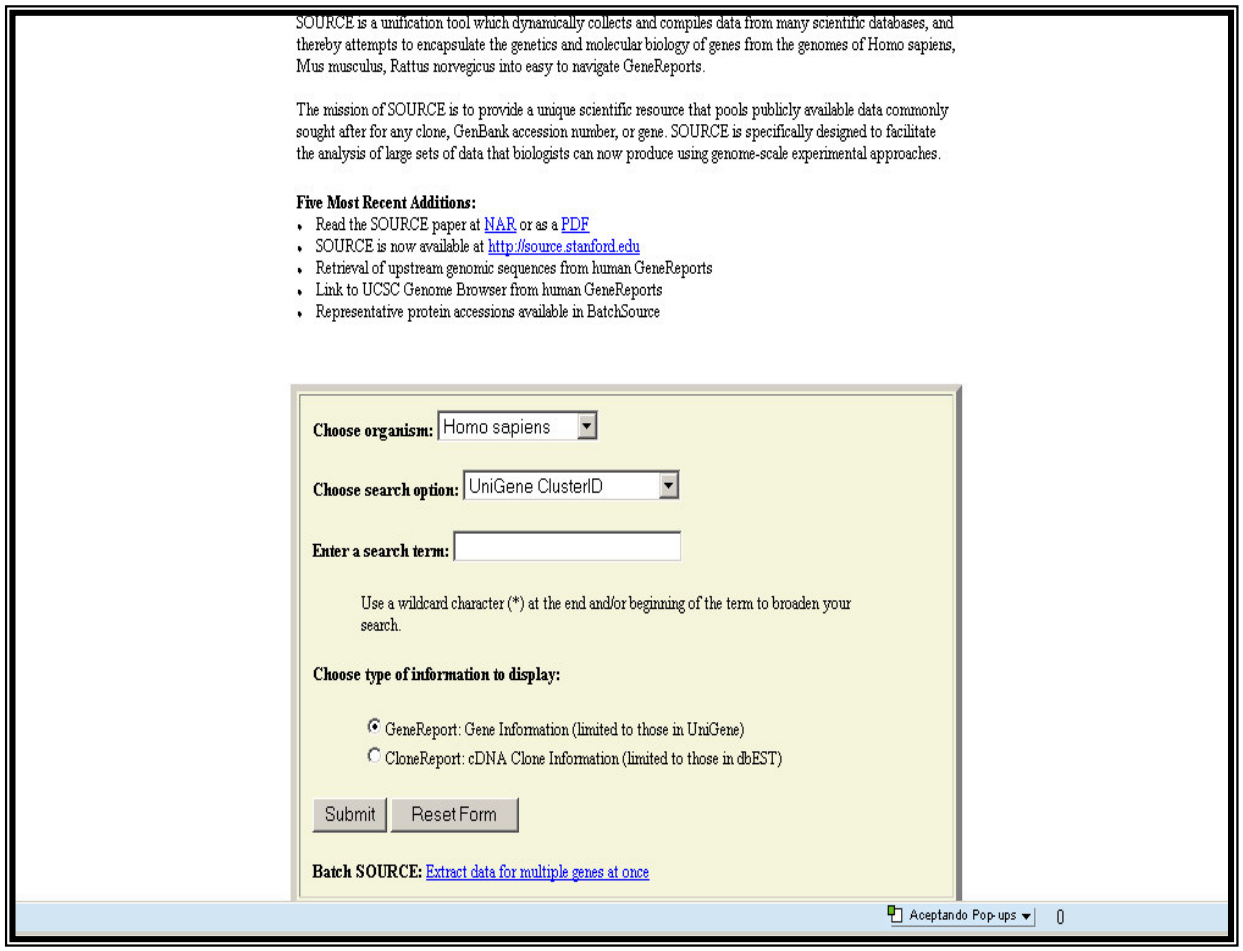


Fig. 7. Sitio web (SOURCE) para la búsqueda por gen o lista de genes en cuanto a su función biológica, biología molecular, etc.

Por otra parte, se realizó la distribución de los genes que comprende la biblioteca de expresión a lo largo de los cromosomas, es decir, los genes expresados a lo largo de cada uno de los cromosomas, y de esta manera obtener una gráfica de su distribución, la cual se realizó con otro programa, el DeltaGraphic. Primeramente se obtuvo el esquema de los cromosomas, así como el tamaño (en pares de bases) de cada uno. Esto se pudo realizar gracias a otro sitio web (http://genome.ucsc.edu/cgi-bin/hgGateway) (Fig. 8), de donde se extrajo la información requerida. En el programa se cargó la longitud de cada uno de los cromosomas, así como el nucleótido en que inicia y termina cada gen. Esto trajo como resultado la representación esquemática de la biblioteca de expresión a lo largo de los cromosomas.

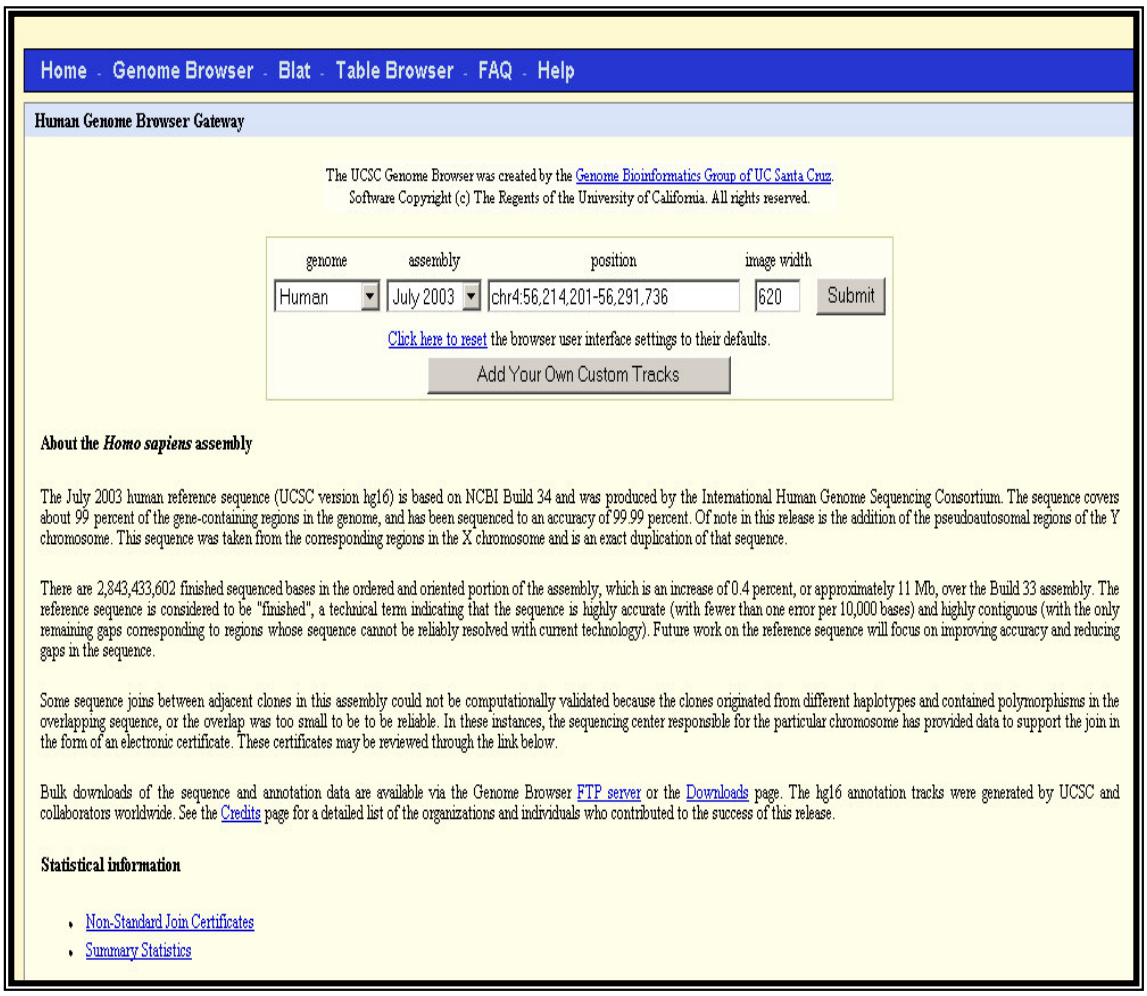


Fig. 8. Página principal del sitio web Genome Browser, de donde se extrajo todo lo referente a los cromosomas.

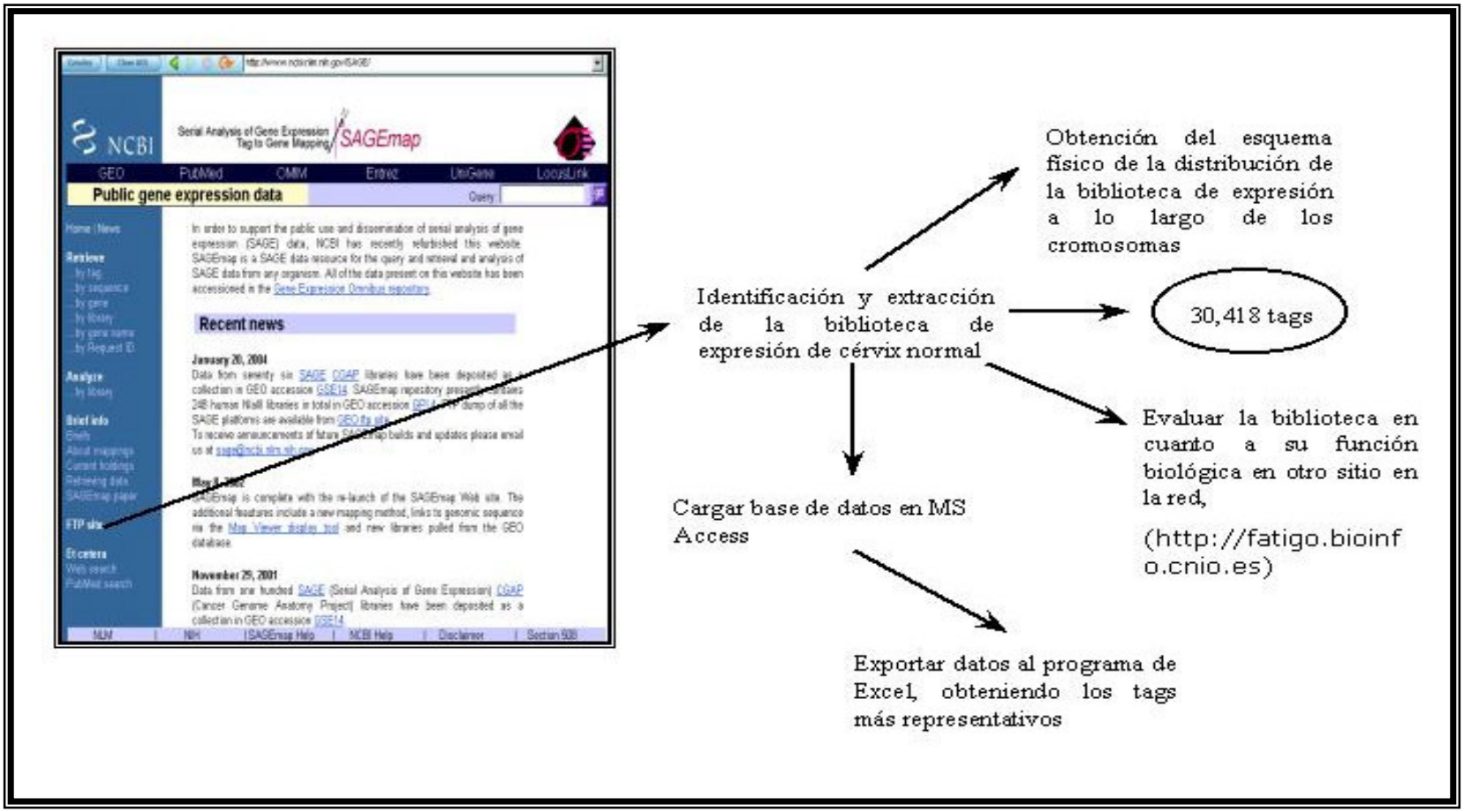
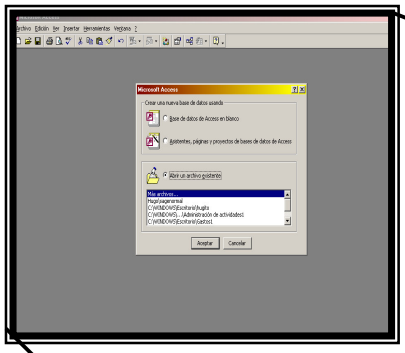


Fig.9. Representación esquemática del proceso de substracción *in silico*. El esquema general muestra como se aplicó la substracción de la Biblioteca de Expresión de Cérvix Normal, y de su tratamiento *in silico* en términos generales.

RESULTADOS Y DISCUSIÓN

Una vez que se realizó el análisis de la biblioteca de expresión de Cérvix en el programa de MCAccess, la tabla de resultados se exportó al programa de Excel, en donde se muestra la cantidad de genes totales de la biblioteca, así como el tag de cada gen, el número total de cada tag, y la normalización del número total de cada tag observados en toda la biblioteca, a tags por millón, para facilitar la comparación entre bibliotecas, su número de identificación de Unigene, y el nombre del gen (tabla 1). Del total de la biblioteca (30,448 tags), alrededor de 6,889 son genes (85.1%); 1,207 corresponden a EST´s (14.9%).



TAG	COUNT	TPM	UNIGNE CLU	GENE NAME
TACCTGCAGA	515	16930.7647	Hs.100000	S100 calcium binding protein A8 (calgranulin A)
TAGGTTGTCT	356	11703.5966	Hs.279860	tumor protein, translationally-controlled 1
TAGGTTGTCT	356	11703.5966	Hs.374596	ESTs, Highly similar to S06590 IgE-dependent histamine-releasing factor
TTTCCTGCTC	276	9073.57486	Hs.13273	KIAA0592 protein
TTTCCTGCTC	276	9073.57486	Hs.139322	small proline-rich protein 3
GAGGGAGTTT	201	6607.92952	Hs.76064	ribosomal protein L27a
GAGGGAGTTT	201	6607.92952	Hs.356342	ESTs, Highly similar to 2113200C ribosomal protein L27a [Homo sapiens] [H.sapiens]
GTGACCACGG	188	6180.55099	Hs.299882	ESTs, Highly similar to N-methyl-D-aspartate receptor 2C subunit precursor [Homo sapiens] [H.sapiens]
GTGGCCACGG	184	6049.0499	Hs.112405	S100 calcium binding protein A9 (calgranulin B)
GGGCTGGGGT	173	5687.42192	Hs.90436	sperm associated antigen 7
GGGCTGGGGT	173	5687.42192	Hs.350068	ribosomal protein L29
GCATAATAGG	168	5523.04557	Hs.356482	ESTs, Weakly similar to putative 60S ribosomal protein L21 [Arabidopsis thaliana] [A.thaliana]
GCATAATAGG	168	5523.04557	Hs.350077	ribosomal protein L21
TCAGATCTTT	161	5292.91867	Hs.108124	ribosomal protein S4, X-linked

TABLA 1. ***Tabla de los 15 genes más representativos de la biblioteca de expresión, así como de los nombres de los genes.***

Los genes más representativos dentro de la misma biblioteca podemos agruparlos en base a su abundancia en dos de las tres familias implicadas en el proceso de diferenciación de keratinocitos; proteínas precursoras de la capa cornificada (SPRRs), proteínas asociadas a filamentos intermedios, y las proteínas asociadas a uniones a calcio (S100s); las cuales se encuentran dentro de una región de 2Mb en la banda del cromosoma 1q21, llamado el complejo de diferenciación epidermal (CDE). Los patrones de expresión de estos genes durante el desarrollo de los keratinocitos se representan en la Fig.10.

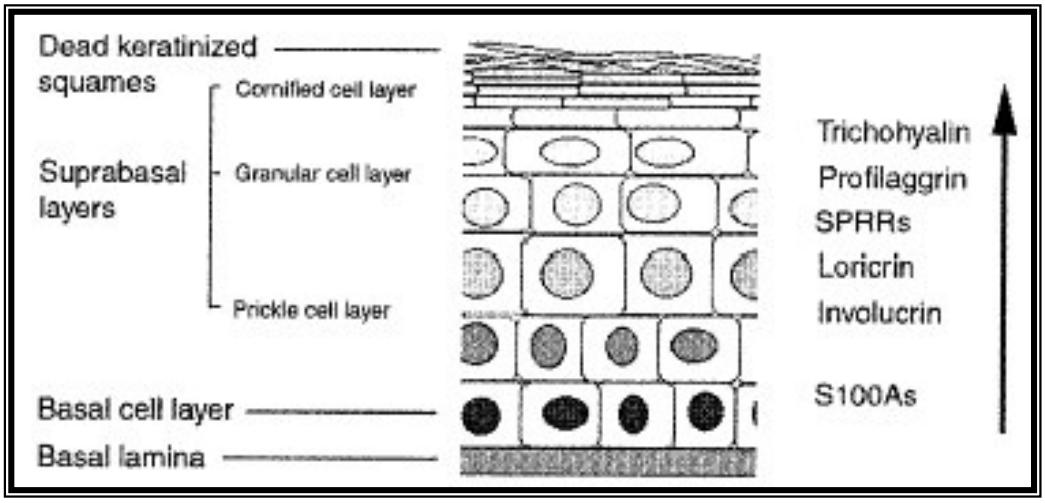


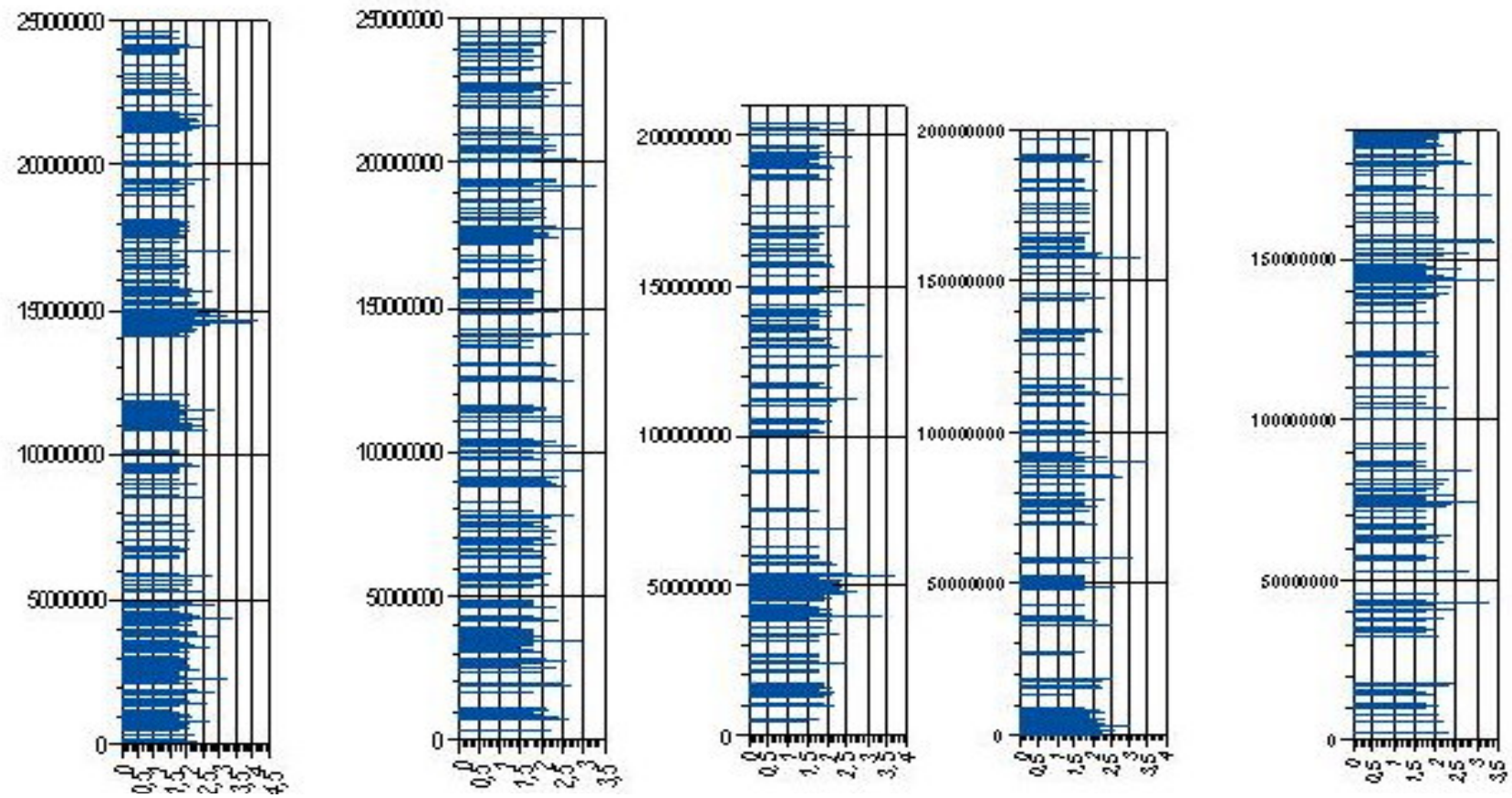
Fig. 10. ***Diagrama mostrando el proceso de diferenciación de keratinocitos. La flecha muestra la dirección del movimiento celular, desde la capa basal hasta las capas superiores***[27]**.**

Laco -localizaciónfuncional yestructuralmentedeestosgenes, aunado asucoordinadaexpresióndurantela diferenciaciónepidermal nosha permitido sugerir que están sujetos a algún tipo de mecanismo coordinadodecontrol transcripcional. Asimismo, suele vadaexpresión en la biblioteca de Cérvix nos permite sugerir que pueden tener un papel importanteenladiferenciacióndel mismo, yaquelaregión1q21 estáfrecuentementeinvolucradaenreorganizacióncromosomal enlos cáncereshumanos [28].

Encuantoasu funciónbiológicaenlaFig. 11, semuestralabiblioteca de expresión de Cérvix, en donde más del 60%de los genes se encuentranreferidosal metabolismo. Enestecontextolabibliotecase corroboró comparándola con otras bibliotecas de tejidos epitelial es normales, así como en otro sitio web (http://source.stanford.edu/cgi-bin/sourceSearch).

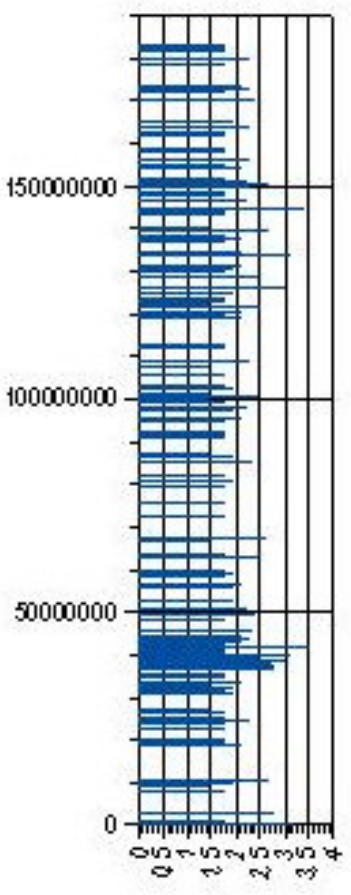
Fig. 11. Representación esquemática de la función biológica de los genes de la Biblioteca de Expresión de Cérvix Normal.

Por otra parte, los datos se sometieron al programa DeltaGraphic de donde se obtuvo la distribución física de la biblioteca de expresión a lo largo de los cromosomas (Fig. 12). Se puede observar pequeños agrupamientos muy particulares, como es el caso del 1q21, del cual se ha hablado de su gran importancia y relación con cánceres, por su rearreglo cromosómico. Aunque estudios citogenéticos en muestras de cáncer cervical han identificado diversos cambios estructurales específicos involucrando a los cromosomas 1, 3, 5, 17, y X. A pesar de eso los análisis de cariotipo convencionales no han definido alteraciones genéticas involucradas en el carcinoma cervical. Otros estudios con otras técnicas han definido pérdidas de material genético involucrando a los cromosomas 2q, 3p, 4p, 4q, 5q, 6q, 11q, 13q, y 18q, y ganancias del 1q, 3q, 5p y 8q en el carcinoma cervical [29]. Recientes estudios por nuestro grupo aplicando Hibridación Genómica Comparativa con líneas celulares derivadas de cáncer cervical muestran ganancias en las regiones cromosómicas 1q23, 3q11, 3q22, 5p y perdidas en el 2q, 4p, 6q, 9q, 19 [30]. Estos datos tienen relación, ya que en la Fig. 12 de la biblioteca de expresión de Cérvix Normal se muestran agrupamientos muy particulares en los cromosomas 1q, 3q, 5p, 6q, 9q, 12, 17, 19, que puedan estar involucrados con esta neoplasia. Sin embargo, para ayudarnos a explicar cómo es que el genoma humano se comporta en este tejido y así encontrar diferencias en cualquier patología, esta biblioteca de expresión debe ser contrastada con su contraparte tumoral. Esta comparación, es el comienzo para discriminar los cambios en la expresión de los genes que están asociados con la infección e integración del VPH en los keratinocitos cervicales, entre otros el factor más importante en el desarrollo de esta neoplasia. Actualmente es muy poca la información en términos de expresión diferencial de esta neoplasia.

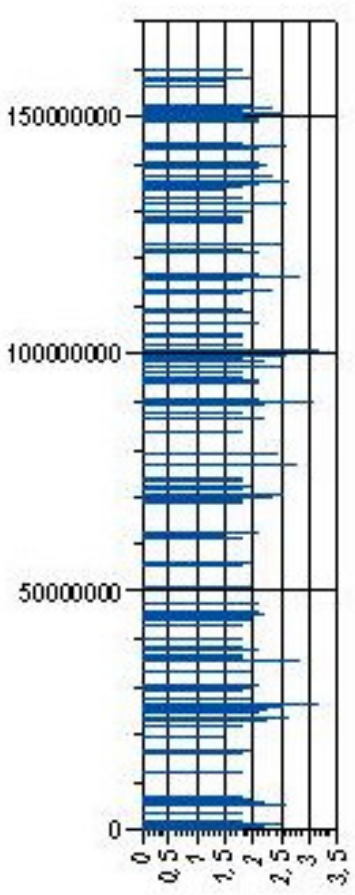


Cromo1 ***Cromo2*** ***Cromo3*** ***Cromo4*** ***Cromo5***

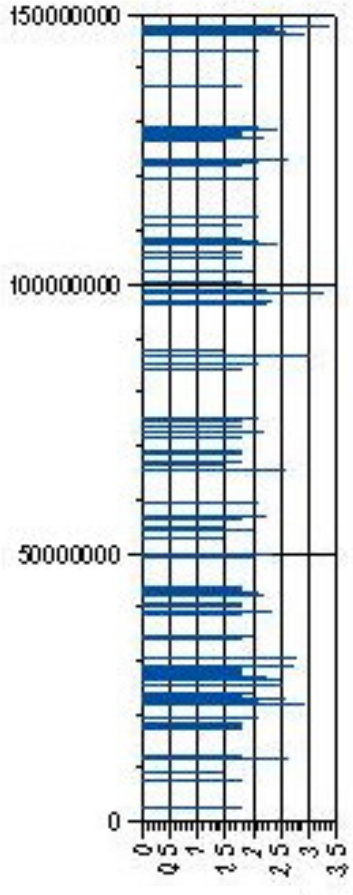
Fig. 12. Caracterización parcial de la Biblioteca de Expresión de Cérvix Normal. Este esquema muestra la distribución de todos los genes de la Biblioteca de Expresión a lo largo de los Cromosomas.



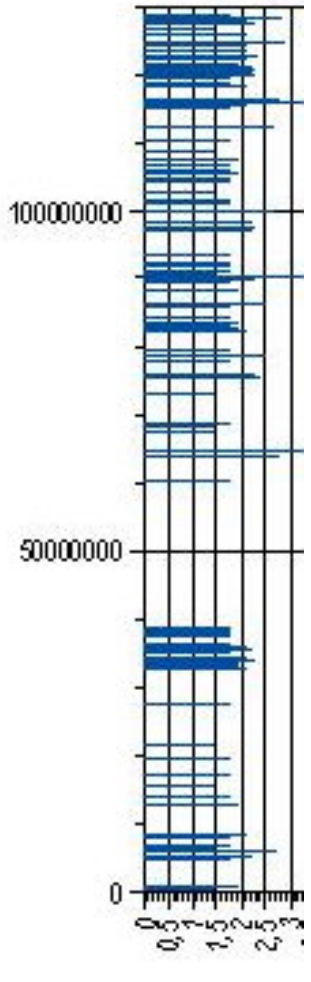
Cromo6



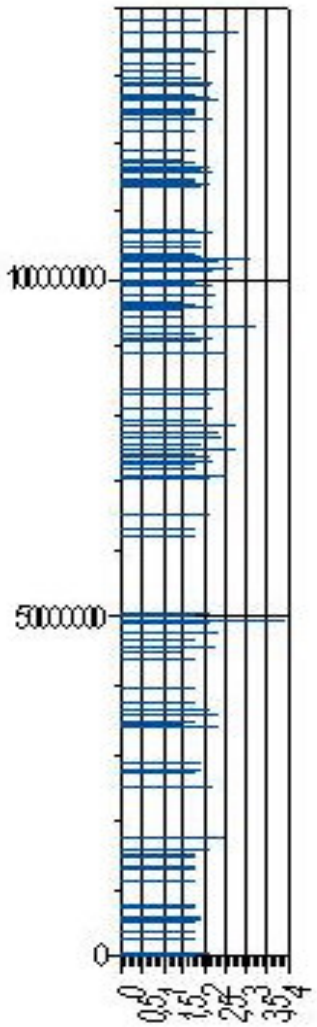
Cromo7



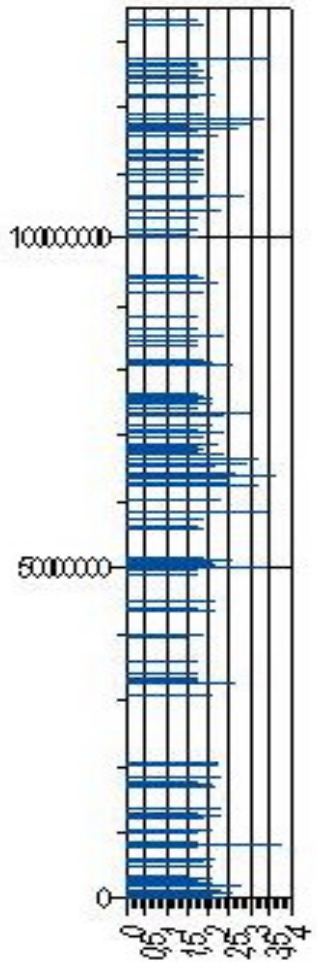
Cromo8



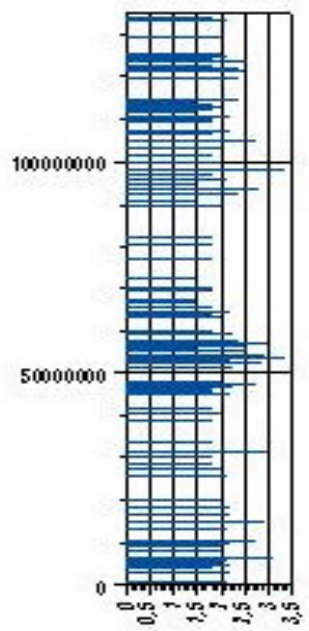
Cromo9



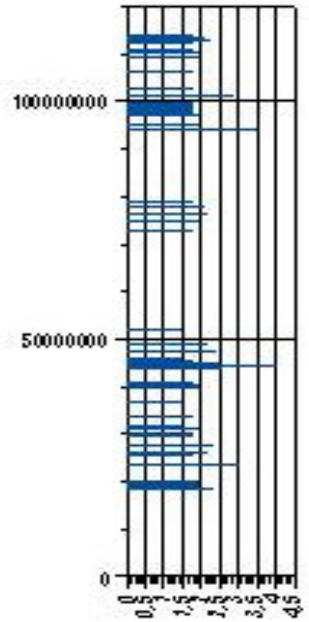
Cromo10



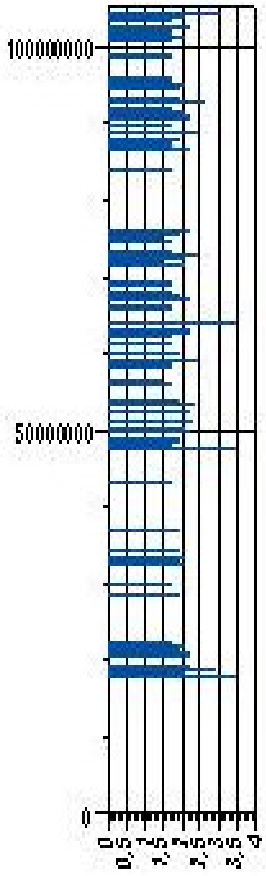
Cromo11



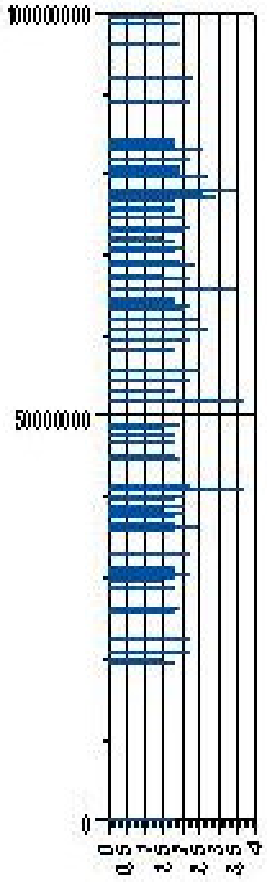
Cromo12



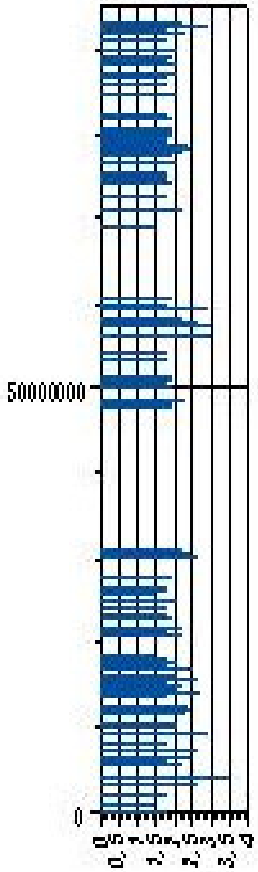
Cromo13



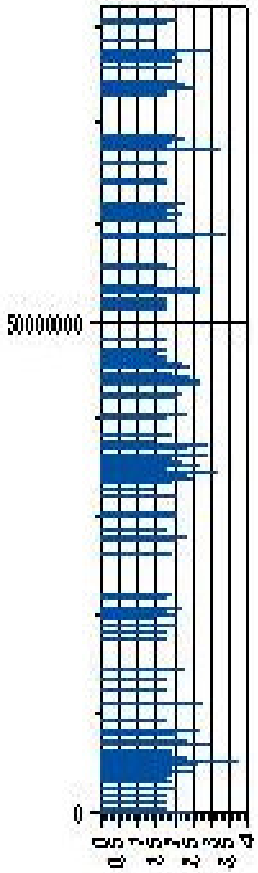
Cromo14



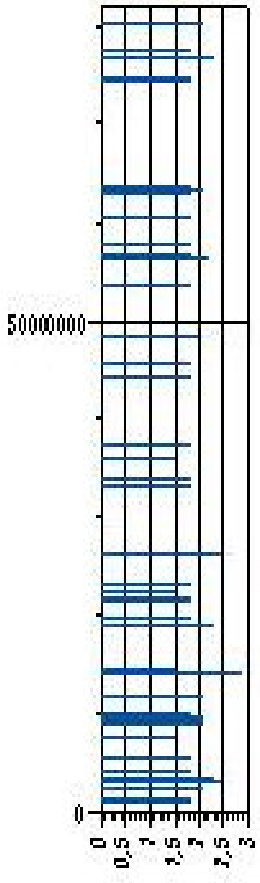
Cromo15



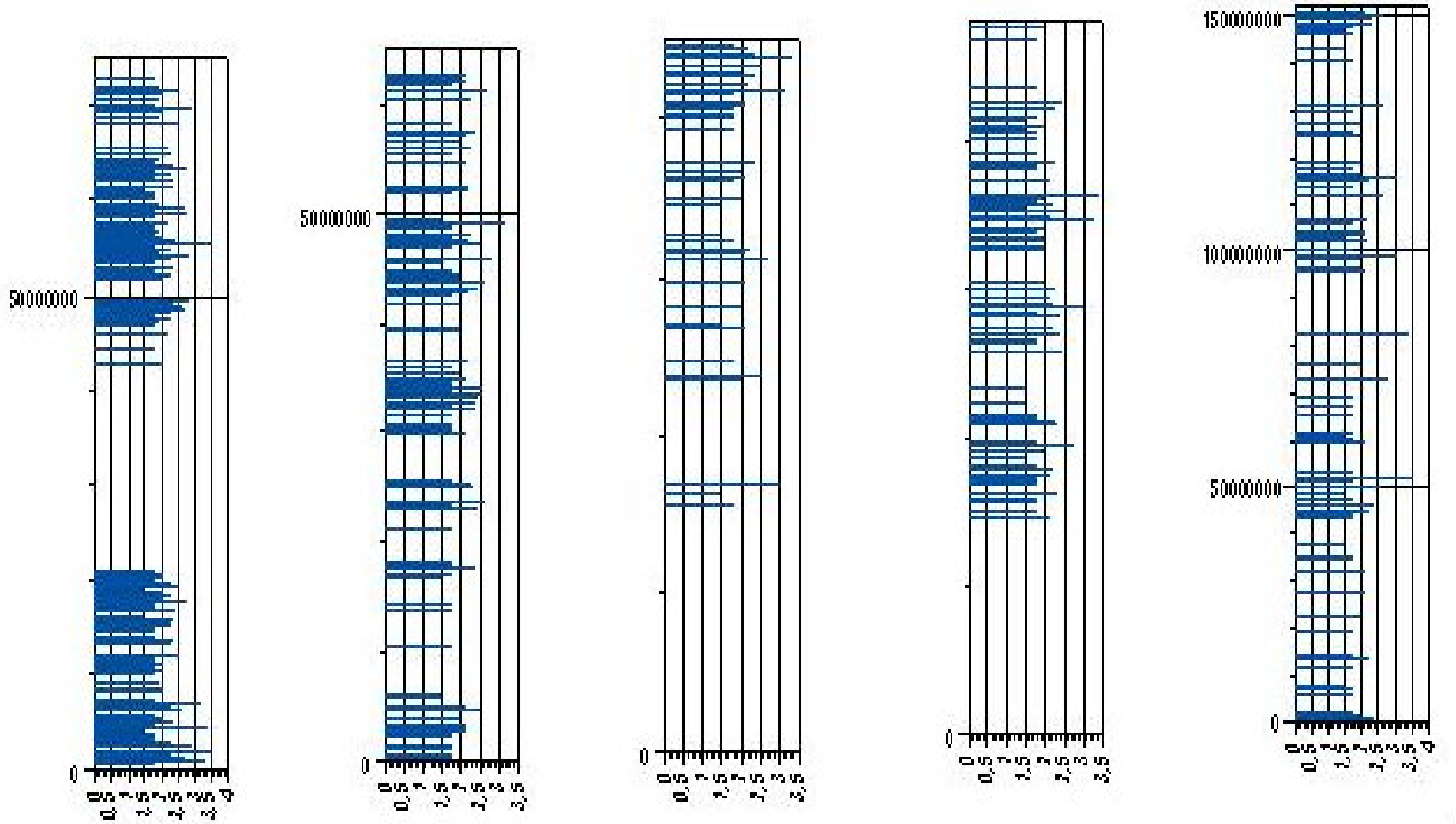
Cromo16



Cromo17



Cromo18



Cromo19 ***Cromo20*** ***Cromo21*** ***Cromo22*** ***Cromo23***

LITERATURA CITADA

1. Rosai J. (1989). Ackerman's Surgical Pathology. Vol. 2. Moscoby Company, Washington, D.C.

2. Kessel R. (1998). Basic Medical Histology. Oxford University.

3. Deaths by cause, sex and mortality stratum in WHO regions, estimates for 1999. World Health Statistics. World Health Organization.www -nt.who.int/whosis

4. Rodríguez, S.A., Ladastida, S., Tapia, R., Kuri, P& Macías C.G. (eds). 1999. México -Salud 2000. Registro Histopatológico de neoplasias en México 1993 -1996. Población derechohabient e del IMSS. Ciencia y Cultura Latinoamericana, México, D.F.

5. Walboomers, J.M., et al. (1999). Human Papillomavirus is a necessary cause of invasive cervical cancer worldwide. J. Pathology;189:12 -19

6. W. Tindle R. (2002) Immune evasion in human papillomav irus-associated cervical cancer. J Nature Reviews;2:59 -70

7. Griffiths, T.R., et al. (1999). Human Papillomavirus and Urological Tumors: Basic Science and Role in Penile Cancer;84:579 -86

8. Alani, R.M & Münger, K (1998). Human Papillomavirus and Associated Malignancies. J Clin Oncol;16:330 -337

9. Tindle, R.W. (2002). Immune evasion in human Papillomavirus -associated cervical cancer. Nature Reviews Cancer;2:59 -65

10. Carulli J, Artniger M, Swain, Root C, Chee L, Tulig C, Guerin J, Osborne M, Stein G, Lian J, Lomedico P (1998). High throughput of differential gene expression. *Jor Cell Biochem Supp;* 30/31: 286-296.

11. Pastorian K, Hawel L, Byus C (2000). Optimization of cDNA representational different analysis for the identification of differentially expresed mRNAs. *Analytical Biochemistry;* 283:89 - 98

12. Velculescu V and Vogelstein B (2000) Analysing uncharted transcriptomes with SAGE. *Trends in Genetics* ; 16:423-425

13. 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 (2001). Discovery of new markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma. *Cancer Research;* 61:4320 -4324

14. Yamamoto M, Wahatsuki T , Hada A, Ryo A (2001). Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods;* 250:45-66

15. Margulies EH, Kardia SL, Innis JW (2000). Identification and prevention of a GC content bias in SAGE libraries. *Nucleic Acids Res*; 15:60-70

16. Polyak K, *et al.* (2001). Gene discovery using the serial analysis of gene expression technique: implications for cancer research. *Journal of Clinical Oncology* ;19:2948 -2958.

17. Sasaki M, Nakahira K, Kawano Y, Katakura H, Yoshimine T, Shimizu K, Kim S, Ikenak a K (2001). MAGE -E1, a new member of the melanoma -associated antigen gene family and its expression in human glioma. *Cancer Res;* 61:4809-4814

18. Polyak K, *et al.* (2001). A SAGE (serial analysis of gene expression) view of breast cancer tumor progression. *Cancer Research,* 61:5697 -5702.

19. Polyak K, *et al.* (2001). HIN -1, a putative cytokine highly expressed in normal but not cancerous mammary epithelial cells. *Proceedings of the National Academy of Sciences,* 98:9796 -9801.

20. Yu, J *et al.* (1999) Identification and c lassification of p53 regulated genes. *Proc. Natl. Acad. Sci.* 96(25):14517-14522

21. Zhang L, Zhou W, Velculescu V, Kern S, Hruban R, Hamilton S, Vogelstein B, Kinzler K (1997). Gene normal expression profiles in normal and cancer cells. *Science*; 276:1268 -1272

22. Velculescu V. E . *et al* . (1999) Analysis of human transcriptomes. *Nature Genetics;* 23:387 -388

23. Leerkes MR, Caballero OL, Mackay A, Torloni H, O'Hare MJ, Simpson AJ, de Souza SJ. (2002) In silico comparison of the transcriptome derived from purified normal breast cells and breast tumor cell lines reveals candidate upregulated genes in breast tumor cells. Genomics. 2002 Feb; 79(2):257 -65.

24. Vasmatzis, G. et al. (1997). Discovery of three genes specifically expressed in human prostate by expressed séquense tag database analysis. Proc. Natl. Acad. Sci. 95:300 -04

25. Itoh, k., et al. (1998). Expression profile of active genes in granulocytes. Blood 15:1432 -41

26. Diehn M , Sherlock G, Binkley G, Jin H, Matese JC, Hernandez-Boussard T, . Rees CA, Cherry JM, Botstein D, . Brown PO and Alizadeh AA (2003) SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data; Nucleic Acids Research, 31(1):219 –223

27. Ruth, R. E., et al. (2001) Subchromosal Positioning of the Epidermal Differentati on Complex (EDC) in Keratinocyte and Lymphoblast Interphase Nuclei. Experimental Cell Research 272: 163-175

28. Agnieszka Pietas, et al. (2002). Molecular Cloning and Characterization of the human S100A14 Gene Encoding a Novel Member of the S100 Family. Genom ics 79:513 -522

29. H. Rao Pulivarthi, et al. (2004). Chromosomal amplification, 3q gain and deletion of 2q33 -q37 are the frequent genetic changes in cervinal carcinoma. BMC Cancer 4:1 -9

30. Hidalgo Alfredo, et al. (2003). Chromosomal imbalances in four new cervix carcinoma derived cell lines. BMC cancer 3:1471 - 2407