



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

ANÁLISIS DE LOS PATRONES DE EXPRESIÓN Y FOSFORILACIÓN
PROTÉICA EN LINEAS CELULARES DE CÁNCER CERVICO UTERINO

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
JUAN CARLOS HIGAREDA ALMARAZ

DIRECTOR DE TESIS
DR. SERGIO ENCARNACIÓN GUEVARA
CENTRO DE CIENCIAS GENÓMICAS

COMITÉ TUTOR
DRA. MARCELA LIZANO SOBERÓN
INSTITUTO NACIONAL DE CANCEROLOGÍA
DR. DIEGO ARENAS ARANDA
CENTRO MÉDICO NACIONAL SIGLO XXI

MÉXICO, D. F. DICIEMBRE DE 2012



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.

Indice

Indice	I
Indice de Figuras y Tablas	IV
Resumen	1
Abstract	3
1. Introducción	5
1.1 El cáncer, definición y características.	5
1.1.1 Mantenimiento de señalización de proliferación.	7
1.1.2 Evasión de supresores de crecimiento.	7
1.1.3 Resistencia a la muerte celular.	9
1.1.4 Habilitación de la inmortalidad replicativa.	10
1.1.5 Inducción de Angiogénesis	12
1.1.6 Activación de Invasión y Metástasis.	13
1.1.7 Reprogramación del Metabolismo Energético.	14
1.1.8 Evasión a la Destrucción Inmune.	15
1.1.9 Características Facultativas.	15
1.2 Análisis integrativo y biología de sistemas.	17
1.3 Análisis Proteómico.	19
1.4 La complejidad del proteoma.	19
1.5 El Cáncer Cérvico Uterino como Modelo de Estudio.	20
2 Planteamiento del problema.	21
2.1 Hipótesis.	22
2.2 Objetivo General.	22
2.3 Objetivos Específicos.	22
3. Materiales y Métodos.	23
3.1 Cultivo Celular.	23
3.2 Análisis Proteómico.	24
3.3 Reconstrucción de redes.	24

3.4 Extensión de redes.	25
3.5 Análisis de promotores.	25
3.6 Análisis de enriquecimiento de vías, ontología génica y cromosomas.	26
3.7 Cuantificación de genes y transcritos a partir de datos de RNA-seq.	26
3.8 Análisis de expresión diferencial a partir de datos de RNA-seq.	26
3.9 Análisis de factores transcripcionales.	27
3.10 Asociación de los blancos transcripcionales con los niveles de expresión de mRNA.	27
3.11 Enriquecimiento de proteínas fosforiladas.	28
3.12 Separación cromatográfica.	28
3.13 Análisis de MS/MS.	28
4. Resultados.	29
4.1 Existe un patrón de expresión proteica conservado en las líneas celulares de cáncer cérvico uterino.	29
4.2 1433Z es un “hub” dentro de la red de interacción proteína-proteína.	32
4.3 Existen características comunes de regulación en las proteínas de la red ampliada.	37
4.4 Existen diferentes perfiles de expresión genética y ontología génica entre las líneas celulares HeLa y NHEK.	40
4.5 La red de regulación de factores de transcripción expresados diferencialmente en células HeLa controla los procesos fundamentales en el mantenimiento del estado neoplásico.	42
4.6. Identificación de vías sobre o sub representadas que gobiernan el fenotipo neoplásico en la línea celular HeLa.	43
4.7 Existe un patrón definido en la activación de circuitos celulares que involucra interconexiones e interferencias, que se encuentran determinados por expresión genética, redes de regulación y vías diferenciales en la línea celular HeLa.	47
4.8 Los análisis de la expresión genética de la línea celular HeLa demuestran la existencia de dominios cromosomales.	54
5. Discusión	57
5.1. El grupo central de cáncer cérvico uterino representa un conjunto de procesos biológicos que permiten la preservación del fenotipo maligno.	57

5.2. La familia de transductores de señales 1433 se erige como un posible segundo “hit” dentro del cáncer cérvico uterino.	58
5.3. Los análisis de enriquecimiento de vías y el análisis de sitios de unión de TF a promotores sugieren una cooperación de c-Myc y E2F1 que mantiene encendidas distintas señales de supervivencia y proliferación.	59
5.4. Al añadir capas de información biológica y aumentar el nivel de resolución es posible encontrar nuevas y más detalladas vías de señalización y de regulación transcripcional.	61
5.5. El análisis de proteínas fosforiladas es un paso crítico en la validación de las predicciones realizadas por el modelo.	67
5.6. Las aberraciones cromosomales en las células promueven su evolución y la conservación del fenotipo maligno.	68
6. Conclusiones	69
7. Perspectivas	70
7.1.- Disección de vías de señalización.	70
7.2.- Análisis de la familia de transductores 1433.	71
7.3.-Establecimiento de un modelo usando información de líneas celulares y biopsias.	72
8. Bibliografía	72
9. Anexo 1 Proteomic patterns of cervical cancer cell lines, a network perspective	
10. Anexo 2 Analysis and Prediction of Pathways in HeLa Cells by Integrating Biological Levels of Organization with Systems-Biology Approaches	
11. Anexo 3 Material suplementario	

Indice de Figuras y Tablas

Fig. 1 Cuando una célula normal (rosa) con un daño en su genoma no es eliminada, la probabilidad de que las mutaciones acumuladas y la presión de selección favorezcan la proliferación sin control y la pérdida de control tisular es mayor. 5

Fig. 2 Propiedades del cáncer. Hanahan y Weimberg (2010) 6

Fig. 3 Redes de señalización intracelular que regulan las operaciones en las células cancerosas. Hanahan y Weimberg (2010) 17

Fig. 4 Se compararon los perfiles de proteínas de seis líneas celulares de cáncer cervical por 2D SDS-PAGE, y se estableció un conjunto de proteínas comunes a ellas que no encontraban en nuestro control. Este conjunto de proteínas fue llamado "núcleo central de cáncer cérvico uterino", (a). A partir de este núcleo, fue reconstruida una red PPI (b), Esta red fue expandida y se utilizó para obtener un enriquecimiento de GO y vías celulares. Finalmente, se llevó a cabo un análisis de los factores de transcripción contenidos en la red extensa utilizando un análisis de ChIP-Seq mediante la base de datos del proyecto ENCODE. 30

Tabla 1 Proteínas identificadas como miembros del "núcleo central del cáncer cérvico uterino". El nombre de la proteína correspondiente a la base de datos UniProt, la puntuación y el valor esperado fueron obtenidos del motor de búsqueda MASCOT. El valor esperado es el número de correspondencias con resultados iguales o mejores que de los esperados al azar. La puntuación $-10 * \text{LOG}_{10}(P)$, donde P es la probabilidad absoluta. Cuanto menor es la expectativa de valor, la puntuación es mejor. 32

Fig. 5 Reconstrucción de red de proteínas expresadas diferencialmente en 6 líneas celulares de cáncer cérvico uterino, en comparación con un control no maligno. Las conexiones entre proteínas representan interacciones proteína-proteína verificadas experimentalmente entre 33 de las 66 proteínas identificadas en este estudio. Las proteínas no incluidas en esta red no poseen interacciones experimentalmente comprobadas. La red fue construida utilizando Cytoscape y el plugin Bisogenet. 34

Fig. 6 Ampliación de la red. Las proteínas pertenecientes a nuestro núcleo de expresión diferencial fueron usadas como cebo para obtener interacciones proteína-proteína probadas experimentalmente mediante la adición de nodos vecinos a una distancia de uno. Esto resultó en una red de 1321 nodos y 9666 bordes. La red se expandió utilizando Cytoscape y el plugin Bisogenet. 35

Tabla 2 Enriquecimiento de Ontología Genica nivel 3, usando como categoría procesos biológicos. El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento. 36

Tabla 3 Enriquecimiento de vías basados en KEGG. El tamaño del conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la base de datos KEGG correspondiente según las vías almacenadas en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas que son parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento 37

Tabla 4 Factores de transcripción significativamente sobre expresados identificados mediante datos de ChIP-Seq pertenecientes a la base de datos ENCODE. Los picos fueron asignados a las regiones correspondientes de secuencias de promotor [-700 a 300 pb], basados en los sitios de inicio de transcripción (TSS) anotados en GENCODE. Los picos cerca de la TSS de los genes de la red extendida y todos los genes GENCODE se refiere a la cantidad de picos encontrados en la región promotora. El P-valor se calculó mediante una prueba hipergeométrica. El tamaño de la muestra se consideró como la cantidad de genes que correspondían con las proteínas de la red extendida. El tamaño de la población es el número de entradas GENCODE anotadas que poseían una anotación completa y estaban separados por una distancia de hasta 500 pares de bases. 38

Tabla 5 Factores de transcripción significativamente sobre expresados identificados mediante datos de ChIP-Seq pertenecientes a la base de datos SwitchGear. Los picos fueron asignados a las regiones correspondientes de secuencias de promotor [-700 a 300 pb], basados en los sitios de inicio de transcripción (TSS) anotados en SwitchGear. Los picos cerca de la TSS de los genes de la red extendida y todos los genes SwitchGear se refiere a la cantidad de picos encontrados en la región promotora. El P-valor se calculo mediante una prueba hipergeométrica. El tamaño de la muestra se consideró como la cantidad de genes que correspondían con las proteínas de la red extendida. El tamaño de la población es el número de entradas GENCODE anotadas que poseían una anotación completa y estaban reportados en SwitchGear con una puntuación mayor a 20. **39**

Fig. 7 Integración de diferentes capas de información biológica dentro de la dinámica celular. Se realizó un análisis del total de transcritos en la línea celular HeLa, este análisis presentó un panorama general de la expresión génica. Posteriormente, se realizó un análisis de expresión diferencial mediante RNA-seq y un análisis de sobre representación de la actividad de FT. Esto permitió la reconstrucción de rutas metabólicas y vías de señalización y regulación de la transcripción celular. Por último, se validó esta reconstrucción con un análisis de enriquecimiento de proteínas fosforiladas mediante fosfoproteómica. **41**

Fig. 8 Patrones de expresión génica en la línea celular HeLa a) La distribución de los transcritos totales muestra que existen dos poblaciones, una de baja abundancia y una segunda, mayor, de alta abundancia. Este análisis muestra que los parámetros utilizados para la búsqueda de transcritos de baja abundancia funciona correctamente. b) Una representación gráfica de los procesos celulares que se distinguen mediante ontología génica (GO), utilizando el dominio “componentes celulares” en nivel 3, la cantidad de elementos recuperados se comparó con el tamaño total de la vía. c) Un diagrama de dispersión que muestra la calidad de los resultados diferenciales del análisis de la expresión diferencial mediante RNA-Seq, utilizando la línea de queratinocitos epiteliales NHEK como control. Se encontro un total de 3.360 genes sobre expresados y 2.129 genes sub-expresados. d) La distribución porcentual del dominio “procesos biológicos” GO nivel 3, representados por los transcritos sobre expresados. e) La distribución porcentual del dominio “procesos biológicos” GO nivel 3 representados por los transcritos sub expresados. Estas gráficas fueron construidos a partir de un resumen de todos los términos GO similares en un circuito celular funcional. **44**

Fig. 9 Redes de expresión de factores transcripcionales a) La distribución porcentual del dominio “procesos biológicos” GO nivel 3 construidos a partir de la red de FT sobre representados, destacando "Proliferación celular", "Metabolismo de los bloques de construcción", "Organización celular", "Angiogénesis", "Metabolismo central" y "Señalización ". Esta tabla fue construida a partir de un resumen de todos los términos GO similares en un circuito celular funcional. b) Hubs que se obtuvieron a partir de la medición de la centralidad “node degree” en la red de FT sobre representados. El color indica la calificación, donde el rojo es el más alto y amarillo el más bajo. c) Hubs que se obtuvieron a partir de la medición de la centralidad “betweenness” en la red de FT sobre representados. El color indica la calificación, donde el rojo es el más alto y amarillo el más bajo. **45**

Tabla 6 Enriquecimiento de vías según PID basados el conjunto de transcritos sobre expresados. El tamaño del conjunto se refiere al número de transcritos que tienen un ID Uniprot en la vía PID correspondiente según el sitio ConsensusPathDB. El número de candidatos se refiere al número de proteínas que forman parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento. **46**

Fig. 10 Red de interconexiones y “crosstalks” entre los circuitos celulares. La red se realizó a partir de datos obtenidos por el análisis de las vías de señalización y regulación. Los nodos representan las proteínas que componen cada una de las vías, cada ruta se indica mediante el uso de colores diferentes. **49**

Fig. 11 Análisis de Meta-vías. Se realizó un análisis combinando las vías de señalización y regulación transcripcional; los bordes indican la relación regulatoria o jerárquica, y los nodos indican vía. Los colores indican cada una de las “características del cáncer”, con dos de las características más representativas resaltadas por nodo. **50**

Fig. 12 Vías metabólicas sobre representadas. Los mapas metabólicos fueron reconstruidos mediante la base de datos KEGG, las redes de FT sobre representadas y los análisis fosfoproteómicos. Los transcritos sobre

expresados se muestran en verde, los transcritos sub expresados en rojo y las fosfoproteínas identificadas en morado. **51**

Fig. 13 Mapas de ciclo celular y vías de moléculas de adhesión celular a) Reconstrucción del ciclo celular y complejo de Mantenimiento del Minicromosoma (MCM). b) Reconstrucción de las vías de moléculas de adhesión celular. Ambos mapas se construyeron la base de datos KEGG, las redes de FT sobre representadas y los análisis fosfoproteómicos. Los transcritos sobre expresados se muestran en verde, los transcritos sub expresados en rojo y las fosfoproteínas identificadas en morado. **52**

Fig. 14 Vía de componente del espliceosoma. Esta vía fue reconstruido e construyeron la base de datos KEGG, las redes de FT sobre representadas y los análisis fosfoproteómicos. Los transcritos sobre expresados se muestran en verde, los transcritos sub expresados en rojo y las fosfoproteínas identificadas en morado. **53**

Fig. 15 Enriquecimiento de vías de señalización, según PID, de los cromosomas que contienen las secuencias de los transcritos sobre expresados. a) Cromosoma 12. b) Cromosoma 17. El tamaño de los nodos representa el número de genes contenidos en cada categoría y el color el P-valor obtenido. Todos los resultados son significativos. **55**

Fig. 16 Enriquecimiento de vías de señalización, según PID, de los cromosomas que contienen las secuencias de los transcritos sub expresados. a) Cromosoma 2. b) Cromosoma 19. El tamaño de los nodos representa el número de genes contenidos en cada categoría y el color el P-valor obtenido. Todos los resultados son significativos. **56**

Tabla 8 Enriquecimiento de vías de señalización, según PID, de las citobandas del Cromosoma 1 que contienen las secuencias de los transcritos sobre expresados. **57**

Tabla 9 Enriquecimiento de vías de señalización, según PID, de las citobandas del Cromosoma 1 que contienen las secuencias de los transcritos sub expresados. **58**

Resumen

El cáncer cérvico uterino (CaCu) representa una oportunidad única para estudiar una transformación maligna debido a que posee un origen bien determinado, el 90.7% de los casos surgen como una consecuencia de la infección de Virus del Papiloma Humano de Alto Riesgo (VPH-AR).

Debido a la eliminación frecuente y espontánea de las secuencias virales, no todos los pacientes infectados con el VPH-AR desarrollan cáncer cervical. Esto indica que la mayoría de infecciones por VPH son sub clínicas y sólo una pequeña fracción de las infecciones por VPH-AR produce lesiones epiteliales tempranas, y una fracción aún más modesta de las lesiones progresa hacia cáncer. En consecuencia, incluso si la infección por VPH-AR puede ser considerada como el “primer hit” que da lugar al CaCu, los pasos que permitan el desarrollo del cáncer aún no se han descrito. Nuestro objetivo ha sido determinar y predecir los principales eventos, metabólicos, de señalización y regulación de Ca Cu mediante la integración y el análisis de secuenciación masiva, microarreglos, expresión y fosforilación proteica utilizando un enfoque de la biología de sistemas.

Se identificó un consenso de 66 proteínas, fueron seleccionadas las entidades que se encontraban presentes en las seis líneas celulares de cáncer, pero no aparecen en el control HaCaT, o aquellas que cambiaron constantemente y de manera significativa. Este patrón proteico es fundamental en el cáncer ya que varios estudios han relacionado su expresión aberrante con esta enfermedad. A partir de este conjunto de proteínas, construimos una red PPI que señaló, a través de un análisis topológico, a algunas proteínas que probablemente desempeñen un papel central en el proceso neoplásico, como 1433Z, esta proteína funciona como un cuello de botella debido a su pleiotropia.

En la línea celular HeLa, encontramos un total de 19,974 transcritos expresados, así como 3,360 sobre-expresados y 2,129 sub expresados en referencia a la línea celular NHEK. Usando estos genes, se reconstruyeron vías y se realizaron análisis de enriquecimientos de Ontología Génica que nos permitieron sugerir cuales vías

de señalización, regulación y metabólicas son cruciales para el mantenimiento del fenotipo neoplásico. Utilizando los datos de microarreglos depositados en la base de datos GEO, se llevó a cabo una búsqueda de actividad de factores transcripcionales. Con los datos resultantes, se construyó una red de interacción proteína-proteína (PPI) y se realizó un análisis topológico de la misma para encontrar los genes clave. Con los datos anteriormente mencionados, se construyó un modelo de las relaciones de las diferentes vías reconstruidas y su correlación con “las características del cáncer. Finalmente, se implementó un análisis de enriquecimiento e identificación de proteínas fosforiladas en la línea celular HeLa que permitiera validar nuestras predicciones.

Nuestros resultados sugieren un fenotipo compartida por las seis líneas celulares de cáncer cervical, como resultado de la sobre expresión de un consenso de las proteínas, destacando a 1433Z. Esta proteína de transducción de señales se ha señalado en otros modelos de cáncer como uno de los responsables de la transformación maligna y la toma de decisiones sobre la vida y la muerte de las células

En la línea celular HeLa, nuestros resultados muestran un sistema celular muy robusto, lo que podría sugerir por qué los ensayos de terapia dirigidos contra dianas moleculares no son tan eficientes como se esperaba. La redundancia de funciones, nos permite especular sobre los mecanismos que promueven las ventajas adaptativas que permiten a las células cancerosas evadir el control tisular.

Abstract

Cervical cancer is a unique opportunity to study the malignant transformation because of its common origin: 90.7% of the cases arise as a consequence of High-Risk Human Papilloma Virus (HR-HPV). Owing to a frequent and spontaneous elimination of viral sequences, not all the patients infected with HR-HPV develop cervical cancer. This indicates that most HPV infections are subclinical and only a small fraction of HR-HPV infections produce early epithelial lesions, and a more modest fraction of those lesions will develop into cancer. Consequently, even if infection by HR-HPV can be considered as the initial hit that gives rise to cervical cancer, the superseding steps that enable cancer development have not yet been described. Our objective is to predict and identify key events, signaling and regulation pathways that have a role in maintaining the phenotype of cervical cancer cell lines by integrating microarray and massive transcriptomics analysis with protein expression and phosphorylation using a systems biology approach.

A consensus of 66 proteins was identified; these electrophoretic entities were present in all six cancer cell lines but did not appear in the HaCaT control, or those which changed consistently and significantly. The proteins of this pattern are important in cancer because several studies have linked their aberrant expression to the disease. Starting from this core set of proteins, we acquired a PPI network that pointed, through topological analysis, to some proteins that may well be playing a central role in the neoplastic process, such as 1433Z, this protein plays the role of a bottleneck.

In the HeLa cell line, we found a total of 19,974 transcripts expressed, as well as 3360 over-expressed and 2129 under-expressed transcripts in reference to the NHEK cell line. Using these genes, we assembled pathways and Gene Ontology enrichment analyses that allowed us to suggest which signaling, regulation and metabolic pathways are crucial to maintaining the neoplastic phenotype. Using microarray data deposited in GEO, a search of transcription-factor response was conducted. With the resulting data, we built protein-protein interaction networks, and centrality measures of those networks were used to find key genes. Finally, an

analysis and identification of phosphorylated protein enrichment in the HeLa cell line was performed to validate our predictions.

Our results suggest a phenotype shared by the six cervical cancer cell lines as a result of the overexpression of a consensus of proteins, highlighting to 1433Z. This signal transduction protein has been reported in other models of cancer as being responsible for malignant transformation and the decision between life and death of cells

In the Hela cell line, our results show a very robust cellular system, which may suggest why therapy trials directed against molecular targets are not as efficient as expected. An interesting hypothesis, the redundancy of functions, allows us to speculate on the mechanisms that promote adaptive advantages that permit cancer cells to evade control tissues.

1. Introducción

1.1 El cáncer, definición y características.

El cáncer es un conjunto de enfermedades que presentan como fenotipo en común un crecimiento descontrolado y trastornos en la diferenciación celular, además de invasión y metástasis en tejidos sanos comprometiendo su función (World Health Organization, 2011).

Para su desarrollo, el cáncer requiere un muy complejo conjunto de condiciones: es conducido por un modelo Darwiniano a nivel celular (Little 2010), que comprende todos los niveles de información celular (genética, epigenética, regulación transcripcional y traduccional, así como modificaciones post-traduccionales y regulación por miRNAs), además, implica comunicación entre diferentes tipos celulares, la relación con el micro ambiente del tejido y finalmente, al organismo en su totalidad (Gentles y Gallahan 2011; Sonnenschein y Soto 2011). Su concepción ha evolucionado de acuerdo al conocimiento y perspectivas del contexto científico en el cual es interpretado.

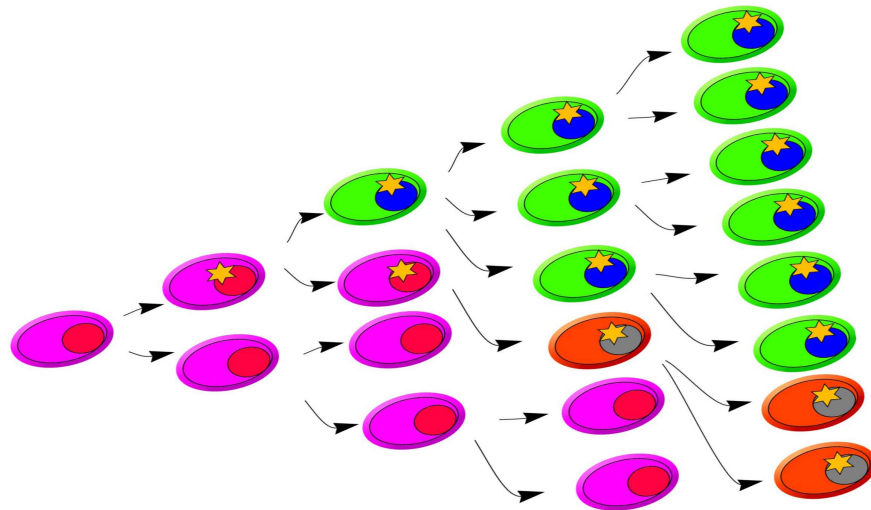
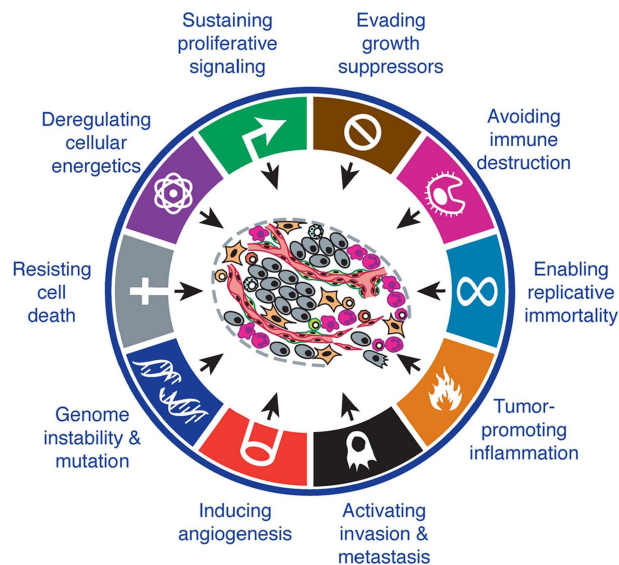


Fig. 1 Cuando una célula normal (rosa) con un daño en su genoma no es eliminada, la probabilidad de que las mutaciones acumuladas y la presión de selección favorezcan la proliferación sin control y la pérdida de control tisular es mayor.

A lo largo de los años se han logrado avances, tanto experimentales como conceptuales que ocasionaron una revolución en la forma en la cual se concibe este padecimiento. Una de las más importantes fue el establecimiento de posibles reglas que gobiernan la transformación de células humanas normales en cáncer. Existe un pequeño número de capacidades adquiridas, moleculares, bioquímicas y celulares que al parecer son compartidas por todos los tipos de cáncer en humanos.

Weinberg y Hanahan (2000) plantearon que existen seis características clave en todos los tipos de cáncer, estas son: El mantenimiento de señalización de proliferación, evasión de supresores de crecimiento, resistir la muerte celular, habilitación de la inmortalidad replicativa, inducción de angiogénesis y activación de invasión y metástasis. Una reciente revisión de los autores involucra dos nuevas características emergentes; la reprogramación del metabolismo energético y la evasión a la respuesta inmune, además de dos características facultativas: Inestabilidad genómica y mutación; e inflamación promovida por tumorigénesis (Hanahan y Weinberg 2011).



. Fig. 2 Propiedades del cáncer. Hanahan y Weimberg (2010)

1.1.1 Mantenimiento de señalización de proliferación

Un tejido es capaz de ejercer un estricto control en la producción y liberación de señales promotoras de crecimiento que permiten la entrada y progresión a través del crecimiento (acumulación de masa) y el ciclo celular con precisión, esto asegura una homeostásis del número celular y el mantenimiento de la función y la arquitectura del tejido (Lelièvre 2009). En general, las señales se transmiten por factores de crecimiento que se unen a receptores de superficie celular, comúnmente, este tipo de receptores contienen dominios intracelulares de tirosina quinasa. Al recibir la señal, el dominio desencadena una cascada de fosforilación que regula la progresión del ciclo y el crecimiento celular, a menudo estas señales son capaces de influenciar a diversos niveles aspectos importantes en el contexto celular, como la sobrevivencia y el metabolismo energético (Sun et al. 2008; Reinhardt y Yaffe 2009)

Las células malignas son capaces de gobernarse a sí mismas, ya que adquirieron la capacidad para mantener la señalización de proliferación de formas alternativas, como la estimulación proliferativa autócrina, estimulación parácrina de las células adyacentes, sobre-expresión de receptores de membrana, alteraciones estructurales en receptores, activación constitutiva de la señalización como B-Raf y desregulación de los circuitos de regulación negativa como en el caso de Ras (Collado y Serrano 2010; Cheng et al. 2008; Davies y Samuels 2010; Jiang et al. 2003; Yuan y Cantley 2008; Wertz y Dixit 2010).

1.1.2 Evasión de supresores de crecimiento

Dentro de un tejido normal, las células poseen una serie de programas bien definidos y de gran alcance que regulan negativamente la proliferación celular, estos programas se dividen en los que dependen de señales internas o externas y aquellos que están en manos de la acción de genes supresores de tumores (Azar et al. 2009).

Hasta el momento se ha encontrado una gran cantidad de genes que cumplen la función de supresores de tumores. Ellos limitan el crecimiento y la proliferación celular de múltiples maneras. Existen dos supresores de tumores arquetipos, p53 y pRb (proteína de Retinoblastoma). Ellos operan como nodos centrales de control dentro de dos circuitos celulares complementarios clave, gobiernan las decisiones de las células para proliferar o disparar la senescencia y programas de muerte (Joerger y Fersht 2008; Manning y Dyson 2011).

Las células cancerosas poseen diferentes arsenales que les permite evadir la regulación por supresores de tumores, por un lado existen aquellos que poseen mutaciones en estos dos guardianes celulares, haciéndolos inactivos o menos eficientes en su labor, o aquellos que intervienen en sus vías para evitar la respuesta adecuada.

Una de las vías mejor conocidas por sus efectos anti-proliferativos es la de TGF- β , sin embargo, las células cancerosas poseen un intrincado mecanismo que les permite evadir esta circuitería. En muchos tipos de tumores en estados avanzados, la señal de TGF- β es redirigida hacia la activación del programa celular llamado transición epitelio-mesénquima (EMT), lo que confiere a las células una serie de rasgos asociados con un alto grado de malignidad (Bierie y Moses 2006; Massagué 2008; Ikushima y Miyazono 2010).

También se han detallado mecanismos anti-proliferativos más específicos, como los contactos célula-célula. Las señales de inhibición por contacto ayudan a mantener el balance entre proliferación y muerte, lo cual permite que las células sean capaces de construir y mantener tejidos arquitectónicamente complejos, sin embargo, en las células cancerosas se ha demostrado, tanto empírica como experimentalmente que estos candados han sido superados (Nelson y Bissell 2006).

Se han reportado múltiples ejemplos en diversos tipos de cáncer, como la supresión de E-caderina (Sun-Mi Park et al. 2008), inactivación de la vía de Salvador-Warts-Hippo (Bin Zhao et al. 2007), desregulación de la expresión de la

familia de proteínas 1433 (Li et al. 2008), alteraciones en la actividad de Merlín (Curto et al. 2007), la pérdida de la expresión de LKB1 (Partanen et al. 2009) o la activación de EphB3 y EphB4 por la efrina-B2 (Astin et al. 2010).

1.1.3 Resistencia a la muerte celular

La muerte celular programada es una barrera a la proliferación descontrolada y al desarrollo del cáncer (Adams y Cory 2007). Las células cancerosas desarrollan una variedad de estrategias para limitar o burlar a la apoptosis, en relación con la anulación de señales de crecimiento, una de las estrategias más comunes es la pérdida de la función supresora de p53, que elimina este sensor crítico de daño de la circuitería inductora de apoptosis.

Alternativamente, las células cancerosas pueden obtener resultados similares aumentando la expresión de reguladores anti-apoptóticos (Bcl-2, Bcl-x) o señales de sobrevivencia (Igf1/2). También puede regular negativamente factores pro-apoptóticos (Bax, Bim, Puma) o eliminando la transducción de los ligandos de la vía extrínseca de muerte inducida (Fulda 2010). La gran cantidad de mecanismos que permiten evadir a la apoptosis pueden ser un reflejo de la redundancia y robustez de las señales inductoras de muerte que las poblaciones de células cancerosas soslayan durante su evolución a la malignidad (Ryder et al. 2012).

Recientemente, el conocimiento del abanico de maquinarias y procesos que dan origen a la “muerte celular programada” se ha ampliado considerablemente, incluyendo a la autofagia e incluso a una nueva forma de muerte conocida bajo el nombre de “necroptosis” (Christofferson y Yuan 2010).

Se sabe que la autofagia representa una respuesta sumamente importante en la fisiología celular, opera a niveles basales dentro de la célula, pero puede ser fuertemente inducida por diversos estados celulares o distintos tipos de estrés, siendo el más estudiado la deficiencia de nutrientes (Levine y Kroemer, 2008). Existen dos conexiones obvias entre la apoptosis y la autofagia, la vía de

señalización de AKT/mTOR (Levine y Kroemer 2008) y la proteína Beclina-1 (Mizushima 2007), lo cual promueve la tesis de que la autofagia puede operar en contra del cáncer en concertación con la apoptosis o de una manera separada.

En las células cancerosas, la respuesta a ciertos tipos de estrés, como falta de nutrientes, radioterapia o drogas citotóxicas pueden desencadenar una respuesta autofagica que posee efectos protectores contra la muerte celular (White y DiPaola 2009; (Apel et al. 2009). Incluso, células cancerosas fuertemente estresadas han mostrado un comportamiento reversible, similar a la dormacia, lo que podría explicar la persistencia y la reaparición de algunos tumores que han sido tratados mediante agentes tóxicos o radiación (Lu et al. 2008)

Finalmente, la necroptosis es un mecanismo de muerte celular recientemente propuesto (Christofferson, et al 2010), el cual se encuentra involucrado en diversas patologías y coincide dentro de las vías canónicas del cáncer con la activación anormal de la vía de NF- κ B que permite la regulación negativa de la proteína RIP1, cuyo trabajo final es tomar la decisión de detonar esta forma de muerte celular.

1.1.4 Habilitación de la inmortalidad replicativa

En 1964, Leonard Hayflick realizo una serie de experimentos en donde probó que las células normales poseen un límite en el número de ciclos sucesivos de crecimiento y división, él lo llamo etapa III y sugirió el nombre de senescencia, posteriormente, este límite fue conocido como límite de Hayflick, en su honor. A diferencia de las células diploides, que poseen un número finito de duplicaciones, Hayflick describió que las células con un número modal de cromosomas diferente al común (heteroploides) pueden efectivamente cultivarse sin límite, son “inmortales”, además de que poseen capacidad de formar tumores (Hayflick 1965).

Actualmente se sabe que este límite se encuentra asociado con dos eventos que frenan la proliferación celular, la senescencia, en la cual las células no son

capaces de proliferar, sin embargo se encuentran viables, y la crisis, en la cual las células parecen morir irremediablemente. Sin embargo, en muy raras ocasiones, algunas células que surgen de una población en crisis parecen poseer un potencial de replicación sin límites. Este fenómeno ha sido llamado inmortalización el cual, junto con la transformación, es un rasgo característico de las células cancerosas.

La principal razón de la senescencia y la posterior crisis, son los daños a la integridad cromosomal ocasionado por el acortamiento de los telómeros, los cuales están formados por secuencias de hexanucleótidos repetidos en tándem, estos se acortan progresivamente con cada división celular de las células no inmortalizadas, y eventualmente pierden la habilidad de proteger los cromosomas de daño por fusión y pérdida de información (Ladetto 2010).

Existen evidencias que indican que una forma de contrarrestar estos candados por parte de las células cancerosas es la expresión de la Telomerasa, la cual es una DNA polimerasa especializada en añadir segmentos de DNA a las regiones teloméricas. Esta enzima se encuentra ausente en las células no inmortalizadas, sin embargo se encuentra expresada y funcional en más del 90% de las células espontáneamente inmortalizadas, incluyendo las células de tumores malignos humanos (Micco et al. 2008).

En contraste, se sugiere que al haber señales oncogénicas que incitan a la proliferación y ya comenzada la crisis por el acortamiento de los telómeros, la vía de p53 se encarga de evitar la proliferación de células con daño genómico, sin embargo, existen evidencias que sugieren que en las células que han perdido la vigilancia genómica mediada por p53, y sobreviven a la erosión inicial de los telómeros, comienza un evento conocido como ciclo ruptura-fusión-puente (RFP). Las alteraciones genómicas resultantes de los ciclos RFP incluyen deleciones y amplificaciones de segmentos cromosomales, los cuales, permiten incrementar la mutabilidad del genoma, acelerando la adquisición de oncogenes y supresores de tumores mutados (Artandi y DePinho 2010).

1.1.5 inducción de Angiogénesis

Al igual que los tejidos normales, los tumores requieren nutrientes y oxígeno como sustento, así como una vía que les permita evacuar los desechos metabólicos y el dióxido de carbono. La vasculatura *de novo* asociada al tumor generada por el proceso de angiogénesis le permite cubrir esta necesidad. Durante la embriogénesis, el desarrollo de la vasculatura involucra el nacimiento de nuevas células endoteliales y su ensamblaje, formando la vasculogénesis, en adición a la angiogénesis de nuevos vasos a partir de los existentes. Siguiendo esta morfogénesis, la vasculatura normal permanece quiescente durante mucho tiempo (Hanahan y Folkman 1996).

La angiogénesis puede ser encendida durante la vida adulta, siendo parte de procesos fisiológicos como la cicatrización de heridas y en el ciclo reproductivo femenino, pero solo de manera pasajera. En contraste, durante la progresión tumoral, el interruptor de la circuitería angiogénica permanece siempre encendido, causando que la vasculatura normalmente quiescente continuamente de origen a nuevos vasos que ayudan sustancialmente a que la neoplasia crezca (Hanahan y Folkman, 1996).

Existen muchas evidencias que indican que el interruptor angiogénico se encuentra gobernado por factores compensatorios que permiten la inducción o se oponen a la angiogénesis (Bergers y Benjamin 2003; Baeriswyl y Christofori 2009). Algunos de esos reguladores son proteínas de señalización, estimuladoras o inhibitoras, que se unen a receptores de superficie celular presentados por las células vasculares endoteliales. Los arquetipos de inductores e inhibidores de la angiogénesis mejor conocidos son el factor de crecimiento endotelial A (VEGFA) y la trombospondina (TSP-1), respectivamente.

El gen de VEGFA codifica para un ligando que se encuentra involucrado en la organización de nuevos vasos sanguíneos durante el desarrollo embriogénico y postnatal, y en la supervivencia homeostática de células endoteliales, así como en situaciones fisiológicas y patológicas en el adulto. La señalización de VEGF

mediante tres receptores tipo tirosina quinasa (VEGFR 1–3), se encuentra fuertemente regulada, lo cual refleja la complejidad de su propósito. Así, la expresión de VEGF es regulada por hipoxia y por señalización oncogénica (Ferrara 2010; Mac Gabhann y Popel 2008; Carmeliet 2005).

Adicionalmente, los ligandos de VEGF pueden ser secuestrados en la matriz extracelular en formas latentes y ser sujetos de liberación y activación por proteasas que degradan dicha matriz (Kessenbrock et al. 2010). Otras señales pro-angiogénicas como es el caso de miembros de la familia de factores de crecimiento de fibroblastos (FGF) han sido implicadas en la angiogénesis sostenida cuando su expresión es crónicamente sobre-regulada (Baeriswyl y Christofori 2009).

1.1.6 Activación de Invasión y Metástasis.

En los carcinomas que surgen de tejidos epiteliales que progresan a los grados patológicos más altos de malignidad, reflejada en la invasión local y metástasis distante, las células cancerosas desarrollan alteraciones típicas en su forma así como en la forma de adherirse a otras células y a la matriz extracelular (ECM). Una de las alteraciones mejor caracterizadas involucra la pérdida de E-caderina, una molécula clave en la adhesión entre células. Al formar uniones adherentes entre células epiteliales adyacentes, la E-caderina ayuda a ensamblar la característica formación laminar de las células epiteliales y a mantener la quiescencia de las células dentro de esas estructuras. El aumento en la expresión de E-caderina se ha establecido precisamente como un antagonista de invasión y metástasis, mientras que la reducción de su expresión puede potenciar estos fenotipos (Cavallaro y Christofori 2004), (Berx y van Roy 2009).

Como se describió párrafos arriba, el programa de “transición epitelio-mesénquima” (EMT) se ha asociado con un alto grado de malignidad, este programa celular tiene lugar normalmente durante la embriogénesis, sin embargo ha adquirido un papel central en el estudio de la metástasis y la resistencia a muerte celular programada (Polyak y Weinberg 2009). Uno de los rasgos bien definidos de la EMT en células de carcinomas epiteliales, es la sobre-regulación

de proteínas mesenquimales, como la vimentina, fibronectina y N-caderina, además de la sub-regulación de proteínas epiteliales, como la E-caderina. Estas células adquieren propiedades similares a las células troncales (Sendurai A Mani et al. 2008).

Se han descrito un conjunto de factores transcripcionales, funcionalmente activos durante la embriogénesis, que poseen funciones pleiotropicas, como Snail, Slug, Twist, y Zeb1/2 que son expresados en diferentes combinaciones en diversos tipos de carcinomas y se ha demostrado experimentalmente que son agentes causales en la iniciación del programa de invasión y metástasis (Yang y Weinberg 2008; Schmalhofer et al. 2009; Micalizzi et al. 2010; Taube et al. 2010).

1.1.7 Reprogramación del Metabolismo Energético

Existe una gran cantidad de mecanismos moleculares intrínsecos y extrínsecos que convergen en la alteración del metabolismo central de las células cancerosas, lo cual provee de un adecuado soporte para las tres necesidades básicas de una célula en acelerada división celular: Una obtención rápida de ATP para mantener las condiciones energéticas (Locasale y Cantley 2010); un notable incremento en la síntesis de macromoléculas (Kaelin y Thompson 2010) y el mantenimiento adecuado del estado redox celular (Hamanaka y Chandel 2010).

El fenotipo metabólico observado en células malignas y tumores mejor caracterizado es el efecto Warburg, en donde existe un desplazamiento de la generación canónica de ATP mediante la fosforilación oxidativa hacia la generación de ATP a través de la glicólisis en condiciones normales de oxígeno (Warburg 1956).

Esta glicólisis aeróbica se encuentra asociada a la actividad de oncogenes, como Ras y Myc, además a supresores de tumores mutantes como p53 (Jones y Thompson, 2009; DeBerardinis et al., 2008) cuyas alteraciones en las células cancerosas han sido seleccionadas por las ventajas adaptativas que les confiere. Esta dependencia de la glucólisis puede ser aún más acentuada en las condiciones de hipoxia que operan dentro de muchos tumores: el sistema de respuesta a hipoxia actúa pleiotropicamente para elevando la expresión de

transportadores de glucosa y diversas enzimas de la vía glucolítica (DeBerardinis et al. 2008; Jones y Thompson 2009).

De este modo, tanto la oncoproteína Ras y el estado de hipoxia, de forma independiente, puede aumentar los niveles de HIF1, HIF2 y algunos de los factores de transcripción, que a su vez regulan a la alza la vía glucolítica (Kroemer y Pouyssegur 2008).

1.1.8 Evasión a la Destrucción Inmune

La evasión a la respuesta inmune es una capacidad recientemente aceptada dentro de las características del cáncer y muchos de sus mecanismos representan un gran misterio. Sin embargo, en la clínica es bien clara desde hace mucho tiempo. Los tumores sólidos que logran desarrollarse han logrado, de alguna forma, eludir la detección de varias ramas del sistema inmune o han sido capaces de limitar el alcance del exterminio inmunológico, y evitar su erradicación.

El hallazgo de un monitoreo inmunológico deficiente en los tumores es sumamente claro en ciertos tipos de cáncer en pacientes inmunocomprometidos (Vajdic y van Leeuwen 2009). Sin embargo, en pacientes inmuno competentes, las células cancerosas son capaces de paralizar a los Linfocitos Cito-Toxicos (CTL) y a las células Asesinas Naturales (NK) mediante la secreción de factores inmunosupresivos como TGF-B (Shields et al. 2010; Li Yang et al. 2010).

Otros mecanismos más sutiles operan a través del reclutamiento de células inflamatorias que son activamente inmunosupresoras, incluyendo las células T reguladoras (Tregs) y las células supresoras mieloides derivadas (MDSCs). Ambos mecanismos son capaces de suprimir la acción de los linfocitos citotoxicos (Ostrand-Rosenberg y Sinha 2009).

1.1.9 Características Facultativas

Como se mencionó anteriormente, se cree que la inestabilidad genómica permite la aparición de mutantes inmortales, estas células, generalmente heteroploides, pueden poseer diversos arreglos cromosomales (Artandi y DePinho 2010) que permiten la rápida evolución de las células malignas mediante la activación de

oncogenes y la inhabilitación de genes supresores de tumores como p53 (Negrini et al. 2010). Se ha encontrado evidencia de que la pérdida de una proteína, como pRb (van Harn et al. 2010) o incluso la mutación de un solo alelo de un gen, como en el caso de BRCA1 (Konishi et al. 2011), es capaz de generar inestabilidad genómica que incrementa la malignidad de las células tumorales.

Esta evidencia nos permite concluir que los defectos en la estructura, vigilancia y reparación del genoma son una ventaja selectiva que permite la progresión tumoral, debido a que aceleran la tasa a la cual las células pre-malignas “evolucionan” acumulando genotipos favorables.

Una segunda característica facultativa implica el estado de inflamación crónica que impera en las lesiones malignas, el cual es propiciado por el sistema inmune, esta situación puede promover la progresión tumoral a través de diversos medios, incluyendo factores de crecimiento que sostienen señales de proliferación, factores de sobrevivencia que limitan la muerte celular, factores proangiogénicos, enzimas que modifican la matriz extracelular que pueden facilitar la angiogénesis, invasión y metástasis y señales inductivas que conducen a la activación de la EMT y otros programas (DeNardo et al. 2010; Grivennikov et al. 2010;).

Una de las vías de señalización críticas en la inflamación promotora de carcinogénesis es NF- κ B, primeramente, por su potencial como inductora de la proliferación celular, sobrevivencia, angiogénesis, motilidad y producción de quimosinas y citosinas en conjunto con STAT3 (Yu et al. 2009; Grivennikov Karin 2010) La inactivación de p53 mediada por NF- κ B y sus consecuentes daños al genoma (Colotta et al. 2009); La reprogramación epigenética causada por un gen blanco de NF- κ B, la proteína KDM6B que elimina la metilación de la lisina 27 de la histona H3 (De Santa et al. 2007), además de su potencial para intensificar la señalización de la señal de Wnt/b-catenina (Umar et al. 2009).

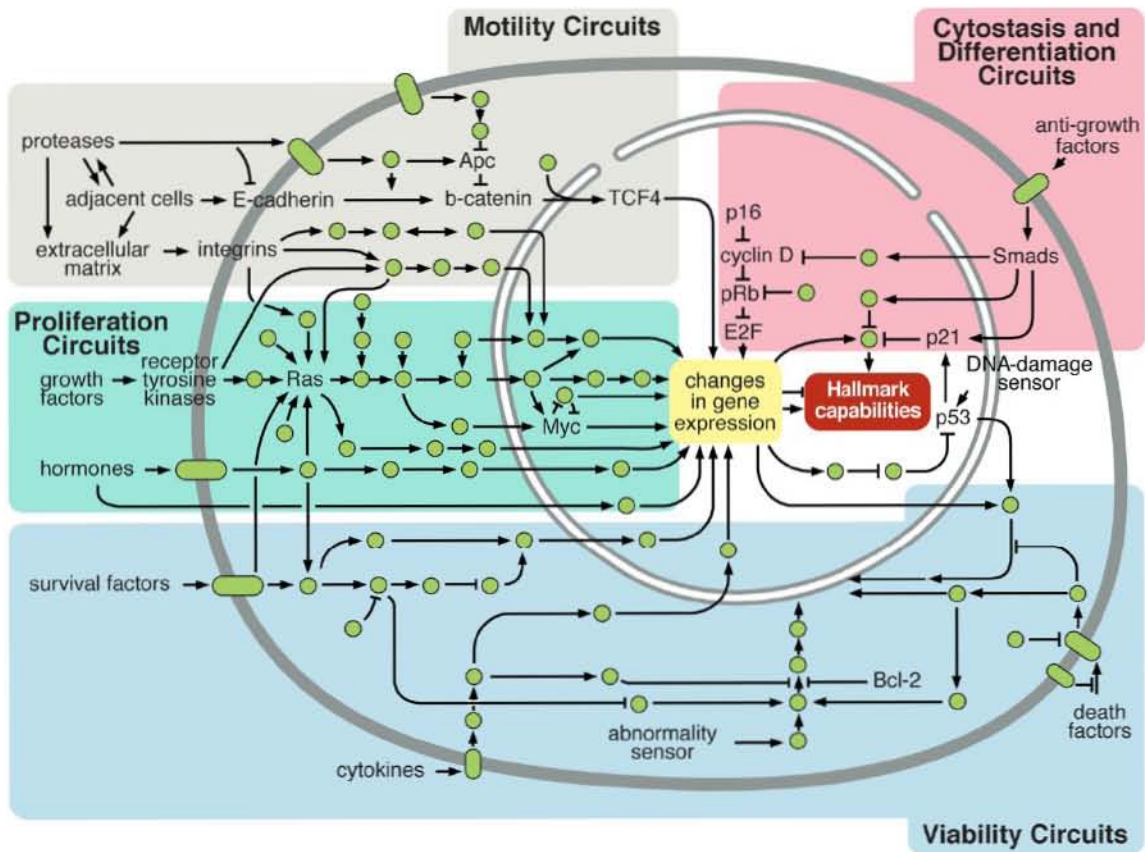


Fig. 3 Redes de señalización intracelular que regulan las operaciones en las células cancerosas. Hanahan y Weimberg (2010)

1.2 Análisis integrativo y biología de sistemas

Con el objetivo de entender la complejidad biológica a diferentes niveles de organización, es necesario integrar el conocimiento generado a partir de diversos tipos de experimentos que permitan recrear el comportamiento del sistema (Mazzocchi 2008; Gray et al. 2010).

Los métodos de identificación molecular masivos, transcriptómica, proteómica y metabolómica, nos permiten realizar una búsqueda masiva de las características que definen un sistema bajo estudio, así como la integración de todo ese conocimiento en modelos simples con un gran poder de predicción. Esos modelos pueden ser contrastados con nuevos datos experimentales para perfeccionarlos.

Actualmente, el reto más grande para la biología de sistemas es el sistema de señalización celular (Kitano 2002; Kitano 2002a; Auffray et al. 2009).

Una vía de señalización consiste de varios eventos secuenciales, los cuales pueden contener modificaciones covalentes, reclutamiento, activación o inhibición alosterica y unión de proteínas (Zubarev et al. 2008). Al identificar una mayor cantidad de interacciones entre distintas vías, se va esfumando la idea de que las señales son transmitidas dentro de circuitos lineales e independientes. La interacción de las vías de señalización se da en el marco de una compleja y robusta red, donde, de alguna forma, una misma señal puede seguir múltiples caminos de acuerdo al contexto celular y una vía no puede ser perturbada sin afectar a las demás (Papin et al. 2005; (Hyduke y Palsson 2010).

La fosforilación reversible en proteínas es la modificación post-traducciona l más usada en la señalización en células humanas, cerca del 1.7% del genoma humano codifican cinasas (Manning et al. 2002), y en un momento dado, cerca del 30% de todas las proteínas celulares pueden estar fosforiladas (Hubbard y Cohen 1993).

Además, la fosforilación puede sugerir que la proteína está o no en su forma activa (como es el caso con diversas enzimas metabólicas), si una señal es activa, o que existe un cambio conformacional que permite alguna actividad específica (Mumby y Brekken 2005).

La interacción de las redes de señalización y regulación genética debe ser entendida como una compleja arquitectura que posee propiedades emergentes, dependientes de la organización espacio-temporal y la compartimentalización de las señales, formando circuitos celulares (Takahashi et al. 2005); (Bauch y Superti-Furga 2006).

Las células son sistemas complejos y dinámicos que utilizan circuitos de señalización molecular para gobernar actividades celulares básicas y coordinar sus acciones (Salinas y Gov 2009). La habilidad de las células para percibir y responder de una forma adecuada al microambiente es la base para la homeostasis, desarrollo, formación y reparación de tejidos e inmunidad. Diversos errores en los procesos del manejo de la información son responsables de diferentes enfermedades, como padecimientos autoinmunes, síndromes

metabólicos, siendo el cáncer la enfermedad sistémica por excelencia (Hornberg et al. 2006; Park et al. 2009; Schadt 2009).

Hornberg (2006) definió que la única forma de comenzar a entender en algún momento el fenómeno complejo que designamos como cáncer, es necesario integrar todo el conocimiento disponible a nivel molecular y celular, para construir un modelo del fenómeno en computadora utilizando los resultados de experimentos cuantitativos. De esta forma, la experimentación precisa y el modelamiento detallado del comportamiento del sistema será capaz de reforzar el uno al otro.

1.3 Análisis Proteómico

El análisis proteómico puede ser conceptualizado como un enfoque experimental que permite explicar la información contenida dentro de las secuencias genómicas en términos de estructura, función y control de procesos biológicos y vías que se llevan a cabo en una célula, tejido u órgano (Gygi y Aebersold 2000).

El proteoma es sumamente dinámico, los tipos de proteínas expresadas, su abundancia, el estado de modificación, su localización subcelular, dependen del estado fisiológico de la célula o tejido. Por lo tanto, el proteoma refleja el estado celular o las condiciones externas encontradas por una célula, y el análisis del proteoma se puede ver como un ensayo de todo el genoma para diferenciar y estudiar estados celulares así como para determinar los mecanismos moleculares que los controlan (Köcher y Superti-furga 2007).

1.4 La complejidad del proteoma

El conocimiento del genoma y sus variantes no permiten comprender los genes a nivel funcional dentro de los genes en los procesos celulares. Las bases biológicas de los procesos celulares no pueden identificarse sólo a partir del estudio del genoma, en particular porque la secuencia de nucleótidos que define a un gen sólo describe el estado estático de la información hereditaria; asimismo, para entender la dinámica de los procesos celulares y la forma en que éstos se alteran en las distintas enfermedades es necesario el estudio de las proteínas y sus

interacciones bajo un estímulo determinado, en razón de que éstas definen la complejidad, el ensamble y el funcionamiento de un organismo en relación con su ambiente.

Se ha estimado que el genoma humano se encuentra conformado por aproximadamente 24 500 genes, a pesar de que sólo se conoce la función de alrededor de 8 000 (Clamp et al. 2007). Sin embargo, se considera que el número de proteínas expresadas puede llegar a ser de 50 000 a 500 000. La abismal diferencia entre el genoma humano y el número de proteínas expresadas se debe a que la mayoría de genes no expresan necesariamente una sola de ellas; se sabe que por procesamiento del mRNA (maduración, edición y splicing alternativo) es posible la obtención de diversas proteínas a partir de un solo gen. Alternativamente, las proteínas pueden sufrir de aproximadamente 400 modificaciones postraduccionales, entre las que destacan la fosforilación, glicosilación, acetilación, desaminación, etc (Blakeley et al. 2010).

1.5 El Cáncer Cérvico Uterino como Modelo de Estudio

El cáncer cérvico uterino (CaCu) representa una importante oportunidad de disectar al proceso carcinogénico, debido a que posee un agente etiológico bien determinado, los Virus de Papiloma Humano de Alto Riesgo (VPH-AR), que se encuentran alrededor del 90.7% de los casos estudiados (Muñoz et al. 2003). Las oncoproteínas *E6* y *E7* de los VPH-AR son capaces de interactuar con los supresores de tumores p53 y pRb además de 300 proteínas conocidas actualmente. Hasta el momento han sido descritos más de 120 tipos de VPH que afectan a humanos, solo una pequeña parte de ellos se consideran de alto riesgo y son asociados con la carcinogénesis. De ellos, los tipos VPH16 y VPH18 son los más prevalentes, presentes en el 54.6% y 11% de los carcinomas cervicales escamosos, respectivamente (Scheffner et al. 1990; Boyer et al. 1996; Yamato et al. 2006).

Los pacientes que presentan tumores infectados con estos dos tipos de VPH son los más ampliamente estudiados, incluso la primera línea celular humana

establecida, HeLa, es positiva para la infección de VPH18. Esta línea celular es la mayor fuente de conocimiento que poseemos sobre biología celular y cáncer (Masters 2002).

Sin embargo, debido a la eliminación espontánea de la infección viral, no todos los pacientes con una infección de VPH-AR desarrollan cáncer. Esto indica que la mayoría de las infecciones son sub clínicas y solo una pequeña fracción puede producir lesiones epiteliales y solo una fracción más modesta de esas lesiones lograra desarrollar cáncer (Molano 2003).

Consecuentemente, la infección de VPH-AR es necesaria pero no suficiente para el desarrollo del CaCu (Perez-Plasencia et al. 2008). Por lo tanto, las causas que permiten el desarrollo de esta enfermedad a partir de una infección con VPH-AR, o incluso sin ella, no son del todo conocidas.

2 Planteamiento del problema

A través de los años, el estudio del cáncer ha acumulado una gran cantidad de información sobre las alteraciones en todos los niveles biológicos en este conjunto de padecimientos.

Sin embargo, el principal reto en la investigación básica y clínica, consiste en un cambio de paradigma en la concepción de esta enfermedad, comenzar a considerarla como un padecimiento sistémico que debe ser abordado desde diversos ángulos y con diferentes aproximaciones que permitan integrar este enorme cuerpo de conocimientos, producto de décadas de intensiva investigación, con las nuevas tecnologías de análisis masivo, y que este conocimiento posea un impacto poderoso en el entendimiento biológico, la capacidad de diagnosis y el poder de prognosis.

El cáncer cérvico uterino es un valioso modelo ya que se conoce el primer hit en el mecanismo celular, la infección por el VPH, lo cual nos otorga un punto de inicio que nos permitiría entender el mantenimiento y el progreso de este trastorno.

2.1 Hipótesis

Debido a que el cáncer es un conjunto heterogéneo de patologías que convergen en un fenotipo común, la proliferación celular sin control debido a alteraciones en diversas vías de regulación, metabólicas, señalización y transducción de señales, es posible que exista un patrón de eventos moleculares que permita entender este padecimiento a nivel sistémico.

2.2 Objetivo General

El objetivo general de este trabajo es realizar un análisis sistémico de líneas celulares de Cáncer Cérvico Uterino (CaCu) mediante la integración de expresión genética, actividad de factores de transcripción, patrones de expresión y fosforilación proteica.

2.3 Objetivos Específicos

2.3.1 Identificar patrones de expresión proteica, ontología génica y vías metabólicas, de regulación y señalización comunes a las seis líneas celulares de CaCu HeLa, SiHa, CaSki, ViBo, C33A y CaLo que permitan identificar factores clave en la biología del CaCu.

2.31.1 Definir los patrones de expresión proteica diferencial de las líneas celulares de CaCu HeLa, SiHa, CaSki, ViBo, C33A y CaLo, en relación a la línea de queratinocitos epiteliales inmortalizados HaCat.

2.3.1.2 Construir una red de interacción proteína-proteína partir de los patrones de expresión proteica que permita definir los actores clave mediante el análisis topológico de la red, así como la identificación de las vías donde participen.

2.3.1.3 Identificar los perfiles de expresión genética y Ontología Génica de la línea celular HeLa con relación a la línea de queratinocitos primarios normales NHEK.

2.3.2 Integrar la expresión génica, las redes de regulación y la expresión de fosfoproteínas dentro de una dinámica de biología de sistemas que analice el flujo de información y nos permita diferenciar los circuitos que dan origen a las células malignas.

2.3.2.1 Reconstruir la red de regulación y factores de transcripción diferencialmente expresados en la línea celular HeLa en relación a epitelio cervical y a la línea de queratinocitos primarios normales NHEK.

2.3.2.2 Identificar las vías metabólicas, de regulación y señalización que presenten alteraciones en su expresión y establecer su relación con el fenotipo maligno de la línea celular HeLa

2.3.2.3 Determinar el patrón de los circuitos celulares que son determinados por expresión genética, redes regulatorias y vías diferenciales en la línea celular HeLa, además de validar los modelos mediante la identificación de proteínas fosforiladas.

3. Materiales y Métodos

3.1 Cultivo Celular

Las líneas celulares CaSki, HeLa, SiHa, C-33A, ViBo y CaLo fueron donadas por el laboratorio de oncología del Centro Médico Siglo XXI del Instituto Mexicano del Seguro Social (IMSS). La línea celular HaCat fue donada por el Centro de Investigación Sobre Enfermedades Infecciosas, el cual pertenece al Instituto Nacional de Salud Pública. Todas las líneas celulares fueron cultivadas en medio libre de suero RPMI-advanced 1640 con rojo fenol (Gibco BRL, USA) y solución antibiótica-antimicótica (10,000 unidades de penicilina, 10 mg de estreptomina y 25 ug de anfotericina B por mL) (Invitrogen, Carlsbad, CA), suplementado con 1% de suero fetal bovino (Invitrogen), y 200 mM de GlutaMax (Invitrogen).

Las células fueron incubadas en un ambiente con 5% de CO² y saturación de humedad a 37°C en frascos de cultivo de 75 cm² (Nalgene, Nunc International, Rochester, NY). Las células fueron cosechadas al 70% de confluencia con solución de Verseno (Tris base 25mM, NaCl 136.8 mM, KCl 5.36 mM, EDTA 1mM pH7.7) y lavadas tres veces en buffer de fosfatos salinos (0.1M Fosfato de Sodio y 0.15 M NaCl por litro, pH 7.2).

3.2 Análisis Proteómico

La extracción de proteínas y los 2D SDS-PAGE fueron realizados como se ha descrito previamente (Salazar et al. 2010), fueron teñidos usando coomasie coloidal (Candiano et al. 2004) y la imagen digital fue obtenida usando un densitómetro GS-800 densitómetro (Bio-Rad, Hercules, CA). Las imágenes digitales fueron analizadas y comparadas usando el software PDQuest 8.0.1 (Bio-Rad). Cada experimento fue realizado por triplicado.

Una vez que la imagen digital de cada gel fue comparada contra el resto, las entidades electroforéticas de interés fueron cortadas, alquiladas, reducidas, digeridas y transferidas a una placa blanco de análisis de MALDI utilizando los robots Proteineer SP II y SP, usando el software SPcontrol 3.1.48.0 v(Bruker Daltonics, Bremen, Germany), mediante el kit DP Chemicals 96 gel digestion kit (Bruker Daltonics) y procesadas en un MALDI-TOF Autoflex (Bruker Daltonics) Fueron realizados 100 disparos satisfactorios en 20 “shotsteps”, el umbral de resolución de pico fue establecido a 1500, la tasa de señal/ruido de tolerancia fue 6, y no se excluyeron contaminantes. El espectro fue anotado mediante el software flexAnalysis 1.2 v SD1 Patch 2 (Bruker Daltonics) y se utilizó el motor de búsqueda MASCOT (Perkins et al. 1999) para comparar los “mass fingerprints” contra la base de datos UniProt (The UniProt Consortium 2010) lanzamiento 2010_09 usando los siguientes parámetros: Taxon-Human, mass tolerance 500 ppm, un miss-cleavage permitido, como modificación fija Carbamidometil y oxidación de metionina como modificación variable.

3.3 Reconstrucción de redes

La reconstrucción de redes fue realizada con la ayuda del “plugin” de Cytoscape (Shannon et al. 2003), Bisogenet (Martin et al. 2010), usando las proteínas identificadas como nodos y añadiendo vértices con los siguientes criterios: Organism> Homo sapiens, protein identifiers only; Data Settings>protein-protein interactions; all data sources y all experimental methods; method> By adding edges connecting input nodes y as Output> Proteins.

3.4 Extensión de redes

La red primaria obtenida con anterioridad fue extendida mediante la técnica bioinformática de la carnada, usando el “plugin” Bisogenet, de Cytoscape, con la opción de expandir la red y el parámetro: Añadir vecinos de nodos ingresados a la distancia de uno. La red que resulto fue sujeta a análisis de mediciones topológicas de centralidad utilizando el plugin CentiScaPe1.1 (Scardoni et al. 2009) de Cytoscape.

3.5 Analisis de promotores

Se descargó la tabla de anotación de genes Gencode Genes-ENCODE (wgEncodeGencodeManual V3) (Raney et al. 2011)) del sitio web del Navegador del Genoma perteneciente a la Universidad de California Santa Cruz. A partir de esta tabla se recuperó el sitio de inicio de la transcripción de todos los genes GENCODE con una puntuación superior a 750, así como su identificador de Ensembl (Hubbard et al. 2009) y el nombre de gen. Debido a que algunos transcritos presentan varios sitios de inicio cercanos, se realizó un filtrado que sólo considero sitios de inicio ubicados a una distancia mayor de 500 pb, lo que redujo la cantidad de lecturas a 18,534. Las secuencias promotoras fueron definidas entre 700 pb corriente arriba y 300 pb corriente abajo del sitio de inicio de la transcripción. Las ubicaciones de promotor correspondientes a los genes de las proteínas de nuestra red se identificaron basándose en el ID de transcripción Ensembl, así como los nombres de genes y sinónimos indicados en los datos UniProt de esas proteínas.

La información de picos fue descargada desde el Sitio de unión de factores de transcripción de la universidad de Yale (TFBS) el cual es una pista del proyecto ENCODE teniendo preferencia por los experimentos pertenecientes a la línea celular HeLa-S3. Un pico se considera en el interior del promotor si el punto medio del pico estaba dentro de los 1.000 pb reportados como la secuencia del promotor. Hipergeométricas pruebas se realizaron con el fin de evaluar la importancia de

estos hallazgos. Tamaño de la muestra se consideró como el número de promotores que contienen TFBS.

3.6 Análisis de enriquecimiento de vías, ontología génica y cromosomas.

Se realizó el análisis de enriquecimiento de “pathway-based sets” de las proteínas considerando todos los nodos de nuestra red extendida. Dicho análisis fue realizado mediante la herramienta “web” ConsensusPathDB (Kamburov et al. 2011) del Instituto de Genética Molecular Max Planck, usando el análisis de sobre-representación. La búsqueda fue definida contra la base de datos KEGG y PID (Kanehisa et al. 2010; Schaefer et al. 2009).

El análisis de enriquecimiento de cromosomas y citobandas fue realizado usando los transcritos sobre y sub expresados. Dicho análisis fue realizado mediante la herramienta “web” DAVID Bioinformatics Resources 6.7 del National Institute of Allergy and Infectious Diseases (NIAID), NIH (Huang et al. 2009).

De igual forma, empleando la misma herramienta, realizamos un análisis de enriquecimiento basado en Ontología Génica (Harris et al. 2004) nivel 3, en la categoría de proceso biológico, se utilizó este nivel debido a que no contiene gran redundancia, y permite encontrar los procesos con suficiente detalle.

3.7 Cuantificación de genes y transcritos a partir de datos de RNA-seq

Los datos fueron obtenidos y procesados como fue descrito (Nagaraj et al. 2011). Las lecturas procesadas fueron alineadas al genoma humano de referencia (hg19/GRCh37, excluyendo haplotipos adicionales) usando TopHat v1.4.1 (Trapnell et al. 2009) y los transcritos y genes fueron tomados de Ensembl release 59 y fueron cuantificados usando Cufflinks v0.8.3 (Trapnell et al. 2012).

3.8 Análisis de expresión diferencial a partir de datos de RNA-seq

Los datos de “RNA-seq paired-end” de las líneas celulares HeLa-S3 y NHEK fueron descargados de la página web del proyecto ENCyclopedia Of DNA Elements (ENCODE) (Raney et al., 2011) (UCSC accession numbers

wgEncodeEH000130 y wgEncodeEH000131, respectivamente). Los archivos Fastq de dos librerías HeLa-S3 75x75 paired-end RNA-seq (números experimentales 10881, 10882) y dos librerías de NHEK 75x75 paired-end RNA-seq (números experimentales 10884, 11586) fueron alineados contra la versión hg19 del genoma humano de referencia usando TopHat v1.4.1 y Gencode annotation Version 12. El índice del genoma humano fue construido utilizando bowtie v0.12.7. Los genes diferencialmente expresados entre NHEK y HeLa fueron identificados usando cuffdiff v 1.3.0. Se estableció un umbral de p-valor del 0.01 para todos los genes diferencialmente significativos.

3.9 Análisis de factores transcripcionales

Se descargaron conjunto de datos del Gen Expression Omnibus (GEO) (Barrett et al. 2011) de la plataforma de microarreglos Affymetrix del “chip” HG-U133A pertenecientes a la línea celular HeLa y epitelio cervical normal con números de acceso GSM246123 y GSM246422, respectivamente los archivos tipo “.cel” fueron cargados a la herramienta “web” MARA (The Fantom Consortium) para normalizar, asignar sitios promotores de Pol III y de unión dentro del conjunto de sondas presentes en el microarreglo y realizar el análisis de actividad de TF.

3.10 Asociación de los blancos transcripcionales con los niveles de expresión de mRNA

Se asoció la información sobre los genes, que fueron reportados como blancos para cada TF diferencialmente activado, como es reportado por MARA, con los niveles de expresión particulares de acuerdo al análisis de expresión diferencial. Reportes individuales de MARA para cada gen fueron analizados mediante programas “ad-hoc” escritos en Perl 5.12.4. Esta información procesada, así como los datos del análisis RNA-Seq, fueron vertidos en una base de datos relacional construida en MySQL Server 5.5 (Community Edition). Subsecuentemente, se condujo una búsqueda de cada TR relevante relacionada a sus blancos.

3.11 Enriquecimiento de proteínas fosforiladas

La proteína extraída, como previamente se ha mencionado, se resuspendió en buffer de incubación y posteriormente se realizó la cromatografía por afinidad a óxido de metal (MOAC) (Wolschin et al. 2005). La fracción enriquecida en fosfoproteínas fue eluida y precipitada (Wessel y Flügge 1984), a continuación, con estas muestras se realizó una electroforesis de una dimensión en geles de poliacrilamida y dodecilsulfato de sodio (SDS-PAGE). Los geles fueron teñidos con Coomassie coloidal. Todas las bandas se escindieron con un bisturí para la posterior digestión triptica.

3.12 Separación cromatográfica

Los péptidos digeridos (8 μ l) fueron desalados y concentrados en columnas Zorbax (Agilent 5065-9913), posteriormente se realizó un análisis en una columna de fase reversa (Agilent Zorbax 300SB C18, 3.5 μ m, 150X0.075 mm). La separación se realizó a 400 nL/min usando un gradiente lineal. La fase móvil A consistió en una solución de ácido fórmico al 0.1% en agua, la fase móvil B fue una solución de ácido fórmico al 0.1% en acetonitrilo. Las condiciones del gradiente en el análisis cromatográfico fueron establecidas de la siguiente manera: De 95% (0 min) a 95% (14min); De 95%(14min) a 60% (54 min); De 60% (54 min) a 20% (56 min); De 20% (56min) a 20% (61 min); De 20% (61min) a 95% (62 min); y de 95% (62min) a 95% (72 min).

3.13 Análisis de MS/MS

Los péptidos fueron analizados por MS/MS utilizando un cromatógrafo de nanoflujo (Agilent 1100 nano pump G2226A) acoplado a una trampa de iones híbrida de triple cuadrupolo (QTRAP 3200, AB Sciex), equipada con una fuente Nanospray II y utilizando Información Dependiente de Adquisición (IDA). La determinación del ión precursor se realizó utilizando una búsqueda tipo "Enhanced MS scan", sobre un rango de masa de 300-1600 m/z a 4,000 amu/s, con un voltaje de ionizador de spray de 3300 aplicado a Picotip FS360-75-15-N con nitrógeno como gas ionizador de spray. Los iones precursores se colisionaron en Q2 usando

energía de colisión tipo “rolling” (con un máximo permitido de CE=80). Las búsquedas de productos iónicos (MS/MS) se realizaron sobre un rango de masas de 100-1700 m/z a 1000 amu/sec, y los voltajes de colisión fueron determinados dinámicamente. Todas las proporciones de masa/carga de los iones precursores fueron confirmadas. La identificación de las fosfoproteínas se llevó a cabo usando el algoritmo MASCOT, utilizando la base de datos de UniProt, usando los parámetros de digestión trípica, MS/MS búsqueda iónica, aplicación de masa monoisotópica, masa de proteína no restringida, tolerancia de masa peptídica de ± 1.2 Da, tolerancia de masa de fragmentos de ± 0.6 Da y un máximo de fragmentos perdidos de 1.

4. Resultados

4.1 Existe un patrón de expresión proteica conservado en las líneas celulares de cáncer cérvico uterino

Se obtuvieron extractos proteicos totales de seis líneas celulares de cáncer cérvico uterino: dos líneas positivas a la infección con VPH tipo 18 (HeLa y CaLo), dos positivas para VPH tipo 16 (SiHa y CasKi), así como dos negativas a la infección por VPH (ViBo y C-33A); y HaCat, una línea celular de queratinocitos epidermales inmortalizados espontáneamente, fue usada como control. Todas las líneas celulares mencionadas fueron analizadas mediante 2D-PAGE (Figura 4).

Este diseño experimental fue concebido para permitir la identificación de las proteínas comunes al fenotipo maligno independientemente de la presencia de variaciones intrínsecas, tales como la presencia del VPH. Utilizamos la línea celular HaCaT como control negativo, ya que es de origen epitelial y tiene un potencial de replicación ampliado, lo que nos permite distinguir entre los eventos biológicos de la transformación y la inmortalidad. Además, de que se ha utilizado ampliamente como control en varios estudios de cáncer cervical (Yoon Pyo Choi et al. 2005)

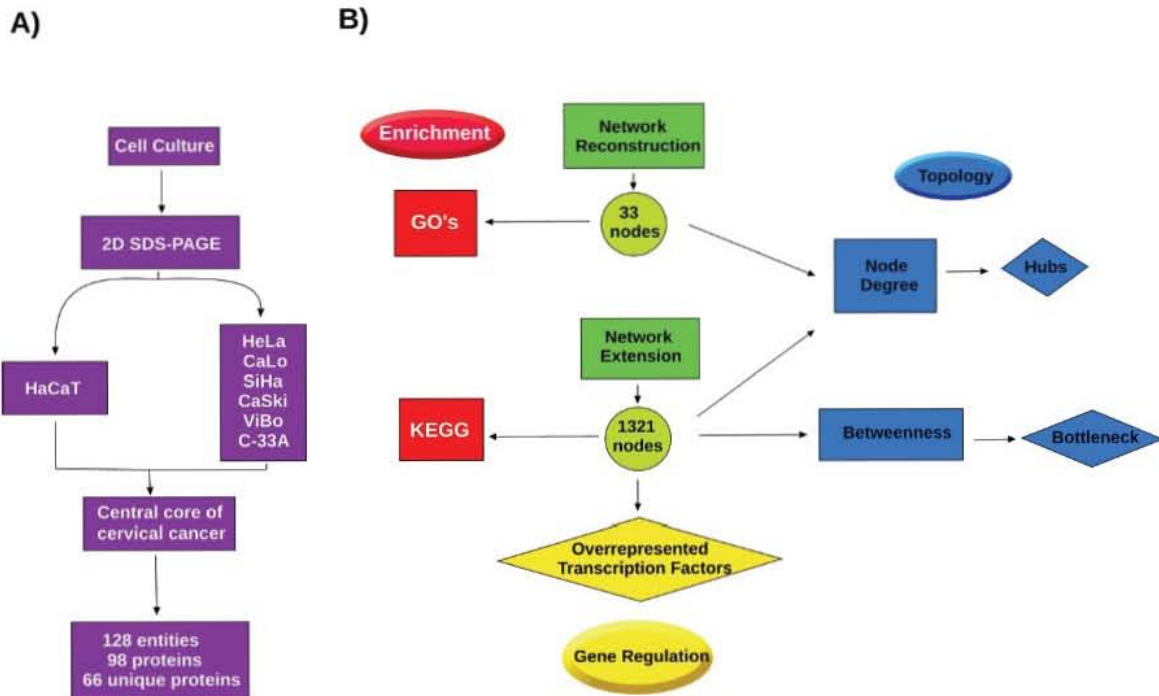


Fig. 4 Se compararon los perfiles de proteínas de seis líneas celulares de cáncer cervical por 2D SDS-PAGE, y se estableció un conjunto de proteínas comunes a ellas que no encontraban en nuestro control. Este conjunto de proteínas fue llamado "núcleo central de cáncer cérvico uterino", (a). A partir de este núcleo, fue reconstruida una red PPI (b), Esta red fue expandida y se utilizó para obtener un enriquecimiento de GO y vías celulares. Finalmente, se llevó a cabo un análisis de los factores de transcripción contenidos en la red extensa utilizando un análisis de ChIP-Seq mediante la base de datos del proyecto ENCODE.

Se realizaron tres análisis independientes y fueron utilizados para definir la expresión proteica en cada una de las líneas celulares analizadas. En promedio, se resolvieron más de 1000 entidades electroforéticas en cada gel 2-DE. Se normalizó la expresión proteica de acuerdo a su densidad óptica (OD). Centramos nuestro análisis en las entidades electroforéticas que se observaron presentes en todas las líneas celulares de cáncer, pero que no fueron observadas en el control HaCat, o aquellas que sufrieron una sobre-regulación de manera significativa (más de 2 veces). Llamamos a este grupo de 127 entidades electroforéticas el "grupo central de cáncer cérvico uterino". De los 127 spots seleccionados, 98 proteínas fueron identificadas mediante espectrometría de masas tipo MALDI-TOF, identificando un total de 66 identidades proteicas únicas (Tabla 1).

Las 66 proteínas únicas fueron clasificadas mediante un análisis de Ontología Génica usando ConsensusPathDB. Las identidades de las proteínas que conforman el grupo central confirman la calidad biológica del mismo, debido a que existe una fuerte correlación entre la expresión alterada de estas proteínas en diversos tipos de cáncer. Con referencia a las funciones encontradas en literatura, las proteínas identificadas fueron divididas en al menos tres grupos:

El primer grupo está conformado por proteínas que se relacionan con migración celular y metástasis, entre ellas anexina 2, que desempeña un papel crucial en el establecimiento de la metástasis en cáncer de próstata participando en la regulación de la adhesión y migración de las células cancerosas a osteoblastos y células endoteliales (Shiozawa et al. 2008); proteínas disulfuro isómeras que se han encontrado fuertemente expresadas en células de glioma invasivo, en las cuales su inhibición conduce a reducir dicho fenotipo (Goplen et al. 2006); vimentina que es una proteína filamentosa intermedia del citoesqueleto y marcador de la Transición Epitelio Mesenquima (EMT), contribuye a la invasión y metástasis en cáncer de próstata (Melissa G Mendez et al. 2010); la ezrina es requerida para la invasión y metástasis en cáncer de seno (Elliott et al. 2005) y finalmente vinculina, que facilita la generación de fuerza contráctil, aumentando la movilidad e invasión (Mierke et al. 2010). Este conjunto de proteínas posiblemente desempeñen un papel en la invasión y metástasis en el cáncer cérvico uterino, similar al que desempeñan en otros tipos de cáncer.

En un segundo grupo, encontramos proteínas relacionadas con la evasión de la apoptosis GRP78, HSP71, HSP7C, HS90B y GRP75. Estas moléculas son activadas como parte del programa de Respuesta a Proteínas no Plegadas (UPR), este programa se ha relacionado a sobrevivencia, proliferación y angiogénesis (Dong et al. 2008). De igual forma, la sobre-expresión de GRP78 es suficiente para conferir a la célula resistencia a la apoptosis en al menos dos tipos celulares, de una forma independiente a su función dentro de la UPR (Ramachandra K Reddy et al. 2003). Estas proteínas pueden funcionar como un enlace entre la evasión de la apoptosis, angiogénesis y proliferación.

Accession number	Name	Score	Expect	Searched-Matched
GRP78_HUMAN	78 kDa glucose-regulated protein	391	1,60E-035	91 – 46
ENOA_HUMAN	Alpha-enolase	327	4,00E-029	82 – 34
GRP78_HUMAN	79 kDa glucose-regulated protein	302	1,30E-026	64 – 32
HSP7C_HUMAN	Heat shock cognate 71 kDa protein	294	8,10E-026	70 – 39
VIME_HUMAN	Vimentin	290	2,00E-025	64 – 33
ENOA_HUMAN	Alpha-enolase	286	5,10E-025	82 – 34
PDIA1_HUMAN	Protein disulfide-isomerase	252	1,30E-021	77 – 30
ANXA2_HUMAN	Annexin A2	249	2,60E-021	65 – 31
TBB5_HUMAN	Tubulin beta chain	247	4,00E-021	73 – 32
ANXA4_HUMAN	Annexin A4	233	1,00E-019	50 – 25
EFTU_HUMAN	Elongation factor Tu, mitochondrial	217	4,00E-018	44 – 22
TPIS_HUMAN	Triosephosphate isomerase	206	5,10E-017	42 – 17
VIME_HUMAN	Vimentin	206	5,10E-017	62 – 23
PGAM1_HUMAN	Phosphoglycerate mutase 1	205	2,10E-016	41 – 18
ATPB_HUMAN	ATP synthase subunit beta, mitochondrial	193	1,00E-015	84 – 44
ACTG_HUMAN	Actin, cytoplasmic 2	192	1,30E-014	50 – 20
PUR9_HUMAN	Bifunctional purine biosynthesis protein PURH	191	1,60E-015	49 – 22
KCRB_HUMAN	Creatine kinase B-type	178	3,20E-014	50 – 19
EF2_HUMAN	Elongation factor 2	154	8,10E-012	67 – 26
HYOU1_HUMAN	Hypoxia up-regulated protein 1	154	8,10E-012	38 – 19
KCRB_HUMAN	Creatine kinase B-type	154	8,10E-012	50 – 19
PPIA_HUMAN	Peptidyl-prolyl cis-trans isomerase A	150	2,00E-011	45 – 16
GELS_HUMAN	Gelsolin	147	4,00E-011	46 – 19
VINC_HUMAN	Vinculin	145	6,40E-011	36 – 19
WDR1_HUMAN	WD repeat-containing protein 1	144	8,10E-011	46 – 16
EZRI_HUMAN	Ezrin	131	1,60E-009	60 – 26
HSP71_HUMAN	Heat shock 70 kDa protein 1A/1B	129	2,60E-009	42 – 21
TBA1C_HUMAN	Tubulin alpha-1C chain	127	4,00E-009	45 – 13
PRDX1_HUMAN	Peroxiredoxin-1	126	5,10E-009	53 – 18
PDIA3_HUMAN	Protein disulfide-isomerase A3	125	6,40E-009	47 – 15
GRP75_HUMAN	Stress-70 protein, mitochondrial	123	1,00E-008	81 – 34
DHSA_HUMAN	Succinate dehydrogenase flavoprotein subunit	122	3,30E-007	39 – 17
RSSA_HUMAN	40S ribosomal protein SA	121	1,60E-008	37 – 11
HS90B_HUMAN	Heat shock protein HSP 90-beta	119	2,6E-08	44 – 19
ESTD_HUMAN	S-formylglutathione hydrolase	114	8,10E-008	25 – 11
TKT_HUMAN	Transketolase	114	8,10E-008	52 – 19
LDHB_HUMAN	L-lactate dehydrogenase B chain	112	1,30E-007	43 – 20
AL1A1_HUMAN	Retinal dehydrogenase 1	110	2,00E-007	58 – 19
IPYR_HUMAN	Inorganic pyrophosphatase	110	2,00E-007	22 – 9
DDX3X_HUMAN	ATP-dependent RNA helicase DDX3X	108	3,20E-007	23 – 13
DHE3_HUMAN	Glutamate dehydrogenase 1, mitochondrial	107	4,00E-007	38 – 13
G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase	107	4,00E-007	55 – 13
ACON_HUMAN	Aconitate hydratase, mitochondrial	105	6,40E-007	54 – 23
ML12A_HUMAN	Myosin regulatory light chain 12A	105	6,40E-007	32 – 11
RIR1_HUMAN	Ribonucleoside-diphosphate reductase large subunit	101	1,60E-006	28 – 15
EF2_HUMAN	Elongation factor 2	99	2,60E-006	60-20
QCR1_HUMAN	Cytochrome b-c1 complex subunit 1, mitochondrial	98	3,30E-006	45 – 15
CNN2_HUMAN	Calponin-2	96	4,50E-006	51 – 13
PCBP1_HUMAN	Poly(rC)-binding protein 1	94	8,30E-006	39 – 14
IMMT_HUMAN	Mitochondrial inner membrane protein	92	1,30E-005	34 – 13
CAPG_HUMAN	Macrophage capping protein	91	1,80E-005	41 – 11
TCTP_HUMAN	Translationally-controlled tumor protein	91	1,80E-005	19 – 8
PSB4_HUMAN	Proteasome subunit beta type-4	89	1,70E-003	20 – 7
TERA_HUMAN	Transitional endoplasmic reticulum ATPase	88	3,10E-005	49 – 15
K2C1_HUMAN	Keratin, type II cytoskeletal 1	87	3,70E-005	46 – 13
LEG1_HUMAN	Galectin-1	84	7,70E-005	41 – 8
1433Z_HUMAN	14-3-3 protein zeta/delta	80	2,30E-004	43 – 14
LMNA_HUMAN	Prelamin-A/C	78	3,40E-004	37 – 11
HNRPL_HUMAN	Heterogeneous nuclear ribonucleoprotein L	77	4,10E-004	27 – 13
TCPH_HUMAN	T-complex protein 1 subunit eta	77	4,00E-004	36 – 15

*Tabla 1 Proteínas identificadas como miembros del "núcleo central del cáncer cérvico uterino". El nombre de la proteína correspondiente a la base de datos UniProt, la puntuación y el valor esperado fueron obtenidos del motor de búsqueda MASCOT. El valor esperado es el número de correspondencias con resultados iguales o mejores que de los esperados al azar. La puntuación $-10 * \text{LOG}_{10}(P)$, donde P es la probabilidad absoluta. Cuanto menor es la expectativa de valor, la puntuación es mejor.*

En un tercer grupo, se ubicaron proteínas involucradas en metabolismo central, como la gliceraldehído 3 fosfato deshidrogenasa, fosfoglicerato mutasa 1, enolasa A, triosafosfato isomerasa y L-lactato deshidrogenasa B. La expresión de este grupo de proteínas puede deberse a las necesidades energéticas propias del metabolismo de las células cancerosas. El incremento en el metabolismo celular ha sido descrito, con un especial énfasis en el aumento en la ingesta de glucosa en comparación con las células normales (Resendis-Antonio et al. 2010).

Otra proteína miembro del grupo de metabolismo central es la galectina-1, la cual ha sido reportada como un factor angiogénico y ha sido encontrada en la vasculatura de algunos tipos de tumores humanos, incluyendo cabeza y cuello, colon, hígado, próstata y cánceres orales (Jung et al. 2007; Thijssen et al. 2010). Todas estas proteínas, identificadas como el grupo central de CaCu se encuentran resumidas en la Tabla 1.

De estos resultados podemos concluir que existe una expresión proteica diferencial con respecto a la línea control y común en las seis líneas celulares procedentes de cáncer cérvico uterino, y que existe una fuerte correlación entre esta expresión anormal y diversos fenotipos asociados a numerosos tipos de cáncer.

4.2 1433Z es un “hub” dentro de la red de interacción proteína-proteína

Con el propósito de ampliar nuestra perspectiva y aumentar la fuerza de la significancia estadística de nuestros resultados, realizamos un interactoma teórico (Figura 5) utilizando el plugin de Cytoscape, Bisogenet, para añadir interacciones experimentalmente verificadas entre las 66 proteínas identificadas.

Lográndose integrar dentro de la red la señal de 33 proteínas, debido probablemente a la falta de evidencia experimental o a la falta de interacciones directas entre nuestras proteínas. Notablemente, debido al análisis de las propiedades topológicas de la red, se encontró que la proteína 1433Z es el nodo con un mayor número de conexiones.

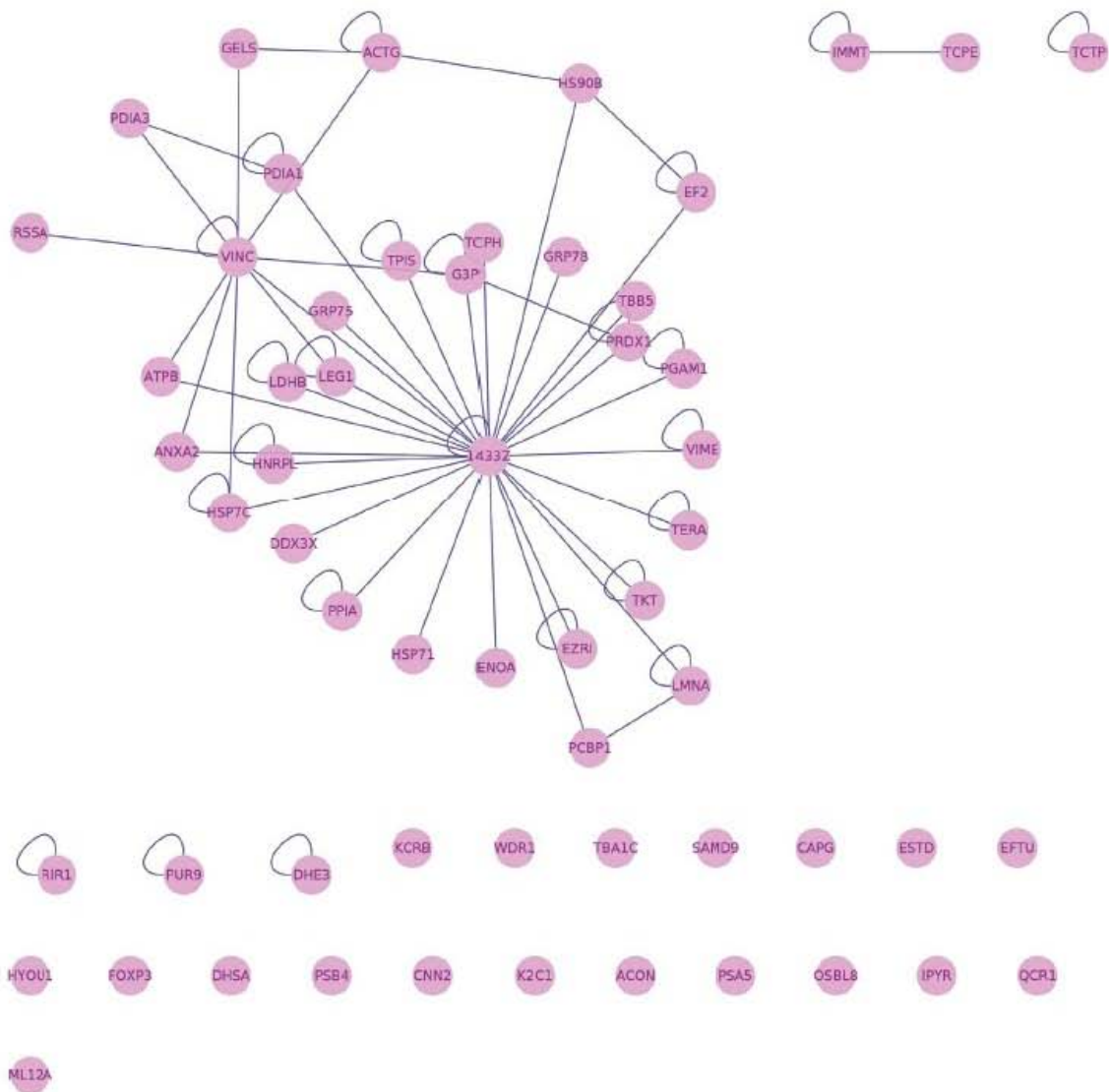


Fig. 5 Reconstrucción de red de proteínas expresadas diferencialmente en 6 líneas celulares de cáncer cérvico uterino, en comparación con un control no maligno. Las conexiones entre proteínas representan interacciones proteína-proteína verificadas experimentalmente entre 33 de las 66 proteínas identificadas en este estudio. Las proteínas no incluidas en esta red no poseen interacciones experimentalmente comprobadas. La red fue construida utilizando Cytoscape y el plugin Bisogenet.

Con el objeto de encontrar posibles vías de interacción, se realizó la expansión de la red, usando las 66 proteínas originales como carnada, buscando como criterio de expansión la interacción con comprobación experimental y la conexión vía el camino más corto con el valor de 1 de nuestras proteínas. El resultado fue una red de 1321 nodos y 9666 aristas (Figura 6)

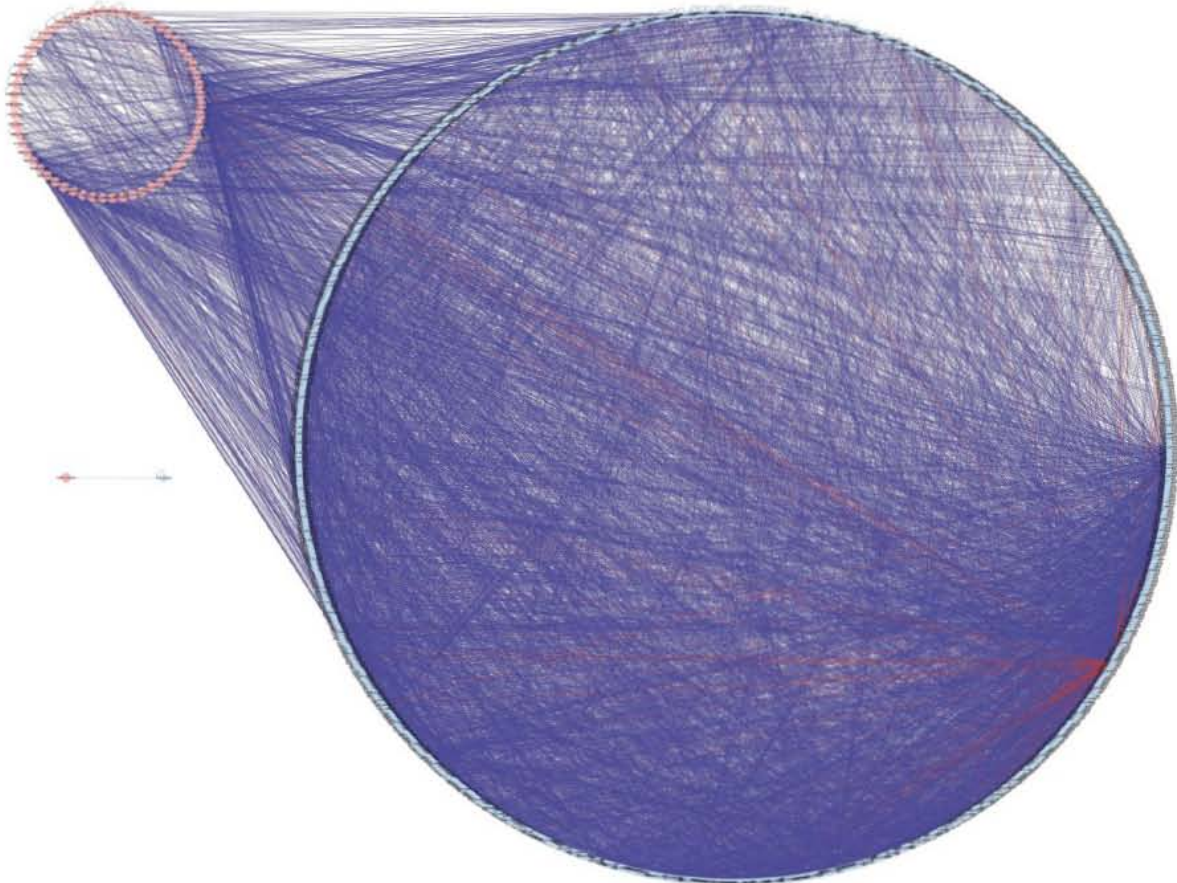


Fig. 6 Ampliación de la red . Las proteínas pertenecientes a nuestro núcleo de expresión diferencial fueron usadas como cebo para obtener interacciones proteína-proteína probadas experimentalmente mediante la adición de nodos vecinos a una distancia de uno. Esto resultó en una red de 1321 nodos y 9666 bordes. La red se expandió utilizando Cytoscape y el plugin Bisogenet.

Un análisis de la distribución de la conectividad, midiendo el “node degree” y el “betweenness”, permitió identificar a las proteínas que poseen una gran interacción con la red, definidas como los “hubs” de la red. Nuevamente la proteína transductora de señales 1433Z, seguida por 1433G son los nodos más conectados. Una razón para utilizar la medición de la propiedad topológica conocida como “betweenness” es porque nos permite explorar la conexión entre la estructura de la red local y la topología global de la red, representando una medida de la influencia que posee un nodo sobre la transferencia de la información (Joy et al. 2005; Hao et al. 2009). Las proteínas más importantes de acuerdo a nuestras mediciones fueron 1433Z, vimentina y vinculina, las cuales forman parte de nuestro grupo central.

El análisis de enriquecimiento de ontología génica nivel 3, categoría de procesos biológicos, muestra 16 GO's sobre-representados (Tabla 2); el análisis de vías de KEGG muestra 16 vías significativas (Tabla 3). Sin embargo, a pesar del gran valor del análisis y del significado biológico de las proteínas identificadas, la búsqueda de GO's es muy general e incluye proceso o vías que comparten proteínas en común con otros proceso que podrían estar o no relacionados con el fenotipo que nosotros deseamos describir (Tabla 3).

<i>gene ontology term</i>	<i>set size</i>	<i>candidates contained</i>	<i>p-value</i>	<i>q-value</i>
GO:0006986 response to unfolded protein	73	8	3,11E-009	7,03E-007
GO:0070841 inclusion body assembly	13	4	4,26E-007	4,82E-005
GO:0006091 generation of precursor metabolites and energy	363	11	2,02E-006	1,52E-004
GO:0032507 maintenance of protein location in cell	80	6	3,14E-006	1,77E-004
GO:0016052 carbohydrate catabolic process	132	7	4,67E-006	1,93E-004
GO:0051651 maintenance of location in cell	87	6	5,13E-006	1,93E-004
GO:0012501 programmed cell death	1338	20	7,18E-006	2,32E-004
GO:0010941 regulation of cell death	1136	18	1,06E-005	2,99E-004
GO:0046907 intracellular transport	855	15	2,08E-005	5,21E-004
GO:0002349 histamine production involved in inflammatory response	2	2	2,56E-005	5,25E-004
GO:0043455 regulation of secondary metabolic process	2	2	2,56E-005	5,25E-004
GO:0002443 leukocyte mediated immunity	184	7	4,05E-005	7,64E-004
GO:0044248 cellular catabolic process	1028	16	4,56E-005	7,92E-004
GO:0050777 negative regulation of immune response	40	4	4,91E-005	7,93E-004
GO:0070727 cellular macromolecule localization	610	12	5,29E-005	7,98E-004
GO:0051235 maintenance of location	133	6	5,79E-005	8,18E-004
GO:0002252 immune effector process	274	8	7,31E-005	9,72E-004

Tabla 2 Enriquecimiento de Ontología Genica nivel 3, usando como categoria procesos biológicos. El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Por lo tanto, estos resultados nos muestran que la proteína transductora 1433Z posee propiedades topológicas que la evidencian como trascendental en la red de expresión proteica y sugieren que posee una fuerte participación en el fenotipo neoplásico, al igual que las proteínas vinculina y vimentina, las cuales se sabe que participan en diversos tipos de cáncer, además de que logramos correlacionar el perfil de expresión proteica aberrante con los análisis de ontología génica y de vías de KEGG.

<i>pathway name</i>	<i>set size</i>	<i>candidates</i>		<i>p-value</i>	<i>q-value</i>	<i>pathway source</i>
		<i>contained</i>				
Small cell lung cancer	84	37		8,67E-009	1,43E-006	KEGG
Protein processing in endoplasmic reticulum	164	57		4,21E-008	3,47E-006	KEGG
Pathways in cancer	325	94		8,81E-008	4,05E-006	KEGG
Apoptosis	87	36		9,82E-008	4,05E-006	KEGG
Neurotrophin signaling pathway	125	46		1,24E-007	4,10E-006	KEGG
T cell receptor signaling pathway	108	41		2,25E-007	6,19E-006	KEGG
Focal adhesion	199	62		9,45E-007	2,23E-005	KEGG
Prostate cancer	89	33		6,11E-006	1,26E-004	KEGG
Pathogenic Escherichia coli infection	54	23		1,13E-005	2,07E-004	KEGG
Endometrial cancer	52	22		1,99E-005	3,28E-004	KEGG
Shigellosis	61	24		3,61E-005	5,42E-004	KEGG
Non-small cell lung cancer	54	22		4,03E-005	5,54E-004	KEGG
Proteasome	44	19		5,17E-005	6,56E-004	KEGG
Bacterial invasion of epithelial cells	70	26		5,59E-005	6,59E-004	KEGG
B cell receptor signaling pathway	75	27		7,62E-005	8,38E-004	KEGG
Leukocyte transendothelial migration	116	37		8,27E-005	8,53E-004	KEGG

Tabla 3 Enriquecimiento de vías basados en KEGG. El tamaño del conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la base de datos KEGG correspondiente según las vías almacenadas en el sitio ConsensusPathDB. El número de candidatos contenidos refiere a la cantidad de proteínas que son parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

4.3 Existen características comunes de regulación en las proteínas de la red ampliada

Se usó la información generada por el proyecto ENCODE para realizar búsquedas entre los factores de transcripción (TF) que poseyeran información de secuenciación masiva de inmunoprecipitación de cromatina (CHIP-Seq) disponibles para sitios sobrerrepresentados de unión de un TF correspondiente en los genes de nuestra red. Se evaluó el enriquecimiento mediante la consideración de sitios de unión que se encuentran hasta 700 pb corriente arriba o 300 pb corriente abajo de todos los sitios de inicio de transcripción descritos en la base de datos GENCODE y que tenían una puntuación mayor que 750. Consideramos estas 1.000 pb, como la secuencia del promotor, ya que se ha informado de que esta es la región de mayor concentración de picos de ChIP-Seq en estudios previos (Robertson et al. 2007)

Para los promotores de los genes que pertenecen a nuestra red ampliada, nos centramos de nuevo en la información proporcionada en GENCODE. De los 1321 nodos de la red se recuperaron 1431 promotores correspondientes a 957 genes únicos de la red. Se encontró mediante una prueba hipergeométrica que los factores de transcripción E2F1, TCF4, c-Myc, Max, E2F6 y NFkB fueron los más significativamente expresados (Tabla 4). Es importante destacar que los ensayos de ChIP-Seq de los factores de transcripción E2F1, c-Myc y Max se realizaron sobre las líneas celulares HeLa-S3 (<http://genome.ucsc.edu/>). El hecho de que estos experimentos se realizaron en una de las líneas celulares que estamos investigando da validez este análisis bioinformático adicional, porque es poco probable que el contexto celular sea significativamente diferente.

<i>Transcription Factor</i>	<i>cell line</i>	<i>tissue of origin</i>	<i>peaks near TSS of the genes of the extended network</i>	<i>peaks near TSS of all the GENCODE genes</i>	<i>p-value</i>
E2F1	HeLa-S3	cervical	187	1445	4,81E-013
TCF4	HCT-116	colorectal	244	2059	1,46E-012
Pol2	HeLa-S3	cervical	312	2878	3,93E-011
c-Myc	HeLa-S3	cervical	123	880	5,52E-011
Max	HeLa-S3	cervical	161	1266	9,77E-011
E2F6	k562	Leukemia	265	2379	1,17E-010
NFKB	GM12878	Lymphoblastoid	111	794	5,02E-010

Tabla 4 Factores de transcripción significativamente sobre expresados identificados mediante datos de ChIP-Seq pertenecientes a la base de datos ENCODE. Los picos fueron asignados a las regiones correspondientes de secuencias de promotor [-700 a 300 pb], basados en los sitios de inicio de transcripción (TSS) anotados en GENCODE. Los picos cerca de la TSS de los genes de la red extendida y todos los genes GENCODE se refiere a la cantidad de picos encontrados en la región promotora. El P-valor se calculo mediante una prueba hipergeométrica. El tamaño de la muestra se consideró como la cantidad de genes que correspondían con las proteínas de la red extendida. El tamaño de la población es el número de entradas GENCODE anotadas que poseían una anotación completa y estaban separados por una distancia de hasta 500 pares de bases.

Además, es importante mencionar que se realizó un análisis similar, pero utilizando las regiones promotoras como se informa en la base de datos de elemento regulador de la transcripción (TRED) (Jiang et al. 2007) y el empleo de la base de datos de Commutation para la adquisición de los sitios de inicio de transcripción. Con este enfoque, se obtuvo un enriquecimiento similar (Tabla 5).

<i>Transcription Factor</i>	<i>cell line</i>	<i>tissue of origin</i>	<i>peaks near TSS of the genes of the extended network</i>	<i>peaks near TSS reported on SwitchGear</i>	<i>p-value</i>
Pol2	HeLa-S3	cervical	608	7766	2,15E-067
TCF4	HCT-116	colorectal	455	5543	1,93E-050
E2F6	k562	Leukemia	488	6343	1,12E-046
Max	HeLa-S3	cervical	313	3375	2,90E-042
c-Myc	HeLa-S3	cervical	250	2395	1,88E-041
NFKB	GM12878	Lymphoblastoid	195	1802	6,78E-034
E2F1-HA	HeLa-S3	cervical	315	4272	7,57E-024

Tabla 5 Factores de transcripción significativamente sobre expresados identificados mediante datos de ChIP-Seq pertenecientes a la base de datos SwitchGear. Los picos fueron asignados a las regiones correspondientes de secuencias de promotor [-700 a 300 pb], basados en los sitios de inicio de transcripción (TSS) anotados en SwitchGear. Los picos cerca de la TSS de los genes de la red extendida y todos los genes SwitchGear se refiere a la cantidad de picos encontrados en la región promotora. El P-valor se calculo mediante una prueba hipergeométrica. El tamaño de la muestra se consideró como la cantidad de genes que correspondían con las proteínas de la red extendida. El tamaño de la población es el número de entradas GENCODE anotadas que poseían una anotación completa y estaban reportados en SwitchGear con una puntuación mayor a 20.

La sobre representación del Factor transcripcional c-Myc es sobresaliente, ya que en la mayoría de los tumores se encuentra altamente expresado, alterando de forma significativa la sobrevivencia y el ciclo celular. Por otro lado, c-Myc necesita dimerizarse para lograr unirse a los sitios promotores en el DNA y la consecuente activación de genes blanco, su pareja canónica es Max, de modo que la presencia de esta proteína en nuestro análisis aporta robustez a los resultados. Los mecanismos que hacen de c-Myc un oncogén sumamente poderoso aun no son muy claros, sin embargo, hay evidencia que indica que la activación de c-Myc se correlaciona con aproximadamente 70% de los cánceres humanos.

La primera explicación posible para la prevalencia de c-Myc en diversos tipos de cáncer es que, los genes que induce, representan la respuesta primaria de casi todas las vías de transducción de señales que se sospecha participan en la transformación maligna (Nilsson y Cleveland 2003). Asimismo, Nilsson (2003), ha sugerido que una pérdida en la capacidad de c-Myc para inducir apoptosis puede convertir a dicho TF en un promotor de crecimiento y transformación celular. Concordantemente, se ha encontrado la abrogación de la función de p53, que es la proteína responsable de la ejecución de la respuesta apoptótica mediante c-Myc en todas las líneas celulares utilizadas en el presente estudio (Scheffner 1991).

Por otro lado, el factor E2F1, controla progresión del ciclo celular y la replicación del ADN, se ha demostrado que es inducido por la actividad de c-Myc (Zeller et al. 2006). E2F1 también ha sido implicado en la sub regulación de p27^{KIP1} mediada por c-Myc, el cual es un inhibidor de quinasas dependiente de ciclinas (Baudino et al. 2003). Por último, la polimerasa 2 aparece como la proteína de unión al ADN que presenta el p-valor más significativo, lo que nos sugiere que los genes de nuestra red están siendo transcritos.

Debido a estos resultados, podemos proponer que en las líneas celulares de cáncer cérvico uterino, existe una red de regulación transcripcional anormal resultado de alteraciones funcionales y de expresión de los TF c-Myc, Max y E2F1.

(Los resultados presentados hasta este punto fueron publicados en el artículo: Higareda-Almaraz J.C., Enriquez-Gasca M. R., Hernandez-Ortiz M., Resendis-Antonio O., Encarnacion-Guevara S. (2011) Proteomic patterns of cervical cancer cell lines, a network perspective. BMC systems biology, 5:96 , ANEXO 1)

Continuando con la investigación, nos centramos en la línea celular HeLa, integrando diferentes capas de información mediante técnicas de biología de sistemas (Figura 7)

4.4 Existen diferentes perfiles de expresión genética y ontología génica entre las líneas celulares HeLa y NHEK

Para conocer la expresión genética en la línea celular HeLa, utilizamos un conjunto de datos de secuenciación masiva de mRNAs generado por el grupo de Nagaraj (Nagaraj et al 2011). El análisis de expresión total arroja un conjunto de 19,974 transcritos codificados por HeLa. La principal diferencia con respecto al análisis del grupo de Nagaraj fue la realización de normalización por cuartiles, lo que mejora la precisión de la búsqueda de transcritos poco abundantes (Fig. 8a). Con los datos de expresión total, nosotros buscamos la representación de los

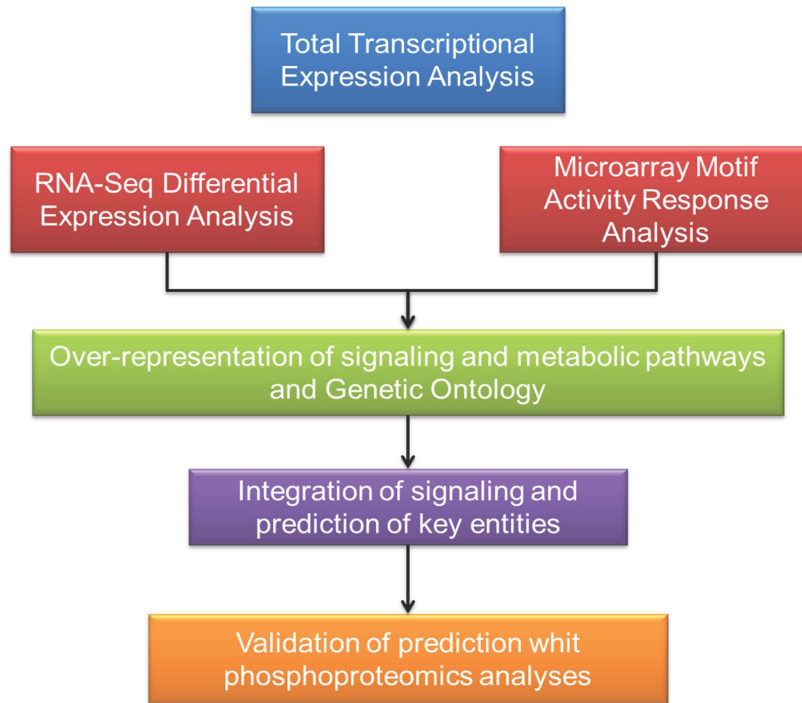


Fig. 7 Integración de diferentes capas de información biológica dentro de la dinámica celular. Se realizó un análisis del total de transcritos en la línea celular HeLa, este análisis presentó un panorama general de la expresión génica. Posteriormente, se realizó un análisis de expresión diferencial mediante RNA-seq y una análisis de sobre representación de la actividad de FT. Esto permitió la reconstrucción de rutas metabólicas y vías de señalización y regulación de la transcripción celular. Por último, se validó esta reconstrucción con un análisis de enriquecimiento de proteínas fosforiladas mediante fosfoproteómica.

procesos celulares mediante Ontología Génica, usando el dominio de “componente celular” en nivel 3, esperando encontrar una alta representación de las vías (Fig. 8b). Cada uno de los términos de GO se representa al menos en un 97% del total reportado para ese término (Tabla S1). Esta métrica nos resultó útil para establecer la calidad biológica de nuestro análisis que nos permitió construir nuestras predicciones a partir de estos datos, además de permitirnos validar nuestro análisis de expresión diferencial.

Posteriormente, para entender los cambios en la expresión genética en la línea celular HeLa con respecto a una célula normal, fue realizado un análisis de expresión diferencial por RNA-Seq (Zhong Wang et al. 2009), usando como control

a la línea celular de queratinocitos epiteliales NHEK. Los resultados muestran un total de 3,360 genes sobre-expresados y 2,129 genes sub expresados (Fig. 8c).

Usando esa información, realizamos un análisis de enriquecimiento de GO con la ayuda del software Consensus Path DB (Kamburov et al., 2011), usando el nivel 3 del domino “proceso biológico”. Incluso a este nivel de resolución, es claro que la expresión diferencial de HeLa favorece fuertemente la proliferación celular sobre la organización tisular (Tabla S2). Las categorías que muestran una clara sobre-representación son ciclo celular, expresión genética, metabolismo de “building blocks” y reorganización del citoesqueleto (Fig 8d). Mientras tanto, dentro de las categorías sub representadas se destacan desarrollo tisular, desarrollo de órganos y sistemas, señalización, adhesión celular, metabolismo de lípidos y muerte celular programada (Fig 8e).

Con estos resultados, sugerimos que existe una fuerte tendencia de HeLa a expresar genes que participan en la evasión del control tisular, lo cual confiere una clara ventaja adaptativa a la proliferación sin barreras.

4.5 La red de regulación de factores de transcripción expresados diferencialmente en células HeLa controla los procesos fundamentales en el mantenimiento del estado neoplásico

Para entender la red transcripcional que gobierna la expresión génica en la línea celular HeLa, usamos datos de microarreglos affymetrix de HeLa y epitelio cervical normal, generados por el grupo de Scotto (Scotto et al. 2008), y depositados en la base de datos GEO. Esos datos fueron analizados en con el “software” MARA, que reporta aquellos factores de transcripción cuya expresión se encuentra regulada diferencialmente. Los blancos de los factores E2F, ZNF143, YY1, ELKA, GABP, NRF1, MYB, NFY, HIF1A, TFDP1 y ELF se observaron sobre-representados, mientras que los blancos transcripcionales de ETS, NFATc, NR1H4, SMAD, TFPC2, HIC1, AR, TBP, SRF y KLF12 sub-representados.

Subsecuentemente, usamos la base de datos generada de blancos transcripcionales generada por MARA para establecer la red de regulación de los

genes obtenidos por el análisis de expresión diferencial. Los blancos fueron obtenidos para cada TF y reconstruimos sus redes regulatorias. Considero importante mencionar que c-Myc, el Factor Nuclear de Hepatocitos 4-alfa, BRCA1, VHL y NEMO se encuentran involucrados en más de una red de TF, es decir, se encuentran sobre-representados, además de encontrarse sobre-expresados.

Después de obtener información de los TF y sus blancos, realizamos un análisis de enriquecimiento mediante Consensus Path DB usando el nivel 3 del dominio “proceso biológico”. En la red de los TF sobre-expresados, encontramos 103 GO “terms” (Tabla S3), resaltando proliferación celular, metabolismo de “building blocks” organización celular, angiogénesis, metabolismo central y señalización (Fig 9a).

En la red de TF sub-expresados, encontramos 159 GO “terms” (Tabla S4) entre los que resaltan, homeostasis de tejidos y misceláneos. Además, con la información generada por las redes de los TF, se construyeron redes de interacción PPI usando el software Cytoscape, a través del “plugin” Hubba (Lin et al. 2008), se definieron los “hubs” usando las propiedades topológicas de “node degree” (Fig 9b) y “betweenness” (Fig 9c). Sorprendentemente c-Myc, HNF4A, BRCA1, VHL y NEMO poseen los valores más altos de estas centralidades. Estos resultados sugieren que un conjunto de genes sobre-expresados en HeLa controlan una vía de regulación particular que no se encuentra presente en el epitelio cervical normal; esto también sugiere una supresión de la expresión de otras redes de regulación. Esta peculiaridad en las redes de regulación puede ser una importante fuente de complejidad que permite fortalecer el sistema celular por presión de selección.

4.6. Identificación de vías sobre o sub representadas que gobiernan el fenotipo neoplásico en la línea celular HeLa

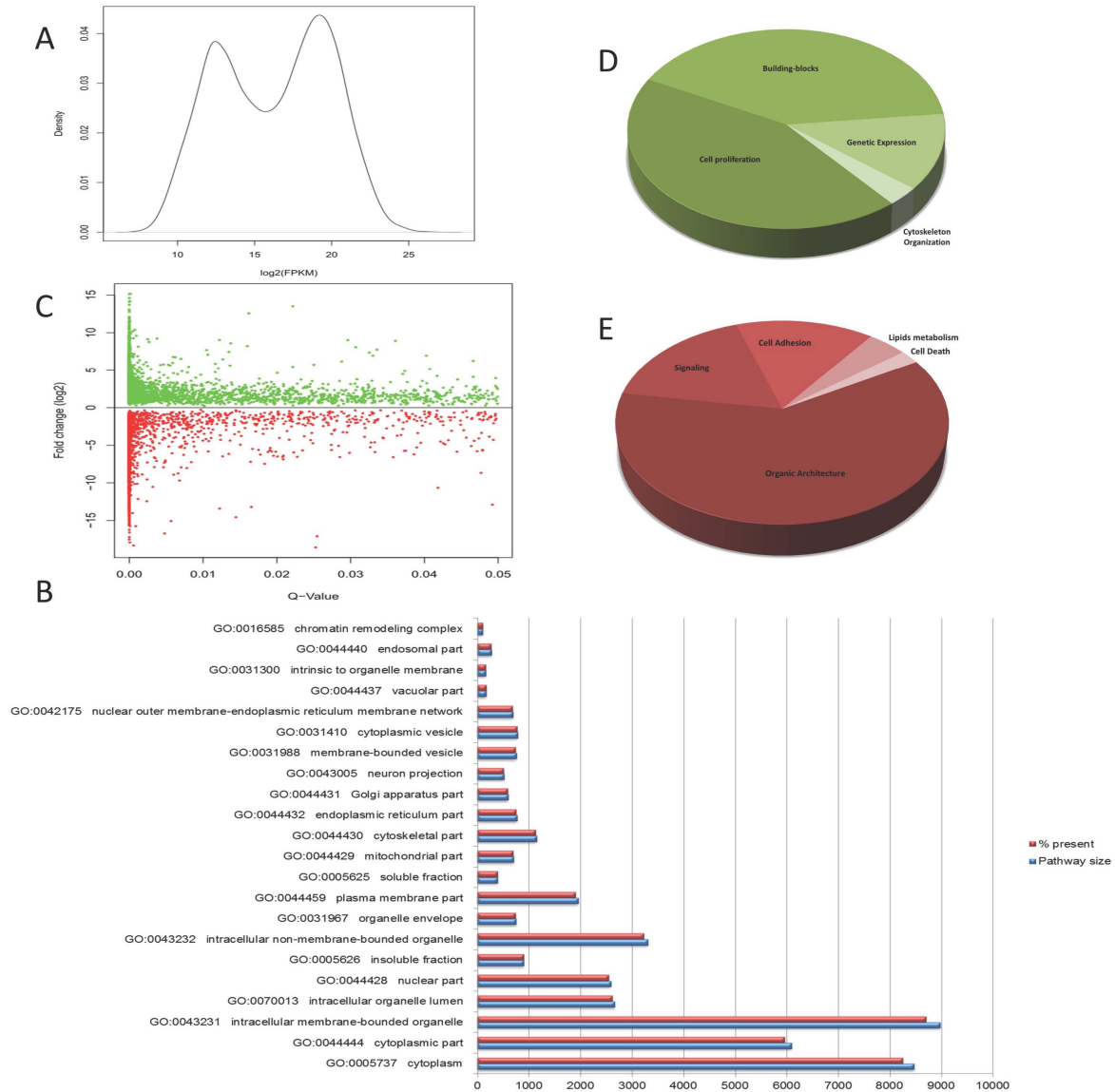


Fig. 8 Patrones de expresión génica en la línea celular HeLa a) La distribución de los transcritos totales muestra que existen dos poblaciones, una de baja abundancia y una segunda, mayor, de alta abundancia. Este análisis muestra que los parámetros utilizados para la búsqueda de transcritos de baja abundancia funciona correctamente. b) Una representación gráfica de los procesos celulares que se distinguen mediante ontología génica (GO), utilizando el dominio “componentes celulares” en nivel 3, la cantidad de elementos recuperados se comparó con el tamaño total de la vía. c) Un diagrama de dispersión que muestra la calidad de los resultados diferenciales del análisis de la expresión diferencial mediante RNA-Seq, utilizando la línea de queratinocitos epiteliales NHEK como control. Se encontró un total de 3.360 genes sobre expresados y 2.129 genes sub-expresados. d) La distribución porcentual del dominio “procesos biológicos” GO nivel 3, representados por los transcritos sobre expresados. e) La distribución porcentual del dominio “procesos biológicos” GO nivel 3 representados por los transcritos sub expresados. Estas gráficas fueron construidos a partir de un resumen de todos los términos GO similares en un circuito celular funcional.

Para lograr entender cuáles vías y procesos son responsables de mantener el fenotipo neoplásico, se transformó la lista completa de expresión diferencial en

identificadores de UniProt no redundantes. La lista se utilizó para realizar un análisis de enriquecimiento de vías mediante ConsensusPathDB, fueron recuperadas 83 vías a partir de los transcritos sobre-expresados según la base de datos de Protein Interacción Database (PID) (Tabla S5).

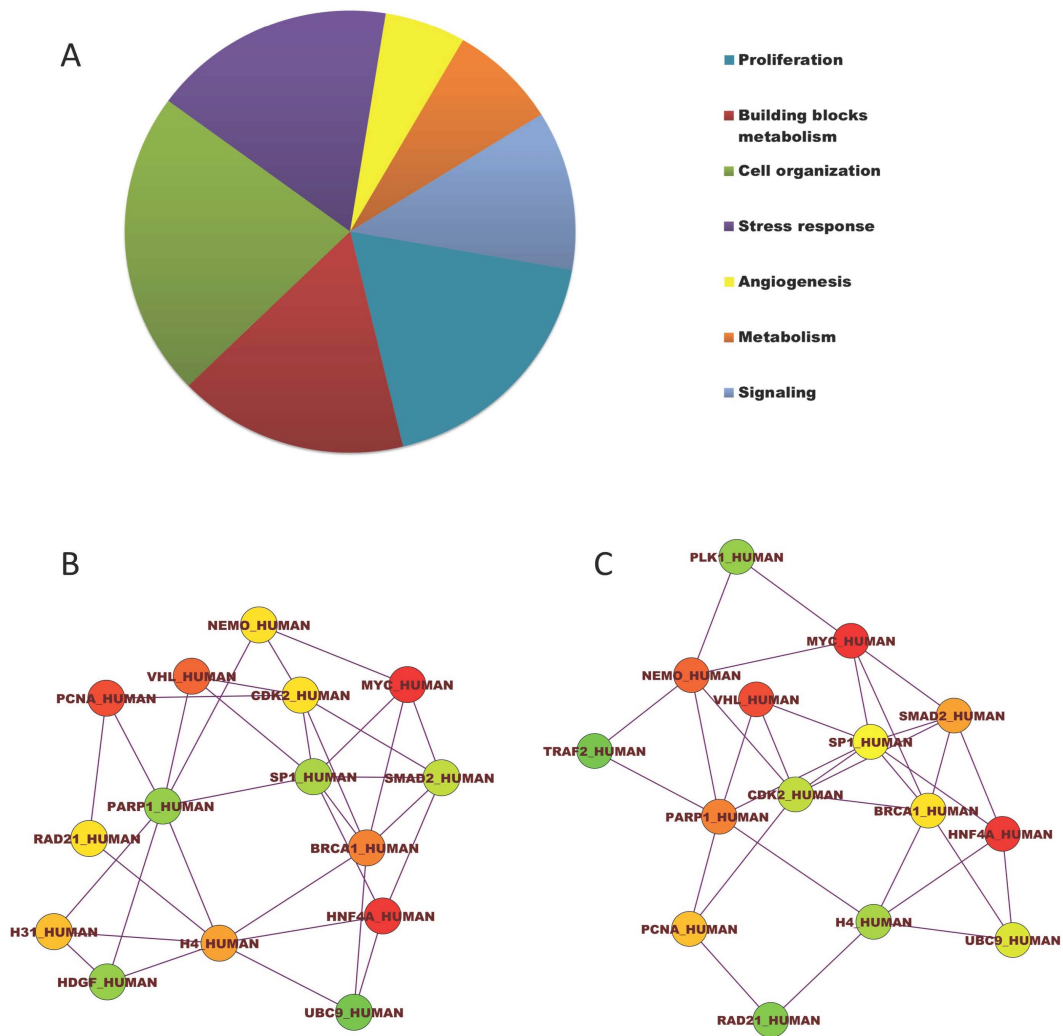


Fig. 9 Redes de expresión de factores transcripcionales a) La distribución porcentual del dominio “procesos biológicos” GO nivel 3 construidos a partir de la red de FT sobre representados, destacando “Proliferación celular”, “Metabolismo de los bloques de construcción”, “Organización celular”, “Angiogénesis”, “Metabolismo central” y “Señalización”. Esta tabla fue construida a partir de un resumen de todos los términos GO similares en un circuito celular funcional. b) Hubs que se obtuvieron a partir de la medición de la centralidad “node degree” en la red de FT sobre representados. El color indica la calificación, donde el rojo es el más alto y amarillo el más bajo. c) Hubs que se obtuvieron a partir de la medición de la centralidad “betweenness” en la red de FT sobre representados. El color indica la calificación, donde el rojo es el más alto y amarillo el más bajo.

Destacan diversas vías de señalización, como ATR, Aurora A y B, algunas redes de regulación como E2F, Myb, blancos de activación transcripcional de c-Myc y efectores directos de p53. En el caso de los transcritos sub expresados, se detectaron 15 vías sobre-representadas según PID, algunas de las más sobresalientes son blancos transcripcionales de deltaNp63, TAp63, y de los miembros de la familia AP1, Fra1 y Fra2 y efectores directos de p53. Es notable que tanto los transcritos sub como los sobre-expresados se encuentren en la vía de efectores de p53 (Tabla S7). Este es uno de los ejemplos del poder de nuestro análisis, permite distinguir desde una perspectiva sistémica componentes de una misma vía que pueden promover el estado maligno por participación activa o por omisión.

Simultáneamente se realizó un análisis por enriquecimiento utilizando la base de datos KEGG, obteniendo 17 vías para los transcritos sobre-expresados (Tabla 6) y 25 para los sub expresados (Tabla S8).

Pathway name	Set size	Candidates	P-value	Q-value
DNA replication	36	28 (77.8%)	7.24E15	1.67E-12
Cell cycle	124	53 (42.7%)	9.42E11	1.09E-08
Homologous recombination	28	20 (71.4%)	7.71E10	5.94E-08
Systemic lupus erythematosus	138	51 (37.5%)	4.46E08	2.58E-06
Fanconi anemia pathway	52	26 (50.0%)	1.32E07	6.12E-06
Mismatch repair	23	15 (65.2%)	6.97E07	2.68E-05
Base excision repair	33	18 (54.5%)	2.27E06	7.49E-05
Transcriptional misregulation in cancer	180	53 (29.8%)	6.84E05	0.00198
Lysine degradation	49	20 (40.8%)	0.00018	0.0037
Purine metabolism	166	49 (29.5%)	0.00016	0.0037
One carbon pool by folate	19	10 (52.6%)	0.00062	0.0135
RNA transport	157	44 (28.0%)	0.00113	0.0217
HTLV-I infection	263	67 (25.5%)	0.00124	0.0221
Nucleotide excision repair	46	17 (37.0%)	0.00172	0.0284
Pyrimidine metabolism	101	30 (29.7%)	0.00253	0.039
Spliceosome	127	35 (27.6%)	0.00463	0.0668
Small cell lung cancer	87	25 (28.7%)	0.00876	0.119

Tabla 6 Enriquecimiento de vías según PID basados el conjunto de transcritos sobre expresados. El tamaño del conjunto se refiere al número de transcritos que tienen un ID Uniprot en la vía PID correspondiente según el sitio ConsensusPathDB. El número de candidatos se refiere al número de proteínas que forman parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Los resultados nos ofrecen una perspectiva sobre diversos eventos que ocurren en una forma diferencial en las células HeLa con respecto a las células normales; algunos de los procesos más significativos se encuentran relacionados a proliferación celular: Replicación del DNA, Ciclo Celular y Recombinación Homóloga. Algunos otros remarcables son Vía de anemia de Fanconi, Regulación deficiente de la transcripción en el cáncer y cáncer de pulmón de células pequeñas, todos ellos guardan estrecha relación con la conservación del fenotipo maligno a través de diferentes tipos de cáncer.

De forma sorprendente, dentro de los genes sub-expresados se encuentran sobre-representadas vías de metabolismo de lípidos, particularmente de esteroides, ácidos linoleico y araquidónico, así como síntesis de ácidos grasos no saturados, así como adhesión celular.

Estos datos complementan y fortalecen los previamente obtenidos mediante el análisis de microarreglos analizados mediante el programa MARA. Sin embargo, a pesar de que consideramos que los enriquecimientos de vías poseen un gran valor en el análisis de la trascendencia biológica de la expresión genética, debemos reconocer que esta forma de análisis posee limitantes, el paso crítico en la formulación de modelos es la validación de los mismos, por lo cual es necesario obtener datos experimentales que confronten al modelo con la realidad y ponga a prueba sus capacidades predictivas.

4.7 Existe un patrón definido en la activación de circuitos celulares que involucra interconexiones e interferencias, que se encuentran determinados por expresión genética, redes de regulación y vías diferenciales en la línea celular HeLa.

Para visualizar la integración de diferentes capas de información biológica que fue obtenida en el presente trabajo, se empleó la herramienta de PID Batch Query tool. Se ingresó la lista de transcritos obtenida mediante el análisis de expresión diferencial, así como la lista de los TF que fueron obtenidos mediante el análisis realizado con el programa MARA. Se obtuvieron un total de 64 vías curadas en

formato biopax nivel 3 (Demir et al. 2010), mostrando una cantidad elevada interconexiones e interferencias entre distintos circuitos celulares. Estas interconexiones permiten el mantenimiento constante de la proliferación, inmortalización y migración celular mediante la expresión diferencial y patrones de control transcripcional que difieren de los observados en las células normales.

Para explorar esta hipótesis, realizamos una reconstrucción de cada una de las redes mencionadas mediante el “plugin” Bisogenet de Cytoscape. Posteriormente, realizamos una búsqueda de nodos altamente conectados y con esos datos construimos dos modelos. Para el primero, construimos una red utilizando cada componente de cada vía (Figura 10). Se exploró la topología de la red usando las medidas de centralidad “node degree” y “betweenness” y con ello determinamos la presencia de “hubs” en todas las vías. Esto nos permitió crear un modelo en donde destaca el hecho de que c-Myc, BRCA1, VEGFA y E2F1 son los nodos más interconectados y poseen una influencia determinante en todas las redes.

Para nuestro segundo modelo, nosotros partimos de la proposición de que hay circuitos celulares bien definidos que pueden ser extrapolable a otras células malignas. Dichos circuitos fueron adaptados al concepto de características del cáncer propuestas por Hanahan y Weinberg, y modelados en un concepto que llamamos meta-vías (Figura. 11). Cada una de las vías que aparecen en las figuras posee un cierto grado de conexión con otras vías que es determinado por la regulación transcripcional, señalización, metabolismo o una combinación de estos mecanismos. Estas meta-vías son capaces de mantener el fenotipo maligno. En la figura 5 cada uno de las vías y meta-vías son relacionadas con su posible papel dentro de las características del cáncer.

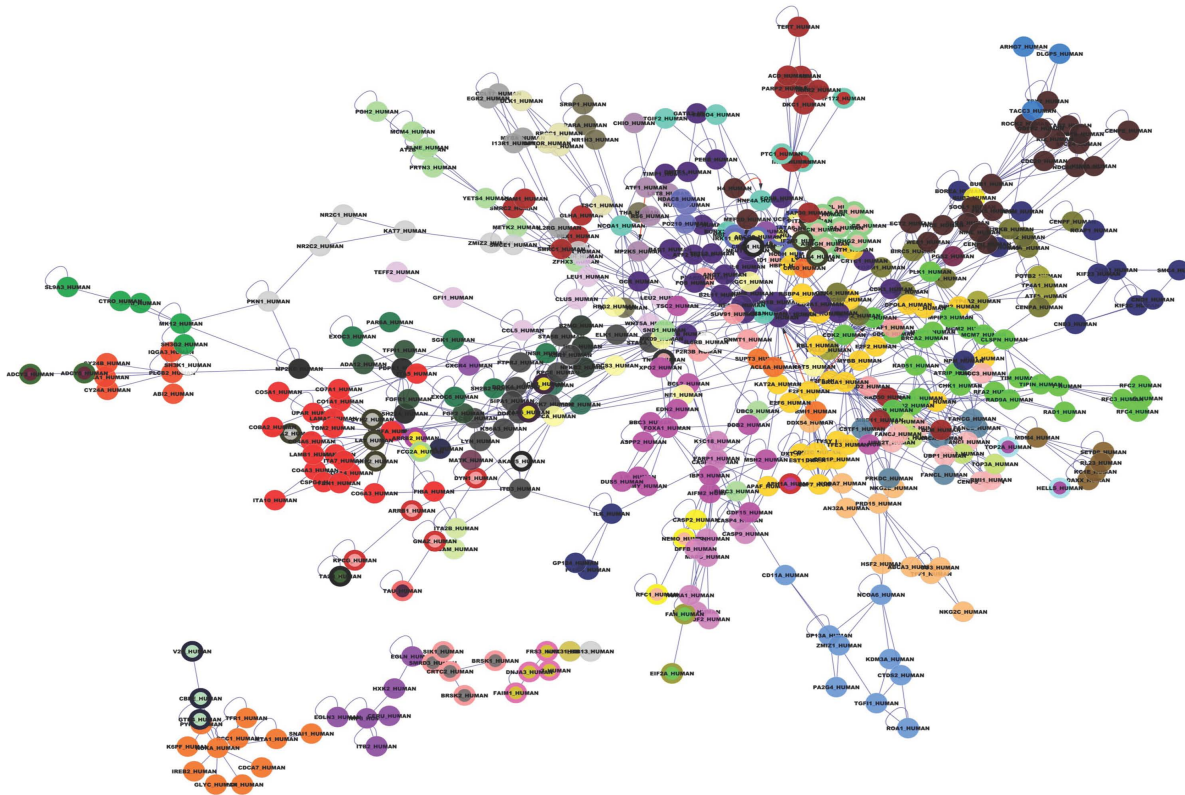


Fig. 10 Red de interconexiones y “crosstalks” entre los circuitos celulares. La red se realizó a partir de datos obtenidos por el análisis de las vías de señalización y regulación. Los nodos representan las proteínas que componen cada una de las vías, cada ruta se indica mediante el uso de colores diferentes.

Para validar los datos que fueron obtenidos mediante expresión diferencial y regulación transcripcional, realizamos un análisis fosfoproteómico en la línea celular HeLa mediante un enriquecimiento con Cromatografía de Afinidad por Oxido de Metal (MOAC), y subsecuentemente la identificación de las proteínas mediante LC/MS-MS.

Fueron identificadas un total de 271 proteínas fosforiladas obtenidas en 3 réplicas experimentales (Tabla S9) repartidas en 40 GO terms nivel 3 (Tabla S10), 21 vías de PID (Tabla S11) y 16 vías de KEGG (Tabla S12). Como era esperado, debido a la baja correlación bien documentada entre transcritos y proteínas expresadas, solo el 17% de las proteínas tienen su equivalencia en transcritos. Sin embargo, se encontró que existe una alta correspondencia entre las vías que fueron

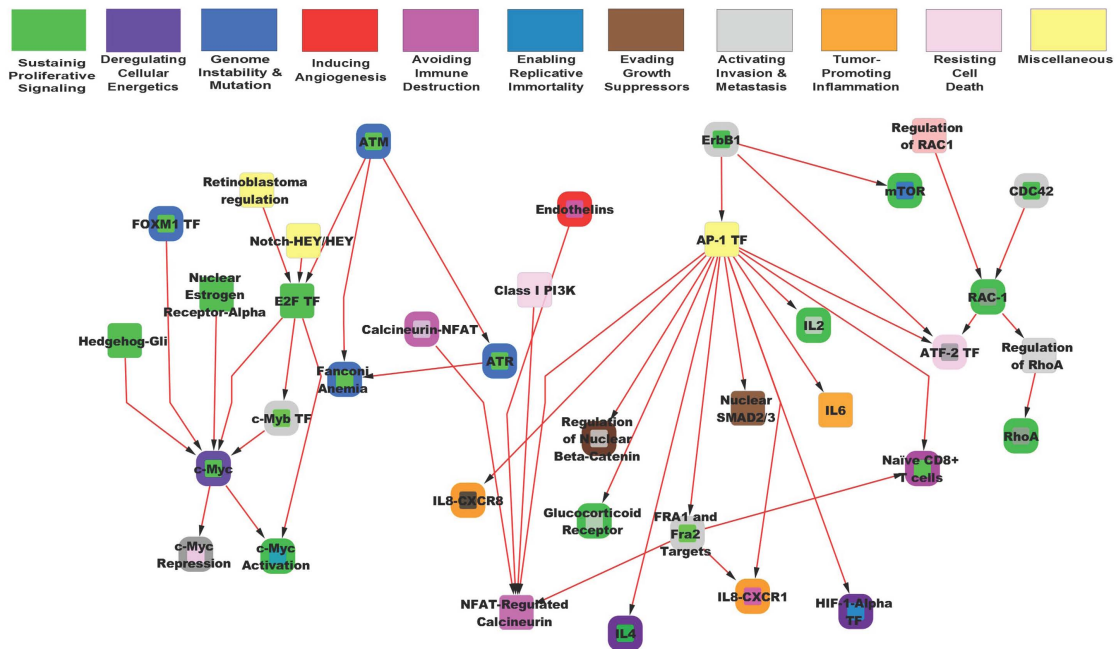


Fig. 11 Análisis de Meta-vías. Se realizó un análisis combinando las vías de señalización y regulación transcripcional; los bordes indican la relación regulatoria o jerárquica, y los nodos indican vía. Los colores indican cada una de las “características del cáncer”, con dos de las características más representativas resaltadas por nodo.

definidas por transcritos y/o TF y aquellos que corresponden a las proteínas fosforiladas, poniendo como ejemplo a blancos transcripcionales validados de c-Myc, eventos de señalización mediados por HDAC clase III, Eventos de señalización de LKB, eventos de señalización de P3K clase I y señalización por la familia FoxO. Otras vías destacadas son aquellas de las cuales, las proteínas identificadas son una consecuencia de las vías derivadas del conjunto de transcritos, como es el caso de la red del factor transcripcional c-Myb y la red del factor de transcripción E2F, cuyos blancos se encuentran dentro de la red de activación transcripcional de c-Myc y simultáneamente en metabolismo central.

Por último, el conjunto de datos que se obtuvo de las tres capas de información fue utilizado para reconstruir la red de señales, regulación y metabolismo que gobierna la línea celular HeLa y que es diferente de aquella que presenta el epitelio cervical normal y la línea de queratinocitos primarios NHEK. Para cumplir ese propósito, se utilizó la herramienta KEGG Mapper tool (Okuda et al. 2008)

para ensamblar el mapa general metabólico (Figura 12), el mapa de ciclo celular (Figura 13a), y los mapas de moléculas de adhesión y adhesión focal (Figura 13b). Basados en esos mapas, fue posible reconocer un patrón claro que puede ser dividido en 5 grupos:

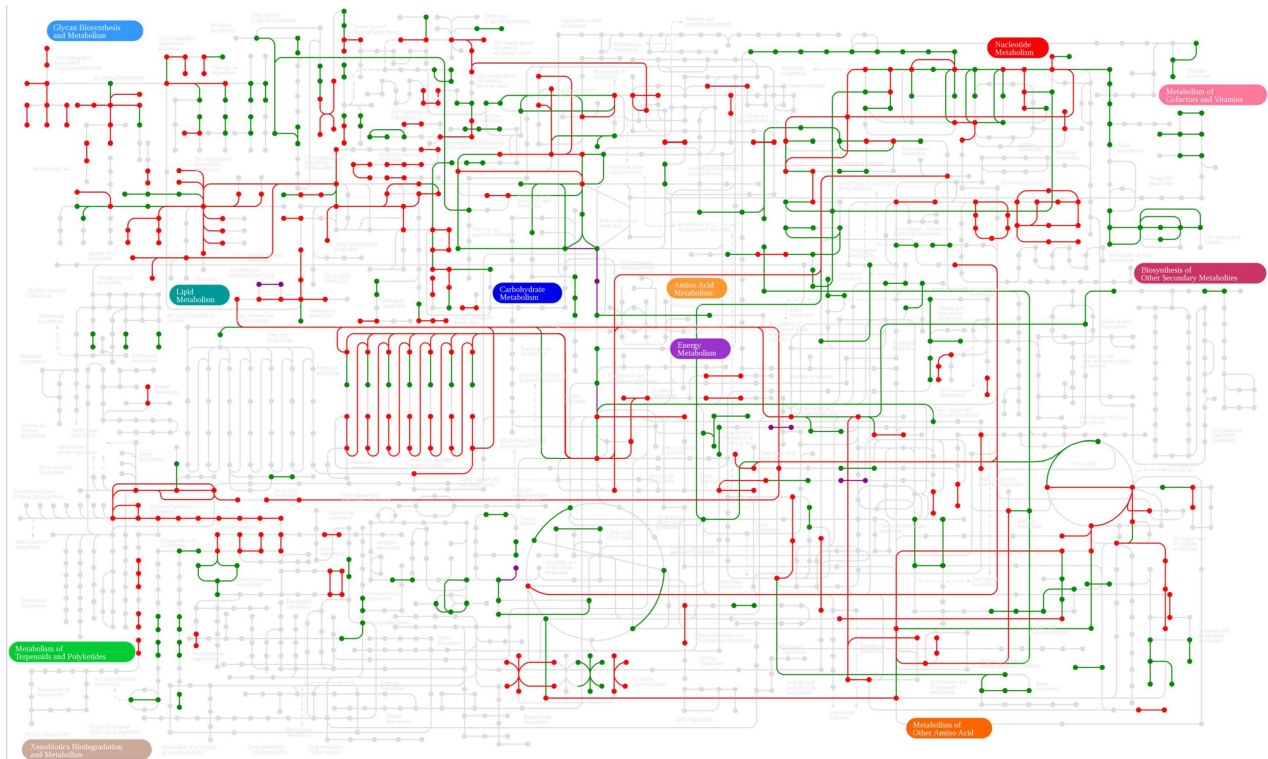


Fig. 12 Vías metabólicas sobre representadas. Los mapas metabólicos fueron reconstruidos mediante la base de datos KEGG, las redes de FT sobre representadas y los análisis fosfoproteómicos. Los transcritos sobre expresados se muestran en verde, los transcritos sub expresados en rojo y las fosfoproteínas identificadas en morado.

1.- Incremento en la proliferación celular gracias a la sobre-expresión de los genes del complejo MCM y diversas proteínas del ciclo celular, como p27, E2F y c-Myc que dan paso a la “Señal de Proliferación Sostenida”. Por otro lado, es complicado integrar las señales completas que encontramos dentro de ciclo celular, para lograr esto necesitaríamos realizar un análisis funcional mediante interferencia y sobre expresión de cada uno de los componentes en los cuales encontramos expresión diferencial para explicar a detalle su función.

2.- Sobre activación del metabolismo central (glicólisis y gluconeogénesis), así como un incremento en las enzimas de la vía de pentosas-fosfato que permiten la

3.- “Activación de la Invasión y Metástasis” causada por la pérdida de expresión de diversas proteínas de adhesión, como CLDN u OCLN, además del incremento en la expresión de ESAM.

(Los resultados presentados hasta este punto se encuentran en el artículo: JC Higareda-Almaraz, IA Valtierra-Gutierrez, M. Hernandez-Ortiz, S. Contreras, E. Hernandez, S. Encarnacion-Guevara. Analysis and prediction of pathways in HeLa cells by Integrating Levels of biological organization with systems-biology approaches, actualmente en revisión en la revista PLOS One, ANEXO 2)

4.- La sobre expresión de transcritos y proteínas fosforiladas que conforman el spliceosoma, el componente U2 se encuentra especialmente sobre representado lo cual es una importante causa de la “Inestabilidad Genómica”(Figura 14).

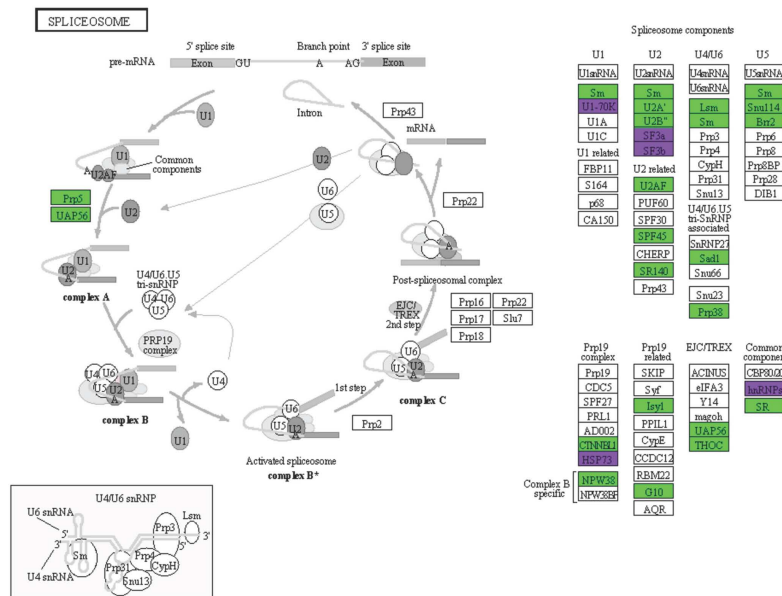


Fig. 14 Vía de componente del spliceosoma. Esta vía fue reconstruido e construyeron la base de datos KEGG, las redes de FT sobre representadas y los análisis fosfoproteomicos. Los transcritos sobre expresados se muestran en verde, los transcritos sub expresados en rojo y las fosfoproteínas identificadas en morado.

5.- La “Evasión de la Apoptosis” en la línea celular HeLa es fuertemente incrementada por la acción de las oncoproteínas E6 y E7, como se ha discutido

anteriormente, pero en adición, se encontró la expresión de los transcritos de cIAPs, TRAFs y Bcl-2, las cuales contribuyen a evitar la apoptosis.

Estos datos nos permiten afirmar que existe una fuerte correlación entre los datos obtenidos informaticamente y la validación mediante la expresión de fosfoproteínas y la vías deducidas a partir de estas,

4.8 Los análisis de la expresión genética de la línea celular HeLa demuestran la existencia de dominios cromosomales.

Para lograr establecer dominios cromosomales únicos en HeLa que pueden estar relacionados a la inestabilidad genómica, se realizó un análisis de expresión cromosomal comparativo entre las líneas celulares HeLa y NHEK mediante la herramienta Web DAVID. Se ingresó la lista de transcritos obtenida mediante el análisis de expresión diferencial. Los resultados arrojados indican que los transcritos sobre expresados se encuentran preferentemente en los cromosomas 3, 9, 15, 17, X, 20, 12 y 1, (Tabla S13) mientras que los sub expresados en los cromosomas 1, 19, 2, 10, 4 y 11 (Tabla S14).

Para conocer los procesos codificados por los transcritos sobre y sub expresados en los diferentes cromosomas, se realizó un análisis de sobre representación de vías de señalización según PID, mediante ConsensusPathDB, anteriormente descrito. El análisis sugiere que en la expresión genética proveniente de estos cromosomas se concentran algunas de las más importantes diferencias fenotípicas entre HeLa y una célula normal.

En los cromosomas 12 y 17 (Figura 15) se encuentra gran parte de los genes sobre expresados que pertenecen a las vías de señalización con mayor efecto pleiotropico, como E2F, c-Myc, p38, Fox y Aurora, además de de vías de remodelación de cromatina, representadas por las HDAC's de clase III (Tabla S15). En los cromosomas 2 y 19 (Figura 16), se encuentra la sub expresión crítica de vías pleiotropicas, como es el caso de Interleucinas 1/2, SMAD2/3, Calmodulina, señalización por calcio y p38 MAPK (Tabla S16).

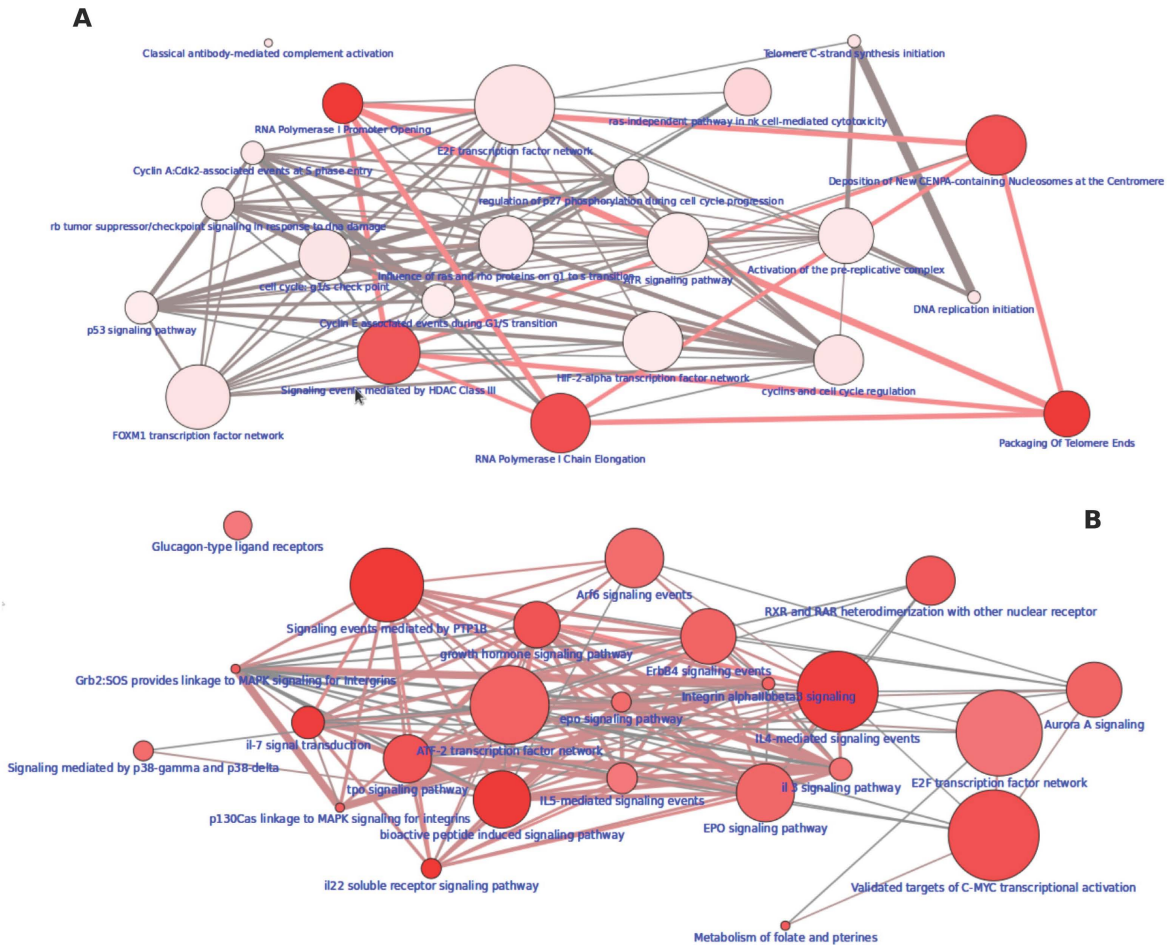


Fig. 15 Enriquecimiento de vías de señalización, según PID, de los cromosomas que contienen las secuencias de los transcritos sobre expresados. a) Cromosoma 12. b) Cromosoma 17. El tamaño de los nodos representa el número de genes contenidos en cada categoría y el color el P-valor obtenido. Todos los resultados son significativos.

Se presta especial atención al cromosoma 1, ya que dentro de él se encuentran transcritos sub y sobre expresados, para ello se analizaron las citobandas de dicho cromosoma, usando la herramienta Web DAVID Bioinformatics Resources 6.7. Se encontró que los transcritos sobre expresados se acumulan en las citobandas 1q21, 1q22, 1q32, 1q44, 1p13 y 1p34. Resaltan procesos como HDAC's de clase III, armado de centromeros mediante CEMPA, ARN Polimerasa I promotor de apertura y elongación de la cadena, activación de ATR en respuesta a estrés y empaquetamiento de telomeros (Tabla 7).

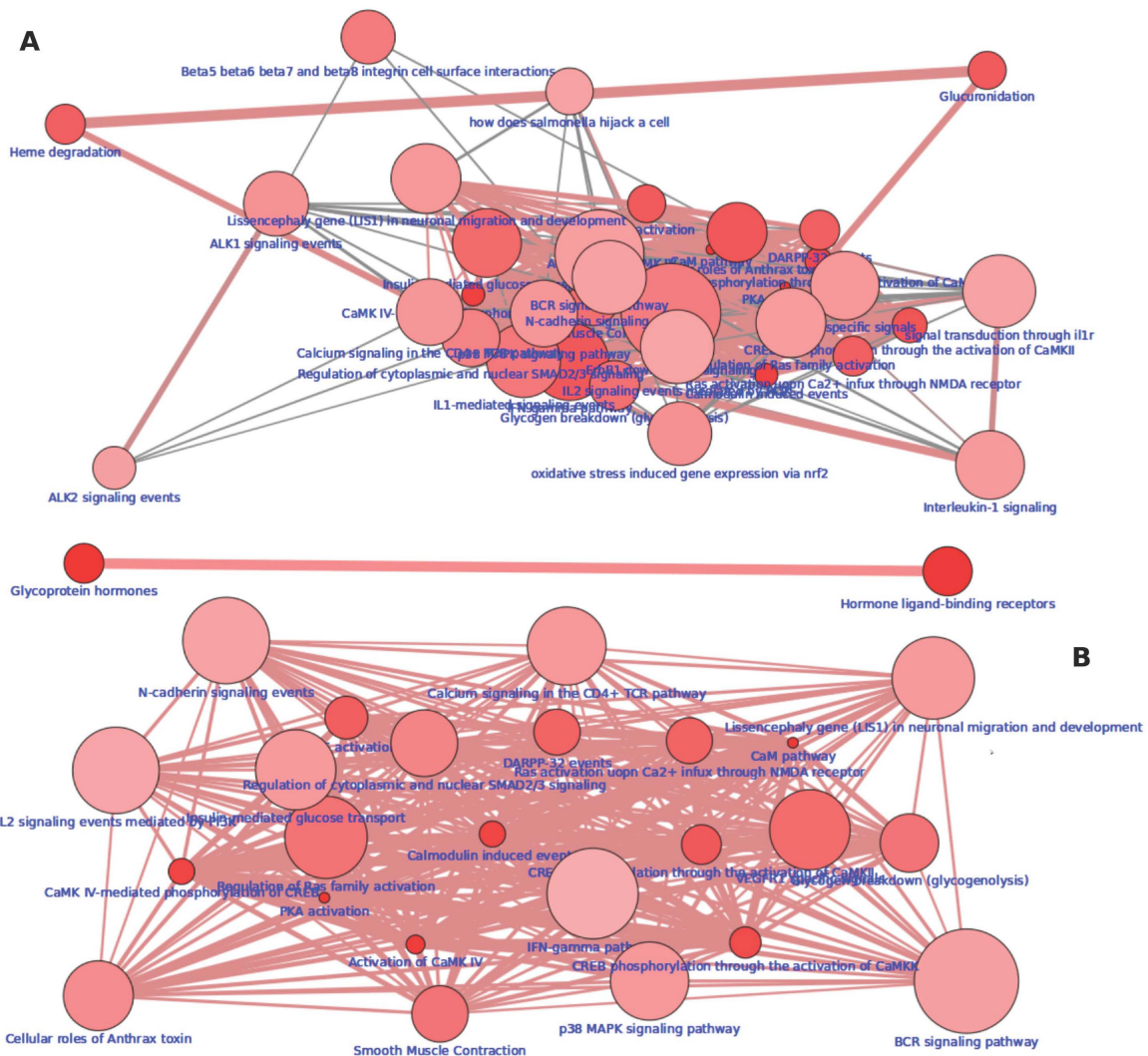


Fig. 16 Enriquecimiento de vías de señalización, según PID, de los cromosomas que contienen las secuencias de los transcritos sub expresados. a) Cromosoma 2. b) Cromosoma 19. El tamaño de los nodos representa el número de genes contenidos en cada categoría y el color el P-valor obtenido. Todos los resultados son significativos.

Por el lado de los transcritos sub expresados, la mayoría se agrupa dentro de las citobandas 1q21, 1q32, 1p13, 1p34 y 1p36. El análisis nos muestra que en los procesos sub regulados destacan Interacciones de unión estrechas, así como las vías de señalización de RhoA y p73 (Tabla 8).

Estos resultados nos sugieren patrones de expresión cromosomal, los cuales poseen un impacto importante en la expresión genética, además de que nos

permite especular que las aberraciones cromosómicas no son aleatorias y promueven la rápida mutación y consecuente micro evolución modelada por la presión de selección tisular.

Vías	# Genes	# Candidatos	Valor-p	Valor-q
RNA Polymerase I Promoter Opening	17	14 (82.4%)	1.74e-21	1.03e-19
Deposition of New CENPA-containing Nucleosomes at the Centromere	36	17 (48.6%)	6.88e-20	2.03e-18
Packaging Of Telomere Ends	21	14 (66.7%)	2.76e-19	5.43e-18
Signaling events mediated by HDAC Class III	40	15 (38.5%)	8.65e-16	1.28e-14
RNA Polymerase I Chain Elongation	35	14 (41.2%)	2.59e-15	3.05e-14
Mitotic Prometaphase	87	10 (11.5%)	1.54e-05	0.000151
Activation of ATR in response to replication stress	33	4 (12.5%)	0.00457	0.0385

Tabla 8 Enriquecimiento de vías de señalización, según PID, de las citobandas del Cromosoma 1 que contienen las secuencias de los transcritos sobre expresados.

5. Discusión

5.1. El grupo central de cáncer cérvico uterino representa un conjunto de procesos biológicos que permiten la preservación del fenotipo maligno.

En la primera parte de este trabajo realizamos un análisis integral derivado de datos obtenidos por técnicas de proteómica, integrando perspectivas de redes biológicas, los datos anotados del genoma humano y técnicas de secuenciación masivas disponibles para el público, como una herramienta que nos permitió la construcción de modelos para inferir un comportamiento biológico (Figura 3).

Como fue definido en el presente estudio, el “grupo central de proteínas del cáncer cérvico uterino” no describe el proceso de inmortalización, sin embargo nos es útil para representar los procesos biológicos involucrados en la viabilidad y proliferación celular. En la tabla 2 se pueden observar los procesos biológicos enriquecidos en nuestro conjunto de proteínas. Estos procesos sugieren un fenotipo maligno, en donde las GO se relacionan con la respuesta a proteínas no plegadas, metabolismo, procesos catabólicos, mantenimiento de ubicación celular y muerte celular. Estas observaciones son consistentes con nuestras conjeturas iniciales, ya que consideramos que el uso de la línea celular HaCaT es un control que nos permiten distinguir entre la inmortalización y transformación, porque es

una línea celular inmortalizada con un cariotipo estable y fenotipo no tumorigénicos.

Vias	# Genes	# Candidatos	Valor-p	Valor-q
Tight junction interactions	62 (33.3%)		0.00298	0.0726
RhoA signaling pathway	474 (8.9%)		0.00391	0.0726
Type I hemidesmosome assembly	82 (25.0%)		0.00547	0.0726
p73 transcription factor network	795 (6.3%)		0.00558	0.0726
regulation of spermatogenesis by crem	102 (20.0%)		0.00862	0.0897

Tabla 9 Enriquecimiento de vías de señalización, según PID, de las citobandas del Cromosoma 1 que contienen las secuencias de los transcritos sub expresados.

Además, la adquisición de las interacciones reportadas entre las proteínas identificadas nos proveyó de una primera idea de los elementos clave en la dinámica neoplásicas, permitiéndonos juzgar la validez de los análisis posteriores mediante la búsqueda de un acuerdo general entre la red inicial original y la red extendida. Es de notar que para la red pequeña el nodo con un valor mayor de interconexión es 1433Z, y esta topología se conserva al expandir la red. Además, es importante mencionar que ambas redes están enriquecidas en procesos similares.

5.2. La familia de transductores de señales 1433 se erige como un posible segundo “hit” dentro del cáncer cérvico uterino.

Cuando se analiza la red extendida, se observaron 6 de los 7 miembros de la familia 1433 (Z, G, B, Q, E, H y S) como los nodos con más conexiones. Esto es consistente con el papel biológico de esta familia como transductores de señales pleiotropicos y ubicuos (Tzivion y Avruch 2001). La identificación de 1433Z y su papel como “hub” y cuello de botella es particularmente relevante, debido a que es bien conocido que se encuentra involucrada en tres procesos importantes, regulación del ciclo celular, transducción de señales y regulador de la apoptosis (Avruch et al. 2001; Zhang et al. 2003; Morrison 2008).

Como fue previamente reportado por Jhonsson y Bates (2006) las proteínas con mayor relevancia dentro del cáncer, exhiben una topología significativamente

diferente cuando son comparadas con proteínas sin relación con la malignidad, también es de resaltar que estas proteínas suelen poseer una mayor proporción de dominios estructurales promiscuos, lo que significa un alto grado de pleiotropía. Esto es particularmente cierto en los miembros de la familia 1433, ya que en su papel como integradores de señales, son capaces de amplificar señales y pueden coordinar respuestas biológicas, como la muerte celular o la sobrevivencia (Niemantsverdriet et al. 2008).

Además, existe una fuerte evidencia de que la sobre-expresión de 1433Z promueve la degradación de p53 mediante la vía proteosomal (Danes et al. 2008), lo cual permite potenciar la acción de la proteína E6 en células infectadas por VPH (Boyer et al. 1996). Al aumentar el recambio de p53, sumándose a los efectos pleiotropicos ocasionados por la sobre-expresión de 1433Z, podría ser la razón por la cual no todas las infecciones de VPH culminan en cáncer. Consideramos tener suficiente evidencia para sugerir que la proteína 1433Z representa un papel de conducción molecular en las líneas celulares cancerosas estudiadas, el balance en la expresión de esta proteína es capaz de determinar el destino celular de una forma independiente a la infección por VPH, sin embargo, potenciada por la presencia de este; causando una desregulación que podría concluir en una transformación maligna.

5.3. Los análisis de enriquecimiento de vías y el análisis de sitios de unión de TF a promotores sugieren una cooperación de c-Myc y E2F1 que mantiene encendidas distintas señales de sobrevivencia y proliferación.

Los resultados del análisis por enriquecimiento de vías según KEGG de las proteínas de nuestra red expandida muestran una sobre-representación de factores de sobrevivencia, relacionadas con la evasión de la apoptosis y adhesión focal relacionada con invasión y metástasis, dos de las más importantes y mejor documentadas características del cáncer; además de varias vías clásicas en cáncer de pulmón de células pequeñas, cáncer de próstata, cáncer endometrial y vías de cáncer (Tabla 3). La presencia de estas vías nos ayudan a fortalecer nuestro análisis, ya que nos permite inferir que las propiedades observadas a nivel

experimental se conservan y la expansión de la red PPI no alteró el significado biológico de los patrones de expresión.

Sin embargo, existen otras vías que no parecen presentar una conexión directa con el cáncer (neurotrofina, vía de señalización de receptores a células T, infecciones bacterianas y vías proteosomales). Posiblemente puedan participar en procesos como estrés celular, inflamación crónica, recambio proteico, sobrevivencia y evasión del sistema inmune.

Como una forma de entender las dinámicas regulatorias de nuestro sistema, fue realizado un análisis de enriquecimiento de sitios de unión de factores transcripcionales entre los promotores de los genes de nuestra red extendida. Los resultados indican que E2F1, c-Myc y Max se encuentran sobre representados. La presencia de estos TF nos brinda coherencia a los resultados obtenidos, ya que la mayoría de la expresión de los genes de nuestra red expandida depende de E2F1, de c-Myc, o de ambos, la sobrerepresentación del compañero canónico de c-Myc, Max, nos muestra que gran parte de los efectores dependientes de este factor transcripcional se encuentran activos. La red de blancos transcripcional de c-Myc se estima en un total de 15% de los genes codificantes del genoma humano, en su red se incluyen genes involucrados en regulación del ciclo celular, metabolismo, biogénesis ribosomal, síntesis de proteínas y funciones mitocondriales, además de reprimir consistentemente genes involucrados en arresto del crecimiento y adhesión celular (Dang et al. 2006), además de que c-Myc se encuentra expresado atípicamente en muchos tipos de cáncer, y especialmente en cáncer cérvico uterino (Sagawa et al. 2001).

De acuerdo a la evidencia, E2F1 puede ser activado por la oncoproteína E7 del VPH-16 de manera independiente de pRB (Hwang et al. 2002), potenciada además por la inactivación de pRB potenciada por las proteínas virales. Los blancos transcripcionales de E2F1 se concentran en síntesis de ADN, como ADN Polimerasa alfa, timidilato sintasa, antígeno nuclear de proliferación celular y ribonucleotido reductasa, así como en genes de regulación como cdc2, ciclina A, y B-myb (De Gregori et al. 1995). Este potencial convierte a E2F1 en un poderoso oncogen cuando pierde su regulación o esta se encuentra comprometida.

5.4. Al añadir capas de información biológica y aumentar el nivel de resolución es posible encontrar nuevas y mas detalladas vías de señalización y de regulación transcripcional.

Comúnmente a este nivel de resolución, muchas preguntas quedan incompletas, es por eso que se buscó integrar una mayor cantidad de información que permitiera realizar un análisis más detallado, con base en este argumento, se buscó realizar un enfoque sistémico utilizando una sola línea celular de cáncer, para este propósito HeLa fue seleccionada por ser la línea celular mejor estudiada, por lo que se cuenta con abundante información sobre ella.

Realizamos el análisis de diferentes capas de información biológica obtenidas de la línea célula HeLa, con el propósito de construir un modelo de vías y meta vías que nos permitiera definir sus relaciones con las características del cáncer, sin perder la perspectiva sistémica (Fig 1).

Hannahan y Weinberg plantearon que existía una intrincada circuitería celular que es capaz de alinearse individualmente con cada característica del cáncer y que probablemente en un futuro cercano, se podría segmentar cada circuito en sub circuitos especializados que soporten propiedades celulares discretas en células normales y reprogramadas, con el objeto de implementar un sello distintivo de la capacidad en células cancerosas, este trabajo es un primer intento en responder al reto que supone poder probar tanto teórica como experimentalmente la existencia de estos sub circuitos y su relación con vías específicas que sustenten funciones celulares específicas.

Se han realizado otros esfuerzos para buscar la integración de diferentes niveles de información en la predicción de vías de señalización que son determinantes para el estado maligno. El grupo de Minn (2012) diseño una metodología novedosa que les permitió la definición de una nueva vía de señalización a partir de un regulador maestro ya conocido y la reconstrucción de la red de dicho regulador mediante un análisis conjunto de genes (GSEA), las interacciones regulatorias son validadas de forma funcional mediante la sobre-expresión o el silenciamiento de algunos genes que participen en la vía propuesta. Esta

estrategia puede ser acoplada a un análisis de sistemas, como el nuestro, lo que permitiría complementar y validar vías nuevas formas no canónicas de dichas vías.

El paso de validación es crítico, debido a que tiene que ser completamente representativo de la red de vías completa, y la técnica más frecuente para realizarlo es la medición directa de mRNA o proteínas (Aldridge et al. 2006). Por lo que planteamos como el primer paso de validación de nuestro trabajo el uso de diferentes tipos de información obtenidos por diferentes grupos de investigación a partir de mRNA y su integración en vías consensuales, siguiendo este razonamiento, a continuación decidimos validar nuestras predicciones iniciales a través de la proteómica, con especial énfasis en la identificación de fosfoproteínas. La validación mediante el análisis fosfoproteómico fue seleccionado debido a que la fosforilación reversible de las proteínas es la modificación postraduccional más comúnmente usado en las células humanas y de mamíferos en general, cerca del 1.7% del genoma humano codifica para proteínas quinasas (Manning et al. 2002) y en un momento dado, cerca del 30% de todas las proteínas del proteoma pueden ser fosforiladas (Hubbard y Cohen 1993). Una fosforilación puede sugerir que una proteína se encuentre en su estado activo, como es el caso de varias enzimas metabólicas, que una señal se encuentra activa, o que existe un cambio conformacional que permite una actividad específica (Mumby y Brekken 2005).

Primeramente, el análisis del total de genes transcritos en la línea celular HeLa arroja un resultado interesante, se encontraron 19,974 genes transcritos lo cual podría suponerse como un número exagerado, sin embargo, se ha demostrado que el promedio de expresión génica encontrado mediante RNA-Seq en diversos tipos de cáncer de mama ronda los 16,245, variando desde 14,648 hasta 18,290 (Eswaran et al. 2012). Al realizar el análisis de ontología génica, observamos que los transcritos cubren al menos el 97% del GO term, lo cual ayuda a comprobar la calidad de la información obtenida y nos prepara para hacer comparaciones con los datos de expresión diferencial.

Al analizar los datos de expresión diferencial de forma clásica, encontramos que existe una sobre-expresión de proteínas oncogénicas que participan en diversos tipos de cáncer, que van desde los factores transcripcionales c-Myc y c-Myb, expresados en una gran variedad de tumores (Musgrove et al. 2008; Fang et al. 2009; Tang et al. 2009; Tanno et al. 2010), pasando por las proteínas de reparación y recombinación de ADN, DNA BRCA1 y 2, amplificadas en cáncer de mama y cáncer de pulmón de células no pequeñas (Antoniou et al. 2002; Rosell et al. 2007), hasta proteínas de “mitotic checkpoint” como BUB1 y BUB3 cuya sola sobre-expresión es capaz de crear inestabilidad genómica (Pinto et al. 2008; Ricke et al. 2011).

Dentro de la perspectiva global a la cual este trabajo se sujeta, encontramos 3,360 genes sobre expresados y 2,129 sub expresados, los GO's y vías resultantes del análisis mediante ConsensusPath nos muestran que en primer lugar existe una clara sobre expresión de genes que coadyuvan ya sea a nivel transcripcional, señalización o metabolismo a mantener la Señal Proliferativa Sostenida (Anexo 3). De igual forma, al analizar los datos concernientes al análisis de factores de transcripción sobre activos, encontramos que las vías de E2F, c-Myc y c-Myb se encuentran claramente regulando positivamente a una gran cantidad de genes que inician mitosis y permiten progresión en el ciclo celular (Palomero et al. 2006; Song et al. 2007; Lefebvre et al. 2010). La sobre expresión de FOXM1 “transcription factor network” se ha relacionado con el aumento en la proliferación celular de carcinomas de próstata en modelos animales (Kalin et al. 2006). Con la integración y validación de los datos obtenidos mediante fosfoproteínas, encontramos que la vía de ciclo celular según KEGG se encuentra sobre-expresada en más del 70% de sus componentes (Fig 7a), así como la totalidad de los transcritos del Complejo de Mantenimiento del Minicromosoma, cuyos niveles incrementados se han relacionado con otros modelos de cáncer (Santin et al. 2005; Majid et al. 2010).

Con relación a la Reprogramación del Metabolismo Energético, podemos observar un claro aumento en la glicólisis y el metabolismo del piruvato-lactato, el cual es expulsado de la célula, acidificando la matriz extracelular y remodelándola al

mismo tiempo (Yamagata et al. 1998; Hirschhaeuser et al. 2011). Sin embargo, la ineficiencia energética de utilizar la glicólisis para obtener ATP se ve compensada al sintetizar substratos para building blocks (lípidos, bases nitrogenadas y péptidos) a partir de los esqueletos de carbono suministrados por la glucosa (Dang 2010; Cairns et al. 2011) (Fig 6).

Una de las enzimas clave en la glucólisis en cáncer es la *Piruvato Quinasa M2*, ya que esta posee una tirosina extra que se encuentra fosforilada, razón que justifica su hallazgo mediante el análisis fosfoproteómico, esta fosforilación inhibe la regulación positiva causada por la *fructuosa 1,6-bisfosfato*, alentando la vía y dejando una gran cantidad de intermediarios fosforilados que pueden ser usados para síntesis anabólica y crecimiento celular (Ferguson y Rathmell 2008). La enorme cantidad de glucosa necesaria para la obtención de energía por parte de la célula es facilitada por la sobreexpresión de los genes que codifican a los transportadores membranales de glucosa GTR3, GTR4, GTR8 y GTR14 (Macheda et al. 2005) y esta permanece dentro de la célula debido a que es fosforilada y convertida en Glucosa-6-F por la HXK2 también sobre expresada en esta línea celular. En este paso, la glucosa-6-F es canalizada dentro de la vía de las pentosas fosfato, en donde encontramos sobre expresadas a las enzimas encargadas de convertirla en ribosas y desoxirribosas (Vizán et al. 2009) (G6PI, K6PF, DEOC, RBSK, KPRA, KPRB y PRPS1) listas para ser utilizadas en el metabolismo de bases nitrogenadas.

A nivel transcripcional, la mayoría de estas enzimas son reguladas por la acción de las redes transcripcionales de c-Myc y las redes del factor de transcripción HIF-1-alfa (Feron 2009), las cuales son encontradas sobre-representadas a nivel de transcrito, FT's y proteínas fosforiladas.

La característica Activación de la Invasión y Metástasis es un tema sumamente complejo, ya que refleja la expresión diferencial de diversas moléculas de adhesión, como las Claudinas, estas últimas presentan un interrelación sumamente compleja y han sido reportadas tanto sub, como sobre expresadas (Fig. 7b) (Prat et al. 2010). En el análisis de expresión diferencial del presente trabajo encontramos sobre expresadas a CLD2 y CLD18, mientras que CLD1,

CLD4 y CLD16 se encuentran sub expresadas. Un signo de que existe la capacidad de realizar metástasis es la aparición de marcadores de Transición Epitelio-Mesénquima (EMT), en el presente trabajo encontramos la sub expresión de OCLN, y la expresión a nivel proteico de VIME y 1433Z (Sarrió et al. 2008; Jing Lu et al. 2009). En cuanto a regulación transcripcional, la vía de las Endotelinas se encuentra sobre regulada debido a la expresión de los genes EDN2 y EDNRA, lo que desemboca en una interferencia entre migración y proliferación celular. Por último, “Signaling events mediated by VEGF pathway” se encuentra activa debido a la sobre-expresión de los factores de crecimiento VEGFA y VEGFB, esta vía permite la angiogénesis, vasculogénesis y crecimiento celular endotelial. Además, representa una interferencia entre diversas características, ya que induce proliferación de células endoteliales, promueve la migración celular, inhibe la apoptosis e induce la permeabilización de vasos sanguíneos (Wouters y Koritzinsky 2008).

A nivel proteico, la familia de transductores de señales 1433 cobra de nuevo un importante papel, ya que su expresión diferencial es preponderante en la proliferación celular, evasión de la apoptosis, adhesión celular, señales mitogénicas y EMT (Seimiya et al. 2000; Porter et al. 2006; Li et al. 2008; Wong et al. 2009). Observamos sub-expresada la forma 1433S en el análisis de expresión diferencial, lo cual concuerda con la mayoría de los modelos, donde esta proteína funciona como supresora de tumores, y su pérdida o disminución correlaciona fuertemente con una pobre prognosis (Benzinger et al. 2005). Adicionalmente se ha observado que es inducida mediante daño al DNA y es requerida para un arresto estable en la fase G2 del ciclo celular, se encuentra regulada directamente por p53 y se ha encontrado silenciada o disminuida en una gran cantidad de carcinomas, su inactivación conduce a la inmortalización de queratinocitos primarios (Hermeking 2003).

A nivel fosfoproteico encontramos fosforiladas las formas 1433B, 1433E, 1433F, 1433G, 1433S y 1433Z. Esta última presenta un antagonismo con la forma 1433S, ya que se ha demostrado que es capaz de sub-regular a p53 en cáncer de mama y su sobre expresión se ha correlacionado con EMT, metástasis y proliferación

celular. Por lo que se propone que esta familia de proteínas transductoras de señales es crítica para el fenotipo maligno (Danes et al. 2008; Niemantsverdriet et al. 2008).

Al realizar el análisis integrativo, se encontró que una gran cantidad de genes del spliceosoma se hallan sobre-expresados, especialmente los del componente U2, donde se encuentra la proteína SRSF1, la cual por si misma se considera oncogénica. Se ha sugerido que cualquier cambio en la estequiometría o actividad de los factores de splicing es capaz de modificar la relación proporcional entre isoformas que normalmente no existen o son menos abundantes en las células normales (Fig. 8). Este fenómeno podría contribuir directa o indirectamente en el desarrollo, progresión y mantenimiento del cáncer. Otra hipótesis sugiere que diversas proteínas con la capacidad de unirse a RNA poseen una gran cantidad de funciones y cambios en su expresión podrían causar efectos oncogénicos no relacionados con su función dentro del spliceosoma (Grosso et al. 2008; Gray et al. 2010). Finalmente, para concretar la integración de las evidencias que poseemos a nivel sistémico, hemos creado un modelo que pretenden demostrar la relación entre los características del cáncer y las vías de señalización, de regulación y metabólicas expresadas diferencialmente por la línea celular HeLa. En la figura 6 observamos lo que hemos llamado “meta-pathways” en donde las vías reconstruidas muestran su relación a nivel de regulación transcripcional, señalización y/o metabolismo. Cada nodo representa una vía y este se encuentra coloreado de acuerdo a las dos características del cáncer más representativos. Una de las primeras cosas que salta a la vista es la gran cantidad de señales que mantienen el estado proliferativo de la célula, además de la gran redundancia que observa en las características. Estos datos sugieren que existe una gran robustez dentro del sistema en donde cada característica se encuentra representada más de una vez y es mantenido por diferentes vías. Dicho comportamiento permitiría explicar el motivo por el cual las terapias dirigidas no han tenido el éxito esperado, ya que al interferir un gen o un conjunto de ellos, sobreviene un cuello de botella (presión de selección) que selecciona positivamente a un fenotipo que permita

mantener las características usando otras vías (Gerlinger y Swanton 2010); Greaves 2010).

5.5. El análisis de proteínas fosforiladas es un paso crítico en la validación de las predicciones realizadas por el modelo.

La elaboración de modelos biológicos con capacidades predictivas es fundamental dentro de la investigación del cáncer, sin embargo, la validación de los modelos sigue siendo complicada, ya que debido a los alcances de los modelos basados en redes y biología de sistemas, se requiere la identificación de varias decenas de mensajeros o proteínas. Los estudios proteómicos basados en espectrometría de masas (MS) son, hasta el momento, una de las mejores técnicas disponibles para dicha validación, ya que permite la detección, y en algunos casos la cuantificación, de una gran cantidad de componentes del sistema analizado de forma reproducible (Sabido et al. 2012).

La mayoría, si no es que todas, de las vías de transducción de señales dependen de la fosforilación reversible de las proteínas para enviar información a través de cascadas de señalización o regulación de proteínas efectoras, como quinasas, factores de transcripción o ubiquitin-ligasas, para obtener finalmente la activación de una vía (Hunter 2000). Por lo cual, la búsqueda de proteínas fosforiladas mediante MS se convierte en la validación ideal para los modelos sistémicos.

En nuestro estudio, se identificaron 272 proteínas fosforiladas que cubren 45 GO's nivel 3, 21 eventos de señalización según PID, 75 vías metabólicas según Reactome y 13 eventos celulares según KEGG. Estos resultados cubren la totalidad de las vías predichas por nuestro modelo, además de aportar datos novedosos sobre la familia de transductores de señales 1433.

De esta forma, consideramos que nuestra validación mediante enriquecimiento de fosfoproteínas seguido por un análisis MS no solo es más preciso y reproducible que un western blot, si no que además arroja valiosa información sobre el sistema de estudio.

5.6. Las aberraciones cromosomales en las células promueven su evolución y la conservación del fenotipo maligno.

Se ha especulado sobre el origen y la participación de la inestabilidad genómica en el cáncer. Existe un debate sobre si es una causa o una consecuencia de la malignidad. Sin embargo, la inserción del cromosoma viral del VPH, la pérdida de función de pRB o una simple mutación alelica como en el caso de BUB1 y BRCA1 son capaces de generar inestabilidad genómica (Negrini et al. 2010, van Harn et al. 2010, Konishi et al. 2011). Se sabe de antemano que a veces, la inestabilidad cromosomal propicia la aparición de células que poseen ciertas formas de inmortalidad, estas células poseen un número modal cromosomal atípico (Artandi y DePinho 2010). Esta inestabilidad cromosomal que permiten la rápida evolución de las células malignas mediante la activación de oncogenes y la inhabilitación de genes supresores de tumores (Negrini et al. 2010).

Esta evidencia nos permite concluir que los defectos en la estructura, vigilancia y reparación del genoma son una ventaja selectiva que permite la progresión tumoral, debido a que aceleran la tasa a la cual las células pre-malignas “evolucionan” acumulando genotipos favorables (Artandi y DePinho 2010). Nuestros resultados muestran que gran parte de los transcritos sobre expresados se concentran en cromosomas que se sabe se encuentran amplificados o tienen zonas de amplificación, ya que la mayoría de los efectores de las vías E2F, c-Myc, p38, Fox y Aurora se encuentran codificados en ellas.

Una mención aparte merece c-Myc, ya que su sobre expresión puede ser explicada por 2 eventos de rearrreglos cromosomales, su locus se encuentra en el cromosoma 8, el cual posee 3 copias normales y 2 cromosomas derivados con la banda 8q24, justo adyacente a los sitios de integración del VPH-18 que posee HeLa lo cual permite su amplificación, sin embargo, esto significa que la amplificación y sobre expresión de c-Myc ocurre después de la inserción del VPH, y seguramente fue provocada por ese evento (Macville et al. 1999).

Estos patrones de expresión cromosomal sugieren que las aberraciones cromosomales son constantes dentro de HeLa, ya que su genoma ha permanecido remarcablemente estable después de años de cultivo continuo (Macville et al. 1999), lo cual sugiere, al contrario de la creencia popular, que todas sus alteraciones genéticas y cromosomales tienen una alta probabilidad de encontrarse en el tumor primario y refleja eventos fundamentales en el desarrollo y mantenimiento del cáncer cérvico uterino. Además nos permite sugerir que las aberraciones cromosómicas no son aleatorias y promueven la rápida mutación y consecuente micro evolución modelada por la presión de selección tisular.

Finalmente, con los datos obtenidos en el presente trabajo, sugerimos que esta robustez sistémica no es conducida al azar, ya que es probable que cuando una célula expresa niveles anormales de proteínas clave, (TF's, transductores de señales, enzimas metabólicas, factores de splicing o incluso, vías que no deberían ser expresadas por ese tipo celular) que les confieran ventajas adaptativas que las habilitan a escapar al gobierno tisular, estos fenotipos serán fijados mediante presión de selección y le permitirán a la célula poseer un arsenal de respuesta a situaciones adversas cada vez más sofisticado.

6. Conclusiones

1.- Existe un patrón de expresión proteica que se conserva en las líneas de cáncer cérvico uterino HeLa, CaLo, SiHa, CaSki, ViBo y C33A. Este patrón involucra proteínas que intervienen en el mantenimiento del fenotipo maligno y se pueden relacionar con las "características del cáncer".

3.- La proteína 1433Z que pertenece a este patrón, posee propiedades topológicas singulares, como un alto node degree y un alto Betweenness, lo que biológicamente puede ser interpretado como un prominente nivel de pleiotropia. Debido a su papel como transductor de señales, esta proteína maneja un gran caudal de información y cualquier variación en su expresión es capaz de causar cambios en el fenotipo celular.

4.- La línea celular HeLa posee perfiles de expresión genética y Ontología Génica que son diferentes de los expresados por la línea celular NHEK.

5.- La red de regulación orquestada por los factores transcripcionales expresados diferencialmente en la línea celular HeLa controla los procesos fundamentales que mantienen el fenotipo neoplásico.

6.- Existe un patrón definido en la activación de circuitos celulares que involucra interconexiones e interferencias, que se encuentran determinados por expresión genética, redes de regulación y vías diferenciales en la línea celular HeLa.

7.- La expresión a nivel cromosomal nos muestra que las múltiples aberraciones que HeLa posee permiten la expresión genética diferencial que es crucial en HeLa para mantener el estado neoplásico. Además es posible concluir que estas alteraciones cromosomales no son al azar, y pueden ser utilizadas como marcadores de diagnóstico y pronóstico.

8.- Nuestro trabajo aporta suficiente evidencia para sostener la hipótesis de que existe una robustez intrínseca en las células cancerosas, cada una de las vías que resultan críticas para mantener el fenotipo neoplásico se encuentran redundantes y existen diversas proteínas que sirven como comodines. Esto puede explicar la resistencia a drogas y a terapias genéticas dirigidas, así como la convergencia de diversos tipos de cáncer a mantener la alteración de ciertas vías de señalización como las dependientes de c-Myc y E2F.

7. Perspectivas

7.1.- Disección de vías de señalización.

Derivado del trabajo bioinformático del presente estudio, surgen incógnitas sobre el funcionamiento no canónico de diversas vías de señalización, por lo cual proponemos una disección de vías de señalización mediante RNA de interferencia y sobre expresión de factores transcripcionales, transductores de señales y

efectores, en cultivos primarios de células epiteliales. El análisis mediante secuenciación masiva de próxima generación, enriquecimiento de proteínas fosforiladas acoplada a espectrometría de masas y proteómica de alta resolución pueden ayudar a establecer los fenotipos obtenidos tras el silenciamiento o sobre expresión y detectar sutiles cambios en el sistema.

Un especial énfasis en el juego de expresión de los FT c-Myc y E2F1 debe ser tomada en cuenta, ya que existe evidencia de regulación de c-Myc sobre miembros de la familia E2F a nivel post-transcripcional a través de miRNAs, por lo cual es necesario realizar un estudio de miRNAs mediante enriquecimiento e identificación por secuenciación masiva, comparando HeLa contra epitelio normal y la posterior sobre expresión de c-Myc y E2F1 en este fondo genético.

7.2.- Análisis de la familia de transductores 1433

En múltiples estudios, incluyendo el presente, la familia de transductores de señales 1433 y en especial 1433z se erigen como un posible “hit” de diversos tipos de cáncer. Poco se sabe de su participación en el contexto del cáncer cérvico uterino. Nuestro estudio fue el primero en sugerir un papel de esta proteína, por lo cual proponemos un análisis de genómica funcional, que incluya modulación de la expresión de todas las proteínas de la familia. El estudio constaría de dos etapas: la construcción de los vectores con cada uno de los miembros de la familia, dependientes de un promotor inducible y un reportero, y su expresión en un cultivo epitelial primario. Posteriormente, para el análisis de los fenotipos, se emplearía secuenciación masiva de próxima generación, enriquecimiento de proteínas fosforiladas acoplada a espectrometría de masas y proteómica de alta resolución, además de análisis de invasividad en matrigel y capacidad tumorigénica en modelos animales.

7.3.-Establecimiento de un modelo usando información de líneas celulares y biopsias

Nuestro modelo se encuentra basado en relativamente poca información, debido principalmente a la falta de datos de secuenciación masiva en diferentes líneas celulares y tumores. Utilizando más recursos, y con acceso a un banco de biopsias, podría ser refinado para incluir diferentes contextos (tipos de infección por VPH, tipos de tumor, grados de lesión) y de esta forma ampliar su capacidad predictiva, además de que podría ser contrastado con nueva información experimental que permitirá pulir los detalles. Al definir las vías de señalización e integrarlas al modelo, el conocimiento biológico del cáncer cérvico uterino podría ayudarnos a encontrar blancos terapéuticos con mayor eficiencia; asimismo, prediciendo el comportamiento y tomando en cuenta la robustez del sistema y la micro evolución, podrían ser superados los actuales fallos en el tratamiento farmacológico y de terapias dirigidas.

8. Bibliografía

- Adams JM, y Cory S. 2007. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26**: 1324–37.
- Aldridge BB, Burke JM, Lauffenburger D A, y Sorger PK. 2006. Physicochemical modelling of cell signalling pathways. *Nature cell biology* **8**: 1195–203.
- Antoniou AC, Pharoah PDP, McMullan G, Day NE, Peto J, Ponder BJ, y Easton DF. 2002. A comprehensive model for familial breast cancer incorporating BRCA1 , BRCA2 and other genes. *British Journal of Cancer* 76–83.
- Apel A, Zentgraf H, Büchler MW, and Herr I. 2009. Autophagy-A double-edged sword in oncology. *International journal of cancer. Journal international du cancer* **125**: 991–5.
- Artandi SE, and DePinho R a. 2010. Telomeres and telomerase in cancer. *Carcinogenesis* **31**: 9–18.
- Astin JW, Batson J, Kadir S, Charlet J, Persad R a, Gillatt D, Oxley JD, and Nobes CD. 2010. Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells. *Nature cell biology* **12**: 1194–204.
- Auffray C, Chen Z, and Hood L. 2009. Systems medicine: the future of medical genomics and healthcare. *Genome medicine* **1**: 2.

- Avruch J, Khokhlatchev a, Kyriakis JM, Luo Z, Tzivion G, Vavvas D, and Zhang XF. 2001. Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade. *Recent progress in hormone research* **56**: 127–55.
- Azar R, Alard A, Susini C, Bousquet C, and Pyronnet S. 2009. 4E-BP1 is a target of Smad4 essential for TGFbeta-mediated inhibition of cell proliferation. *The EMBO journal* **28**: 3514–22.
- Baeriswyl V, and Christofori G. 2009. The angiogenic switch in carcinogenesis. *Seminars in cancer biology* **19**: 329–37.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall K a, Phillippy KH, Sherman PM, et al. 2011. NCBI GEO: archive for functional genomics data sets-- 10 years on. *Nucleic acids research* **39**: D1005–10.
- Bauch A, and Superti-Furga G. 2006. Charting protein complexes, signaling pathways, and networks in the immune system. *Immunological reviews* **210**: 187–207.
- Baudino T a, Maclean KH, Brennan J, Parganas E, Yang C, Aslanian A, Lees J a, Sherr CJ, Roussel MF, and Cleveland JL. 2003. Myc-mediated proliferation and lymphomagenesis, but not apoptosis, are compromised by E2f1 loss. *Molecular cell* **11**: 905–14.
- Benzinger A, Muster N, Koch HB, Yates JR, and Hermeking H. 2005. Targeted proteomic analysis of 14-3-3 sigma, a p53 effector commonly silenced in cancer. *Molecular & cellular proteomics* **4**: 785–95.
- Bergers G, and Benjamin LE. 2003. Tumorigenesis and the angiogenic switch. *Nature reviews. Cancer* **3**: 401–10.
- Berx G, and van Roy F. 2009. Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harbor perspectives in biology* **1**: 1–27.
- Bierie B, and Moses HL. 2006. Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. *Nature reviews. Cancer* **6**: 506–20.
- Blakeley P, Siepen J a, Lawless C, and Hubbard SJ. 2010. Investigating protein isoforms via proteomics: a feasibility study. *Proteomics* **10**: 1127–40.
- Boyer SN, Wazer DE, and Band V. 1996. E7 Protein of Human Papilloma Virus-16 Induces Degradation of Retinoblastoma Protein through the Ubiquitin-Proteasome Pathway. *Cancer Research* **56**: 4620–4624.
- Cairns R a, Harris IS, and Mak TW. 2011. Regulation of cancer cell metabolism. *Nature reviews. Cancer* **11**: 85–95.
- Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L, and Righetti PG. 2004. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**: 1327–33.
- Carmeliet P. 2005. Angiogenesis in life, disease and medicine. *Nature* **438**: 932–6.

- Cavallaro U, and Christofori G. 2004. Cell adhesion and signalling by cadherins and ig-cams in cancer. *Nature reviews. Cancer* **4**: 118–132.
- Cheng N, Chytil A, Shyr Y, Joly A, and Moses HL. 2008. Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion. *Molecular cancer research : MCR* **6**: 1521–33.
- Choi YP, Kang S, Hong S, Xie Xianhe, and Cho NH. 2005. Proteomic analysis of progressive factors in uterine. *Proteomics* 1481–1493.
- Christofferson DE, and Yuan J. 2010. Necroptosis as an alternative form of programmed cell death. *Current opinion in cell biology* **22**: 263–8.
- Clamp M, Fry B, Kamal M, Xie Xiaohui, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, and Lander ES. 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 19428–33.
- Collado M, and Serrano M. 2010. Senescence in tumours: evidence from mice and humans. *Nature reviews. Cancer* **10**: 51–7.
- Colotta F, Allavena P, Sica A, Garlanda C, and Mantovani A. 2009. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis* **30**: 1073–81.
- Consortium TU. 2010. The Universal Protein Resource (UniProt) in 2010. *Nucleic acids research* **38**: D142–8.
- Curto M, Cole BK, Lallemand D, Liu C-H, and McClatchey AI. 2007. Contact-dependent inhibition of EGFR signaling by Nf2/Merlin. *The Journal of cell biology* **177**: 893–903.
- Danes CG, Wyszomierski SL, Lu J, Neal CL, Yang W, and Yu D. 2008. 14-3-3 Zeta Down-Regulates P53 in Mammary Epithelial Cells and Confers Luminal Filling. *Cancer research* **68**: 1760–7.
- Dang CV, O'Donnell KA, Zeller KI, Nguyen T, Osthus RC, Li F. 2006 The c-Myc target gene network. *Seminars in cancer biology* **16**(4):253–64
- Dang CV. 2010. Rethinking the Warburg effect with Myc micromanaging glutamine metabolism. *Cancer research* **70**: 859–62.
- Davies M a, and Samuels Y. 2010. Analysis of the genome to personalize therapy for melanoma. *Oncogene* **29**: 5545–55.
- DeBerardinis RJ, Lum JJ, Hatzivassiliou G, and Thompson CB. 2008. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell metabolism* **7**: 11–20.
- DeGregori J, Kowalik T, Nevins JR. 1995 Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G1/S-regulatory genes. [Internet]. *Molecular and cellular biology* **15**(8):4215–24.

- DeNardo DG, Andreu P, and Coussens LM. 2010. Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity. *Cancer metastasis reviews* **29**: 309–16.
- Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, et al. 2010. The BioPAX community standard for pathway data sharing. *Nature biotechnology* **28**: 935–42.
- Dong D, Ni M, Li J, Xiong S, Ye W, Virrey JJ, Mao C, Ye R, Wang M, Pen L, et al. 2008. Critical role of the stress chaperone GRP78/BiP in tumor proliferation, survival, and tumor angiogenesis in transgene-induced mammary tumor development. *Cancer research* **68**: 498–505.
- Elliott BE, Meens J a, SenGupta SK, Louvard D, and Arpin M. 2005. The membrane cytoskeletal crosslinker ezrin is required for metastasis of breast carcinoma cells. *Breast cancer research : BCR* **7**: R365–73.
- Eswaran J, Cyanam D, Mudvari P, Reddy SDN, Pakala SB, Nair SS, Florea L, Fuqua S a W, Godbole S, and Kumar R. 2012. Transcriptomic landscape of breast cancers through mRNA sequencing. *Scientific reports* **2**: 1–11.
- Fang F, Rycyzyn M a, and Clevenger CV. 2009. Role of c-Myb during prolactin-induced signal transducer and activator of transcription 5a signaling in breast cancer cells. *Endocrinology* **150**: 1597–606.
- Ferguson EC, and Rathmell JC. 2008. New roles for pyruvate kinase M2: working out the Warburg effect. *Trends in biochemical sciences* **33**: 359–362.
- Feron O. 2009. Pyruvate into lactate and back: from the Warburg effect to symbiotic energy fuel exchange in cancer cells. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* **92**: 329–33.
- Ferrara N. 2010. Pathways mediating VEGF-independent tumor angiogenesis. *Cytokine & growth factor reviews* **21**: 21–6.
- Fulda S. 2010. Evasion of apoptosis as a cellular stress response in cancer. *International journal of cell biology* **2010**: 1–6.
- Gentles AJ, and Gallahan D. 2011. Systems biology: confronting the complexity of cancer. *Cancer research* **71**: 5961–4.
- Gerlinger M, and Swanton C. 2010. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British journal of cancer* **103**: 1139–43.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2967073&tool=pmcentrez&rendertype=abstract> (Accessed March 10, 2012).
- Goplen D, Wang Jian, Enger PØ, Tysnes BB, Terzis a J a, Laerum OD, and Bjerkgvig R. 2006. Protein disulfide isomerase expression is related to the invasive properties of malignant glioma. *Cancer research* **66**: 9895–902.
- Gray MW, Lukes J, Archibald JM, Keeling PJ, and Doolittle WF. 2010. Cell biology. Irremediable complexity? *Science* **330**: 920–1.

- Greaves M. 2010. Cancer stem cells: back to Darwin? *Seminars in cancer biology* **20**: 65–70.
- Grivennikov S, and Karin M. 2010. Dangerous liaisons: STAT3 and NF- κ B collaboration and crosstalk in cancer. *Cytokine & growth factor reviews* **21**: 11–19.
- Grivennikov SI, Greten FR, and Karin M. 2010. Immunity, inflammation, and cancer. *Cell* **140**: 883–99.
- Grosso AR, Martins S, and Carmo-Fonseca M. 2008. The emerging role of splicing factors in cancer. *EMBO reports* **9**: 1087–93.
- Gygi SP, and Aebersold R. 2000. Mass spectrometry and proteomics. *Current Opinion in Chemical Biology* 489–494.
- Hamanaka RB, and Chandel NS. 2010. Mitochondrial reactive oxygen species regulate cellular signaling and dictate biological outcomes. *Trends in biochemical sciences* **35**: 505–13.
- Hanahan D, and Folkman J. 1996. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**: 353–64.
- Hanahan Douglas, and Weinberg R a. 2011. Hallmarks of cancer: the next generation. *Cell* **144**: 646–74.
- Hanahan Douglas, and Weinberg RA. 2000. The Hallmarks of Cancer. *Cell* **100**: 57–70.
- Hao P, Zheng S, Ping J, Tu K, Gieger C, Wang-Sattler R, Zhong Y, and Li Y. 2009. Human gene expression sensitivity according to large scale meta-analysis. *BMC bioinformatics* **10 Suppl 1**: S56.
- van Harn T, Foijer F, van Vugt M, Banerjee R, Yang F, Oostra A, Joenje H, and te Riele H. 2010. Loss of Rb proteins causes genomic instability in the absence of mitogenic signaling. *Genes & development* **24**: 1377–88.
- Harris M a, Clark J, Ireland a, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research* **32**: D258–61.
- Hayflick L. 1965. The limited in vitro lifetime of human diploid cell strains. *Experimental cell research* **37**: 614–636.
- Hermeking H. 2003. The 14-3-3 cancer connection. *Nature reviews. Cancer* **3**: 931–43.
- Hirschhaeuser F, Sattler UG a, and Mueller-Klieser W. 2011. Lactate: a metabolic key player in cancer. *Cancer research* **71**: 6921–5.
- Hornberg JJ, Bruggeman FJ, Westerhoff HV, and Lankelma J. 2006. Cancer: a Systems Biology disease. *Bio Systems* **83**: 81–90.

- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* 2009;4(1):44-57.
- Hubbard MJ, and Cohen P. 1993. On target with a new mechanism for the regulation of protein phosphorylation. *Trends in biochemical sciences* **18**: 172–7.
- Hubbard TJP, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, et al. 2009. Ensembl 2009. *Nucleic acids research* **37**: D690–7.
- Hunter T, 2000. Signaling-2000 and beyond. *Cell* **100**:113–27.
- Hwang SG, Lee D, Kim J, Seo T, Choe J. 2002 Human papillomavirus type 16 E7 binds to E2F1 and activates E2F1-driven transcription in a retinoblastoma protein-independent manner. *J Biol Chem* **277**:2923-2930.
- Hydruke DR, and Palsson BØ. 2010. Towards genome-scale signalling network reconstructions. *Nature reviews. Genetics* **11**: 297–307.
- Ikushima H, and Miyazono K. 2010. TGFbeta signalling: a complex web in cancer progression. *Nature reviews. Cancer* **10**: 415–24.
- J. Partanen, A. Nieminen JK. 2009. 3D view to tumor suppression Lkb1, polarity and arrest of oncogenic c-Myc. *Cell Cycle* **8**: 716–724.
- Jiang C, Xuan Z, Zhao F, and Zhang MQ. 2007. TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic acids research* **35**: D137–40.
- Jiang D, Ying W, Lu Yinglin, Wan J, Zhai Y, Liu W, Zhu Y, Qiu Z, Qian X, and He F. 2003. Identification of metastasis-associated proteins by proteomic analysis and functional exploration of interleukin-18 in metastasis. *Proteomics* 724–737.
- Joerger AC, and Fersht AR. 2008. Structural biology of the tumor suppressor p53. *Annual review of biochemistry* **77**: 557–82.
- Jones RG, and Thompson CB. 2009. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes & development* **23**: 537–48.
- Jonsson PF, and Bates P a. 2006. Global topological features of cancer proteins in the human interactome. *Bioinformatics (Oxford, England)* **22**: 2291–7.
- Joy MP, Brock A, Ingber DE, and Huang S. 2005. High-betweenness proteins in the yeast protein interaction network. *Journal of biomedicine & biotechnology* **2005**: 96–103.
- Jr WGK, and Thompson CB. 2010. Q & A Clues from cell metabolism. **465**: 3–5.
- Jung E-J, Moon H-G, Cho BI, Jeong C-Y, Joo Y-T, Lee Y-J, Hong S-C, Choi S-K, Ha W-S, Kim JW, et al. 2007. Galectin-1 expression in cancer-associated stromal cells correlates tumor

- invasiveness and tumor progression in breast cancer. *International Journal of Cancer* **120**: 2331–38.
- Kalin TV, Wang I-C, Ackerson TJ, Major ML, Detrisac CJ, Kalinichenko VV, Lyubimov A, and Costa RH. 2006. Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice. *Cancer research* **66**: 1712–20.
- Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, and Herwig R. 2011. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic acids research* **39**: D712–7.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, and Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research* **38**: D355–60.
- Kessenbrock K, Plaks V, and Werb Z. 2010. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**: 52–67.
- Kitano H. 2002a. Computational systems biology. *Nature* **420**.
- Kitano H. 2002b. Systems biology: a brief overview. *Science (New York, N.Y.)* **295**: 1662–4.
- Konishi H, Mohseni M, Tamaki A, Garay JP, Croessmann S, Karnan S, Ota A, Wong HY, Konishi Y, Karakas B, et al. 2011. Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 17773–8.
- Kroemer G, and Pouyssegur J. 2008. Tumor cell metabolism: cancer's Achilles' heel. *Cancer cell* **13**: 472–82.
- Köcher T, and Superti-furga G. 2007. Mass spectrometry – based functional proteomics : from molecular machines to protein networks. *Nature Methods* **4**: 807–815.
- Ladetto M. 2010. Telomere disrupts, CLL progresses. *Blood* **116**: 1821–2.
- Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, Wang K, Sumazin P, Kustagi M, Bisikirska BC, et al. 2010. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Molecular systems biology* **6**: 1–10.
- Lelièvre S a. 2009. Contributions of extracellular matrix signaling and tissue architecture to nuclear mechanisms and spatial organization of gene expression control. *Biochimica et biophysica acta* **1790**: 925–35.
- Levine B, and Kroemer G. 2008. Autophagy in the pathogenesis of disease. *Cell* **132**: 27–42.
- Li F-Q, Mofunanya A, Harris K, and Takemaru K-I. 2008. Chibby cooperates with 14-3-3 to regulate beta-catenin subcellular distribution and signaling activity. *The Journal of cell biology* **181**: 1141–54.

- Li Z, Zhao J, Du Y, Park HR, Sun S-Y, Bernal-Mizrachi L, Aitken A, Khuri FR, and Fu Haian. 2008. Down-regulation of 14-3-3zeta suppresses anchorage-independent growth of lung cancer cells through anoikis activation. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 162–7.
- Lin C-Y, Chin C-H, Wu H-H, Chen S-H, Ho C-W, and Ko M-T. 2008. Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology. *Nucleic acids research* **36**: W438–43.
- Little MP. 2010. Cancer models, genomic instability and somatic cellular Darwinian evolution. *Biology direct* **5**: 19; discussion 19.
- Locasale JW, and Cantley Lewis C. 2010. Altered metabolism in cancer. *BMC biology* **8**: 88.
- Lu J, Guo H, Treekitkarnmongkol W, Li P, Zhang J, Shi B, Ling C, Zhou X, Chen T, Chiao PJ, et al. 2009. 14-3-3zeta Cooperates with ErbB2 to promote ductal carcinoma in situ progression to invasive breast cancer by inducing epithelial-mesenchymal transition. *Cancer cell* **16**: 195–207.
- Lu Z, Luo RZ, Lu Yiling, Zhang X, Yu Q, Khare S, Kondo S, Kondo Y, Yu Y, Mills GB, et al. 2008. The tumor suppressor gene ARHI regulates autophagy and tumor dormancy in human ovarian cancer cells. **118**: 3917–3929.
- Mac Gabhann F, and Popel AS. 2008. Systems biology of vascular endothelial growth factors. *Microcirculation (New York, N.Y. : 1994)* **15**: 715–38.
- Macheda ML, Rogers S, and Best JD. 2005. Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. *Journal of cellular physiology* **202**: 654–62.
- Macville M, Schröck E, Padilla-Nash H, Keck C, Ghadimi BM, Zimonjic D, Popescu N, Ried T. 1999. Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. [Internet]. *Cancer research* **59(1)**:141–50.
- Majid S, Dar A a, Saini S, Chen Yi, Shahryari V, Liu J, Zaman MS, Hirata H, Yamamura S, Ueno K, et al. 2010. Regulation of minichromosome maintenance gene family by microRNA-1296 and genistein in prostate cancer. *Cancer research* **70**: 2809–18.
- Mani SA, Guo W, Liao M, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al. 2008. The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* 704–715.
- Manning AL, and Dyson NJ. 2011. pRB, a tumor suppressor with a stabilizing presence. *Trends in cell biology* **21**: 433–41.
- Manning G, Plowman GD, Hunter T, and Sudarsanam S. 2002. Evolution of protein kinase signaling from yeast to man. *Trends in biochemical sciences* **27**: 514–20.
- Martin A, Ochagavia ME, Rabasa LC, Miranda J, Fernandez-de-Cossio J, and Bringas R. 2010. BisoGenet: a new tool for gene network building, visualization and analysis. *BMC bioinformatics* **11**: 91.

- Massagué J. 2008. TGF β in cancer. *Cell* **134**: 215–230.
- Masters JR. 2002. HeLa cells 50 years on: the good, the bad and the ugly. *Nature reviews. Cancer* **2**: 311–5.
- Mazzocchi F. 2008. Complexity in biology: Exceeding the limits of reductionism and determinism using complexity theory. *EMBO reports* **9**: 10–14.
- Mendez MG, Kojima S-I, and Goldman RD. 2010. Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **24**: 1838–51.
- Micalizzi DS, Farabaugh SM, and Ford HL. 2010. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *Journal of mammary gland biology and neoplasia* **15**: 117–34.
- Micco R, Cicalese A, Fumagalli M, Dobрева M, Verrecchia A, Pelicci G, and Fagagna F. 2008. DNA damage response activation in mouse embryonic fibroblasts undergoing replicative senescence and following spontaneous immortalization. *Cell Cycle* **7**: 3601–3606.
- Mierke CT, Kollmannsberger P, Zitterbart DP, Diez G, Koch TM, Marg S, Ziegler WH, Goldmann WH, and Fabry B. 2010. Vinculin facilitates cell invasion into three-dimensional collagen matrices. *The Journal of biological chemistry* **285**: 13121–30.
- Minn AJ, Bevilacqua E, Yun J, and Rosner M. 2012. Identification of novel metastasis suppressor signaling pathways for breast cancer. *Cell Cycle* **11**: 2452–2457.
- Mizushima N. 2007. Autophagy: process and function. *Genes & development* **21**: 2861–73.
- Molano, M. Van Den Brule, A. Plummer, M. Weiderpass, E. Posso, H. Arslan A, and Meijer, CJ. Muñoz, N. Franceschi S. 2003. Determinants of Clearance of Human Papillomavirus Infections in Colombian Women with Normal Cytology: A Population-based, 5-Year Follow-up Study. *American Journal of Epidemiology* **158**: 486–494.
- Morrison DK. 2008. The 14-3-3 proteins : integrators of diverse signaling cues that impact cell fate and cancer development. *Trends in Cell Biology* **19**: 16–23.
- Mumby M, and Brekken D. 2005. Phosphoproteomics: new insights into cellular signaling. *Genome Biol.* **6**
- Musgrove EA, Sergio CM, Loi S, Inman CK, Anderson LR, Gardiner-garden M, Alles C, Pinese M, Caldon CE, Schu J, et al. 2008. Identification of Functional Networks of Estrogen- and c-Myc-Responsive Genes and Their Relationship to Response to Tamoxifen Therapy in Breast Cancer. *PloS one* **3**.
- Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJF, and Meijer CJLM. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *The New England journal of medicine* **348**: 518–27.

- Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, and Mann M. 2011. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology* **7**: 548.
- Negrini S, Gorgoulis VG, and Halazonetis TD. 2010. Genomic instability--an evolving hallmark of cancer. *Nature reviews. Molecular cell biology* **11**: 220–8.
- Nelson CM, and Bissell MJ. 2006. Of Extracellular Matrix, Scaffolds, and Signaling: Tissue Architecture Regulates Development, Homeostasis, and Cancer. *Annual review of cell and developmental biology* **22**: 287–309.
- Niemantsverdriet M, Wagner K, Visser M, and Backendorf C. 2008. Cellular functions of 14-3-3 zeta in apoptosis and cell adhesion emphasize its oncogenic character. *Oncogene* **27**: 1315–9.
- Nilsson J a, and Cleveland JL. 2003. Myc pathways provoking cell suicide and cancer. *Oncogene* **22**: 9007–21.
- Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S, and Kanehisa M. 2008. KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic acids research* **36**: W423–6.
- Ostrand-Rosenberg S, and Sinha P. 2009. Myeloid-derived suppressor cells: linking inflammation and cancer. *Journal of immunology (Baltimore, Md. : 1950)* **182**: 4499–506.
- Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, Margolin A, Barnes KC, O'Neil J, Neuberg D, Weng AP, et al. 2006. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 18261–6.
- Papin J a, Hunter T, Palsson BO, and Subramaniam S. 2005. Reconstruction of cellular signalling networks and analysis of their properties. *Nature reviews. Molecular cell biology* **6**: 99–111.
- Park J, Lee D-S, Christakis N a, and Barabási A-L. 2009. The impact of cellular networks on disease comorbidity. *Molecular systems biology* **5**: 262.
- Park S-M, Gaur AB, Lengyel E, and Peter ME. 2008. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & development* **22**: 894–907.
- Perez-Plasencia C, Duenas-Gonzalez A, and Alatorre-Tavera B. 2008. Second hit in cervical carcinogenesis process: involvement of wnt/beta catenin pathway. *International archives of medicine* **1**: 10.
- Perkins DN, Pappin DJ, Creasy DM, and Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551–67.
- Pinto M, Vieira J, Ribeiro FR, Soares MJ, and Henrique R. 2008. Overexpression of the mitotic checkpoint genes BUB1 and BUBR1 is associated with genomic complexity in clear cell kidney carcinomas. *Cellular Oncology* **30**: 389–395.

- Polyak K, and Weinberg R a. 2009. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nature reviews. Cancer* **9**: 265–73.
- Porter GW, Khuri FR, and Fu Haian. 2006. Dynamic 14-3-3/client protein interactions integrate survival and apoptotic pathways. *Seminars in cancer biology* **16**: 193–202.
- Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, and Perou CM. 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research*. **12**: R68.
- Raney BJ, Cline MS, Rosenbloom KR, Dreszer TR, Learned K, Barber GP, Meyer LR, Sloan C a, Malladi VS, Roskin KM, et al. 2011. ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic acids research* **39**: D871–5.
- Reddy RK, Mao C, Baumeister P, Austin RC, Kaufman RJ, and Lee AS. 2003. Endoplasmic reticulum chaperone protein GRP78 protects cells from apoptosis induced by topoisomerase inhibitors: role of ATP binding site in suppression of caspase-7 activation. *The Journal of biological chemistry* **278**: 20915–24.
- Reinhardt HC, and Yaffe MB. 2009. Kinases that control the cell cycle in response to DNA damage : *Current Opinion in Cell Biology*.
- Resendis-Antonio O, Checa A, and Encarnación S. 2010. Modeling core metabolism in cancer cells: surveying the topology underlying the Warburg effect. *PloS one* **5**: e12383.
- Ricke RM, Jeganathan KB, and van Deursen JM. 2011. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *The Journal of cell biology* **193**: 1049–64.
- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**: 651–657.
- Rosell R, Skrzypski M, Jassem E, Taron M, Bartolucci R, Sanchez JJ, Mendez P, Chaib I, Perez-Roca L, Szymanowska A, et al. 2007. BRCA1: a novel prognostic factor in resected non-small-cell lung cancer. *PloS one* **2**: e1129.
- Ryder C, McColl K, Zhong F, and Distelhorst CW. 2012. Acidosis Promotes Bcl-2 Family Mediated Evasion of Apoptosis: Involvement of Acid-Sensing G Protein-Coupled Receptor GPR65 Signaling to MEK/ERK. *The Journal of biological chemistry*.
- Sagawa Y, Nishi H, Isaka K, Fujito A, Takayama M. 2001 The correlation of TERT expression with c-myc expression in cervical cancer. *Cancer Lett*, **168**:45-50.
- Salazar E, Díaz-Mejía JJ, Moreno-Hagelsieb G, Martínez-Batallar G, Mora Y, Mora J, and Encarnación S. 2010. Characterization of the NifA-RpoN Regulon in *Rhizobium etli* in Free Life and in Symbiosis with *Phaseolus vulgaris*. *Applied and environmental microbiology* **76**: 4510–20.

- Salinas S, and Gov N. 2009. The complexity of living: when biology meets theory. Conference on Systems Dynamics of Intracellular Communication. *EMBO reports* **10**: 953–7.
- De Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, and Natoli G. 2007. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130**: 1083–94.
- Santin AD, Zhan F, Bignotti E, Siegel ER, Cane S, Bellone S, Palmieri M, Anfossi S, Thomas M, Burnett A, et al. 2005. Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium : identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology* **331**: 269 – 291.
- Sarrió D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, and Palacios J. 2008. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer research* **68**: 989–97.
- Scardoni G, Petterlini M, and Laudanna C. 2009. Analyzing biological network parameters with CentiScaPe. *Bioinformatics (Oxford, England)* **25**: 2857–9.
- Schadt EE. 2009. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**: 218–23.
- Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, and Buetow KH. 2009. PID: the Pathway Interaction Database. *Nucleic acids research* **37**: D674–9.
- Scheffner M. 1991. The State of the p53 and Retinoblastoma Genes in Human Cervical Carcinoma Cell Lines. *Proceedings of the National Academy of Sciences* **88**: 5523–5527.
- Scheffner M, Werness B a, Huibregtse JM, Levine a J, and Howley PM. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**: 1129–36.
- Schmalhofer O, Brabletz S, and Brabletz T. 2009. E-cadherin, beta-catenin, and ZEB1 in malignant progression of cancer. *Cancer metastasis reviews* **28**: 151–66.
- Scotto L, Narayan G, Nandula SV, Subramaniam S, Kaufmann AM, Wright JD, Pothuri B, Mansukhani M, Schneider A, Arias-Pulido H, et al. 2008. Integrative genomics analysis of chromosome 5p gain in cervical cancer reveals target over-expressed genes, including Drosha. *Molecular cancer* **7**: 58.
- Seimiya H, Sawada H, Muramatsu Y, Shimizu M, Ohko K, Yamane K, and Tsuruo T. 2000. Involvement of 14-3-3 proteins in nuclear localization of telomerase. *The EMBO journal* **19**: 2652–61.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**: 2498–504.

- Shields JD, Kourtis IC, Tomei A a, Roberts JM, and Swartz M a. 2010. Induction of lymphoidlike stroma and immune escape by tumors that express the chemokine CCL21. *Science (New York, N.Y.)* **328**: 749–52.
- Shiozawa Y, Havens AM, Jung Y, Ziegler AM, Pedersen E a, Wang Jingcheng, Wang Jianhua, Lu G, Roodman GD, Loberg RD, et al. 2008. Annexin II/annexin II receptor axis regulates adhesion, migration, homing, and growth of prostate cancer. *Journal of cellular biochemistry* **105**: 370–80.
- Song X-H, Liang B, Liu G-F, Li R, Xie J-P, Du K, and Huang D-Y. 2007. Expression of a novel alternatively spliced variant of NADP(H)-dependent retinol dehydrogenase/reductase with deletion of exon 3 in cervical squamous carcinoma. *International journal of cancer. Journal international du cancer* **120**: 1618–26.
- Sonnenschein C, and Soto AM. 2011. Why systems biology and cancer? *Seminars in cancer biology* **21**: 147–9.
- Sun F, Fu Hanjiang, Liu Q, Tie Y, Zhu J, Xing R, Sun Z, and Zheng X. 2008. Downregulation of CCND1 and CDK6 by miR-34a induces cell cycle arrest. *FEBS letters* **582**: 1564–8.
- Takahashi K, Arjunan SNV, and Tomita M. 2005. Space in systems biology of signaling pathways--towards intracellular molecular crowding in silico. *FEBS letters* **579**: 1783–8.
- Tang Y, Simoneau AR, Liao W, Yi G, Hope C, Liu F, Li S, Xie Jun, Holcombe RF, Journak F a, et al. 2009. WIF1, a Wnt pathway inhibitor, regulates SKP2 and c-myc expression leading to G1 arrest and growth inhibition of human invasive urinary bladder cancer cells. *Molecular cancer therapeutics* **8**: 458–68.
- Tanno B, Sesti F, Cesi V, Bossi G, Ferrari-Amorotti G, Bussolari R, Tirindelli D, Calabretta B, and Raschellà G. 2010. Expression of Slug is regulated by c-Myb and is required for invasion and bone marrow homing of cancer cells of different origin. *The Journal of biological chemistry* **285**: 29434–45.
- Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, et al. 2010. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 15449–54.
- The Fantom Consortium & Riken Omics Science Center. 2009. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genetics* **41**: 553–562.
- Thijssen VL, Barkan B, Shoji H, Aries IM, Mathieu V, Deltour L, Hackeng TM, Kiss R, Kloog Y, Poirier F, et al. 2010. Tumor cells secrete galectin-1 to enhance endothelial cell activity. *Cancer research* **70**: 6216–24.
- Trapnell C, Pachter L, and Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* **25**: 1105–11.

- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, and Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**: 562–78.
- Tzivion Guri, and Avruch Joseph. 2001. 14-3-3 Proteins : Active Cofactors in Cellular Regulation by Serine / Threonine Phosphorylation *. *Journal of Biological Chemistry* **227**: 3061–64.
- Umar S, Sarkar S, Wang Y, and Singh P. 2009. Functional cross-talk between beta-catenin and NFkappaB signaling pathways in colonic crypts of mice in response to progastrin. *The Journal of biological chemistry* **284**: 22274–84.
- Vajdic CM, and van Leeuwen MT. 2009. Cancer incidence and risk factors after solid organ transplantation. *International journal of cancer*. **125**: 1747–54.
- Vizán P, Alcarraz-Vizán G, Díaz-Moralli S, Solovjeva ON, Frederiks WM, and Cascante M. 2009. Modulation of pentose phosphate pathway during cell cycle progression in human colon adenocarcinoma cell line HT29. *International journal of cancer*.
- Wang Z, Gerstein M, and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* **10**: 57–63.
- Warburg O. 1956. On the Origin of Cancer Cells. **123**: 309–314.
- Wertz IE, and Dixit VM. 2010. Regulation of death receptor signaling by the ubiquitin system. *Cell death and differentiation* **17**: 14–24.
- Wessel D, and Flügge UI. 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* **138**: 141–3.
- White E, and DiPaola RS. 2009. The double-edged sword of autophagy modulation in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15**: 5308–16.
- Wolschin F, Wienkoop S, and Weckwerth W. 2005. Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics* **5**: 4389–97.
- Wong EWP, Sun S, Li MWM, Lee WM, and Cheng CY. 2009. 14-3-3 Protein regulates cell adhesion in the seminiferous epithelium of rat testes. *Endocrinology* **150**: 4713–23.
- World Health Organization, 2011
- Wouters BG, and Koritzinsky M. 2008. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nature Reviews Cancer* **8**: 851–64.
- Yamagata M, Hasuda K, Stamato T, and Tannock IF. 1998. The contribution of lactic acid to acidification of tumours: studies of variant cells lacking lactate dehydrogenase. *British journal of cancer* **77**: 1726–31.

- Yamato, K Fen, J Kobuchi, H Nasy, Y Yamada, T Nishishara, T Ikeda, Y Kizaki, M and Yoshinouchi M. 2006. Induction of cell death in human papillomavirus 18-positive cervical cancer cells by E6 siRNA. *Cancer Gene Therapy* 234–241.
- Yang J, and Weinberg R a. 2008. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Developmental cell* **14**: 818–29.
- Yang L, Pang Y, and Moses HL. 2010. TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends in immunology* **31**: 220–7.
- Yu H, Pardoll D, and Jove R. 2009. STATs in cancer inflammation and immunity: a leading role for STAT3. *Nature reviews. Cancer* **9**: 798–809.
- Yuan TL, and Cantley L C. 2008. PI3K pathway alterations in cancer: variations on a theme. *Oncogene* **27**: 5497–510.
- Zeller KI, Zhao X, Lee CWH, Chiu KP, Yao F, Yustein JT, Ooi HS, Orlov YL, Shahab A, Yong HC, et al. 2006. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 17834–9.
- Zhang P, Chan SL, Fu W, Mendoza M, Mattson MP, Drive NS, and Hopkins J. 2003. TERT suppresses apoptotis at a premitochondrial step by a mechanism requiring reverse transcriptase activity and 14- 3-3 protein binding ability. *FASEB journal*. 767–69.
- Zhao B, Wei X, Li W, Udan RS, Yang Q, Kim J, Xie Joe, Ikenoue T, Yu J, Li L, et al. 2007. Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes & development* **21**: 2747–61.
- Zubarev R a, Nielsen ML, Fung EM, Savitski MM, Kel-Margoulis O, Wingender E, and Kel A. 2008. Identification of dominant signaling pathways from proteomics expression data. *Journal of proteomics* **71**: 89–96.

RESEARCH ARTICLE

Open Access

Proteomic patterns of cervical cancer cell lines, a network perspective

Juan Carlos Higareda-Almaraz^{1†}, María del Rocío Enríquez-Gasca^{1,2†}, Magdalena Hernández-Ortiz¹, Osbaldo Resendis-Antonio¹ and Sergio Encarnación-Guevara^{1*}

Abstract

Background: Cervical cancer is a major mortality factor in the female population. This neoplastic is an excellent model for studying the mechanisms involved in cancer maintenance, because the Human Papilloma Virus (HPV) is the etiology factor in most cases. With the purpose of characterizing the effects of malignant transformation in cellular activity, proteomic studies constitute a reliable way to monitor the biological alterations induced by this disease. In this contextual scheme, a systemic description that enables the identification of the common events between cell lines of different origins, is required to distinguish the essence of carcinogenesis.

Results: With this study, we sought to achieve a systemic perspective of the common proteomic profile of six cervical cancer cell lines, both positive and negative for HPV, and which differ from the profile corresponding to the non-tumourgenic cell line, HaCaT. Our objectives were to identify common cellular events participating in cancer maintenance, as well as the establishment of a pipeline to work with proteomic-derived results. We analyzed by means of 2D SDS-PAGE and MALDI-TOF mass spectrometry the protein extracts of six cervical cancer cell lines, from which we identified a consensus of 66 proteins. We call this group of proteins, the “central core of cervical cancer”. Starting from this core set of proteins, we acquired a PPI network that pointed, through topological analysis, to some proteins that may well be playing a central role in the neoplastic process, such as 14-3-3 ζ . *In silico* overrepresentation analysis of transcription factors pointed to the overexpression of c-Myc, Max and E2F1 as key transcription factors involved in orchestrating the neoplastic phenotype.

Conclusions: Our findings show that there is a “central core of cervical cancer” protein expression pattern, and suggest that 14-3-3 ζ is key to determine if the cell proliferates or dies. In addition, our bioinformatics analysis suggests that the neoplastic phenotype is governed by a non-canonical regulatory pathway.

Background

The definition of cancer has evolved according to the knowledge and perspective of the scientific context in which it is conceived. It has changed from a highly heterogeneous disease seen from a cell type and tissue of origin point of view, to the conception of cancer as an illness that involves the deregulation of various pathways that govern key, and somewhat common, cellular processes [1]. Particularly, in 2000 Hanahan and Weinberg suggested that all cancer types represent a manifestation of six essential alterations in cell physiology that

collectively coordinate the malignant phenotype: self-sufficiency in growth signals, insensitivity to growth inhibitors, evasion of programmed cell death, increase of the replicative potential, sustained angiogenesis and tissue invasion and metastasis [2]. Furthermore, in a recent review published by the same two authors, they proposed two emerging hallmarks: reprogramming of energy metabolism and evading immune destruction; besides suggesting genomic instability and mutations, as well as tumor-promoting inflammation, as enabling characteristics [3].

Regardless of the latter, several cancer types have been more intensively studied due to their penetrance in the human population, such as prostate and breast cancer. However, we see in cervical cancer a unique opportunity to study the malignant transformation because of its

* Correspondence: encarnac@ccg.unam.mx

† Contributed equally

¹Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México. Apdo, Postal 565-A, Cuernavaca, Morelos, CP 62210, México
Full list of author information is available at the end of the article

common origin: 90.7% of the cases arise as a consequence of High-Risk Human Papilloma Virus (HR-HPV) infection, according to a study carried out in nine countries of diverse cervical cancer incidences [4]. HR-HPVs encode the E6 and E7 oncoproteins, which interact with the well known and very common, tumor suppressor proteins p53 and pRB, respectively; among several other proteins that together impart a very strong oncogenic potential to the virus [5-7].

To date 120 types of human papillomaviruses have been identified, which can be subdivided into low-risk types, found mainly in genital warts, and high-risk types, which are frequently associated with invasive cervical cancer [8]. Among the high-risk types HPV16 and HPV18 are the most prevalent, present in 54.6% and 11% of squamous cervical carcinomas, respectively [4]. This is part of the reason why, cervical cancers derived from patients infected with those viral types, have been intensively studied, and one of the best studied human cell line, HeLa, is positive for HPV18 [9]. Likewise, it is pertinent to mention that there are other cell lines which originated from HPV negative cervical cancers that have also been widely studied and therefore, enable the visualization of alterations in protein expression, common to many cervical cancer cell lines independently of their origin.

Cervical cancer is one of the most common types of cancer and a major mortality factor of women worldwide [4]. Because most of the cases are a consequence of viral infections, cervical cancer is a disease that has been successfully addressed in developed countries thanks to preventive medicine [10]. Nonetheless, when left unattended, a persistent viral infection combined with a strong and constitutive expression of viral oncoproteins E6 and E7 are highly inductive steps towards the malignant transformation of cervical epithelium [5]. However, owing to a frequent and spontaneous elimination of viral sequences, not all the patients infected with HR-HPV develop cervical cancer [11]. This indicates that most HPV infections are subclinical and only a small fraction of HR-HPV infections produce early epithelial lesions, and a more modest fraction of those lesions will develop into cancer. Consequently, even if infection by HR-HPV can be considered as the initial hit that gives rise to cervical cancer, the superseding steps that enable cancer development have not yet been described [12].

Currently, an accurate prediction of the evolution of the tumor is one of the biggest challenges for clinical oncology. Because of this, the conception of an integrative model that enables the prediction of future states of a system has become of vital importance. We consider that one of the best approaches currently available to accomplish this task are protein-protein interactions

(PPI), because they encompass the scaffold of molecular pathways and cellular processes, besides being capable of revealing the dynamic and interactive function of human proteins [13]. Furthermore, available databases of some organisms have promoted the construction of networks which become the starting point to explore and infer the fundamental principles by which the cell orchestrates its response to different kinds of perturbations.

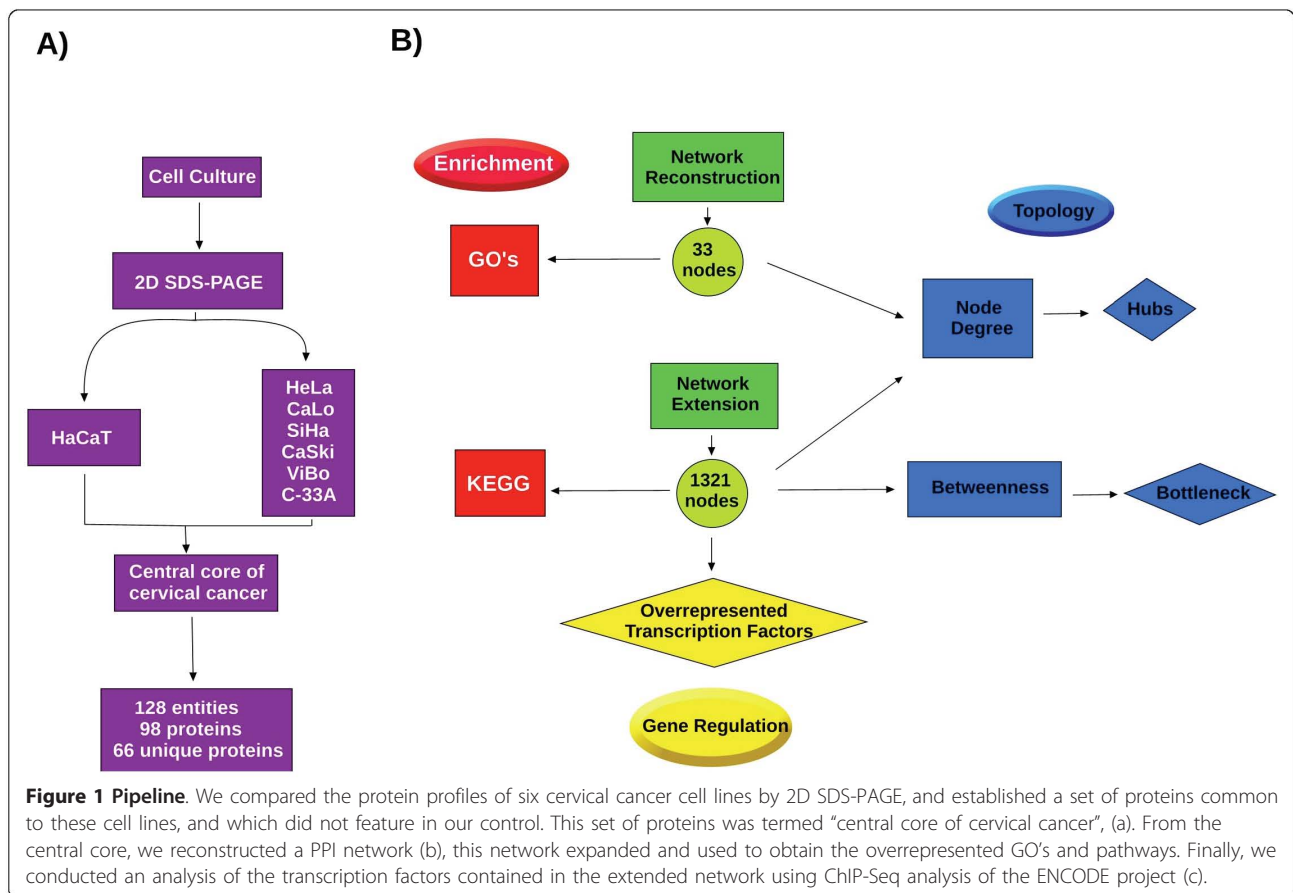
In this work, we developed a pipeline for the functional analysis of differentially expressed proteins in cervical cancer compared with a non-cancerous control and obtained from 2D SDS-PAGE. We used PPI networks, pathways and GOs enrichment analysis, succeeded by an analysis of the promoter gene sequences to get a more systemic point of view of cervical cancer. We consider that the main advantage of following this approach is that it enabled to explore and locate important proteins for the biological phenomenon, which could have been overlooked as a consequence of the experimental methodology used, because some proteins are present in low concentrations or possess isoelectric points or molecular weights located outside the electrophoretic resolution range utilized in 2D SDS-PAGE.

Results

A conserved protein expression pattern exists in cervical cancer cell lines

Protein extracts of the six cancer cell lines: two cell lines which are positive for infection with HPV type 18 (HeLa and CaLo), two positive for HPV type 16 (SiHa and CasKi), and two HPV negative (ViBo and C-33A); and HaCaT, a spontaneously immortalized keratinocyte cell line we used as control, were obtained and analyzed by 2D SDS-PAGE (Figure 1). In our opinion, this experimental design permitted the identification of proteins that are common to the neoplastic phenotype regardless of the presence of intrinsic variations, such as a virus. We set the HaCaT cell line as a control because it is of epithelial origin and has an extended replicative potential, which allows us to distinguish between the biological events of transformation and immortality. Also, it has been widely used as a control in several studies of cervical cancer [14].

Three independent analyses were performed and used for protein profiling. For each sample, 2-D gel electrophoresis was carried out in a pH range of 3 to 10. The Coomassie-stained gels were acquired with densitometer GS-800 (Bio-Rad) and analyzed using PD-Quest software v 8.0.1. On average, more than 1,000 protein spots were revealed in each 2-DE gel. The quantity of each spot in a gel was normalized as a percentage of the total quantity in the map, according to its optical density (OD) value. We focused on the electrophoretic entities



which were present in all six cancer cell lines but did not appear in the HaCaT control, or those which changed consistently and significantly (more than 2-fold). We call this group of 127 proteins, the “central core of cervical cancer”. Among the 127 spots analyzed, 98 proteins were identified based on their tryptic peptide masses using the MALDI-TOF equipment, resulting in 66 unique proteins. (Table 1).

Since this study pointed to a “central core of cervical cancer”, it was necessary to understand the nature of the proteins involved in the process. To this end, 66 proteins were classified via GO (Gene Ontology) using QuickGo of EMBL-EBI database [15]. The identities of the proteins of the central core confirm that these are important in cancer because several studies have linked their aberrant expression to the neoplastic phenotype. With reference to their assigned functions, these proteins can be divided into at least three groups:

The first group includes proteins related to cell migration and metastasis, like anxin 2 that plays a crucial role in establishing metastasis of prostate cancer (PCa) by regulating the adhesion and migration of PCa to osteoblasts and endothelial cells [16]; protein disulfide-isomerase which is strongly expressed by invasive glioma

cells, and its inhibition led to reduced glioma cell migration and invasion [17]; vimentin, an intermediate filament cytoskeletal protein and a marker of Epithelial Mesenchymal Transition (EMT), contributes to invasion and metastasis in prostate cancer [18]; ezrin is required for invasion and metastasis of mammary carcinoma [19]; and finally vinculin, which facilitates contractile force generation, enhancing cell invasion [20]. This group of proteins could be a sign of invasion and metastasis in cervical cancer, similar to their role in other cancer types.

In a second group, we placed proteins related to evasion of apoptosis: GRP78, HSP71, HSP7C, HS90B and GRP75. These proteins are activated as part of the Unfolded Proteins Response (UPR), which has been largely related to survival, cell proliferation and angiogenesis [21]. Likewise, it has been shown that overexpression of GRP78 is sufficient to confer apoptosis resistance in at least two cell types, regardless of its function within the UPR [22]. We believe these proteins could be a link between different hallmarks of cancer, like apoptosis evasion, angiogenesis and proliferation.

In a third group, we find proteins involved in or associated with central metabolism like glyceraldehyde 3

Table 1 Proteins identified as members of the “central core of cervical cancer”

Accession number	Name	Score	Expect	Searched-Matched	Coverage
GRP78_HUMAN	78 kDa glucose-regulated protein	391	1.60E-035	91 - 46	58%
ENOA_HUMAN	Alpha-enolase	327	4.00E-029	82 - 34	82%
GRP78_HUMAN	79 kDa glucose-regulated protein	302	1.30E-026	64 - 32	47%
HSP7C_HUMAN	Heat shock cognate 71 kDa protein	294	8.10E-026	70 - 39	48%
VIME_HUMAN	Vimentin	290	2.00E-025	64 - 33	63%
ENOA_HUMAN	Alpha-enolase	286	5.10E-025	82 - 34	82%
PDIA1_HUMAN	Protein disulfide-isomerase	252	1.30E-021	77 - 30	55%
ANXA2_HUMAN	Annexin A2	249	2.60E-021	65 - 31	72%
TBB5_HUMAN	Tubulin beta chain	247	4.00E-021	73 - 32	62%
ANXA4_HUMAN	Annexin A4	233	1.00E-019	50 - 25	59%
EFTU_HUMAN	Elongation factor Tu, mitochondrial	217	4.00E-018	44 - 22	52%
TPIS_HUMAN	Triosephosphate isomerase	206	5.10E-017	42 - 17	69%
VIME_HUMAN	Vimentin	206	5.10E-017	62 - 23	50%
PGAM1_HUMAN	Phosphoglycerate mutase 1	205	2.10E-016	41 - 18	58%
ATPB_HUMAN	ATP synthase subunit beta, mitochondrial	193	1.00E-015	84 - 44	72%
ACTG_HUMAN	Actin, cytoplasmic 2	192	1.30E-014	50 - 20	48%
PUR9_HUMAN	Bifunctional purine biosynthesis protein PURH	191	1.60E-015	49 - 22	45%
KCRB_HUMAN	Creatine kinase B-type	178	3.20E-014	50 - 19	48%
EF2_HUMAN	Elongation factor 2	154	8.10E-012	67 - 26	33%
HYOU1_HUMAN	Hypoxia up-regulated protein 1	154	8.10E-012	38 - 19	22%
KCRB_HUMAN	Creatine kinase B-type	154	8.10E-012	50 - 19	48%
PPIA_HUMAN	Peptidyl-prolyl cis-trans isomerase A	150	2.00E-011	45 - 16	81%
GELS_HUMAN	Gelsolin	147	4.00E-011	46 - 19	30%
VINC_HUMAN	Vinculin	145	6.40E-011	36-19	20%
WDR1_HUMAN	WD repeat-containing protein 1	144	8.10E-011	46 - 16	31%
EZRI_HUMAN	Ezrin	131	1.60E-009	60 - 26	34%
HSP71_HUMAN	Heat shock 70 kDa protein 1A/1B	129	2.60E-009	42 - 21	33%
TBA1C_HUMAN	Tubulin alpha-1C chain	127	4.00E-009	45 - 13	37%
PRDX1_HUMAN	Peroxiredoxin-1	126	5.10E-009	53 - 18	58%
PDIA3_HUMAN	Protein disulfide-isomerase A3	125	6.40E-009	47 - 15	28%
GRP75_HUMAN	Stress-70 protein, mitochondrial	123	1.00E-008	81 - 34	43%
DHSA_HUMAN	Succinate dehydrogenase flavoprotein subunit	122	3.30E-007	39 - 17	21%
RSSA_HUMAN	40S ribosomal protein SA	121	1.60E-008	37 - 11	34%
HS90B_HUMAN	Heat shock protein HSP 90-beta	119	2.6e-08	44 - 19	27%
ESTD_HUMAN	S-formylglutathione hydrolase	114	8.10E-008	25 - 11	41%
TKT_HUMAN	Transketolase	114	8.10E-008	52 - 19	32%
LDHB_HUMAN	L-lactate dehydrogenase B chain	112	1.30E-007	43 - 20	36%
AL1A1_HUMAN	Retinal dehydrogenase 1	110	2.00E-007	58 - 19	34%
IPYR_HUMAN	Inorganic pyrophosphatase	110	2.00E-007	22 - 9	35%
DDX3X_HUMAN	ATP-dependent RNA helicase DDX3X	108	3.20E-007	23 - 13	20%
DHE3_HUMAN	Glutamate dehydrogenase 1, mitochondrial	107	4.00E-007	38 - 13	29%
G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase	107	4.00E-007	55 - 13	42%
ACON_HUMAN	Aconitate hydratase, mitochondrial	105	6.40E-007	54 - 23	30%
ML12A_HUMAN	Myosin regulatory light chain 12A	105	6.40E-007	32 - 11	63%
RIR1_HUMAN	Ribonucleoside-diphosphate reductase large subunit	101	1.60E-006	28 - 15	18%
EF2_HUMAN	Elongation factor 2	99	2.60E-006	60-20	29%
QCR1_HUMAN	Cytochrome b-c1 complex subunit 1, mitochondrial	98	3.30E-006	45 - 15	29%
CNN2_HUMAN	Calponin-2	96	4.50E-006	51 - 13	48%
PCBP1_HUMAN	Poly(rC)-binding protein 1	94	8.30E-006	39 - 14	44%
IMMT_HUMAN	Mitochondrial inner membrane protein	92	1.30E-005	34 - 13	20%
CAPG_HUMAN	Macrophage capping protein	91	1.80E-005	41 - 11	30%

Table 1 Proteins identified as members of the “central core of cervical cancer” (Continued)

TCTP_HUMAN	Translationally-controlled tumor protein	91	1.80E-005	19 - 8	40%
PSB4_HUMAN	Proteasome subunit beta type-4	89	1.70E-003	20 - 7	34%
TERA_HUMAN	Transitional endoplasmic reticulum ATPase	88	3.10E-005	49 - 15	21%
K2C1_HUMAN	Keratin, type II cytoskeletal 1	87	3.70E-005	46 - 13	29%
LEG1_HUMAN	Galectin-1	84	7.70E-005	41 - 8	51%
1433Z_HUMAN	14-3-3 protein zeta/delta	80	2.30E-004	43 - 14	54%
LMNA_HUMAN	Prelamin-A/C	78	3.40E-004	37 - 11	22%
HNRPL_HUMAN	Heterogeneous nuclear ribonucleoprotein L	77	4.10E-004	27 - 13	16%
TCPH_HUMAN	T-complex protein 1 subunit eta	77	4.00E-004	36 - 15	30%
OSBL8_HUMAN	Oxysterol-binding protein-related protein 8	73	1.10E-003	56 - 13	21%
TCPE_HUMAN	T-complex protein 1 subunit epsilon	71	1.70E-003	30 - 11	23%
HYOU1_HUMAN	Hypoxia up-regulated protein 1	70	2.20E-003	28 - 13	14%
PSA5_HUMAN	Proteasome subunit alpha type-5	69	2.70E-003	17 - 6	25%
FOXP3_HUMAN	Forkhead box protein P3	68	3.40E-003	41 - 12	34%

The entry name corresponds to the UniProt database, the score and expect value were obtained from the search engine Mascot. The expect value is the number of matches with equal or better scores that are expected to occur by chance alone. Score as $-10 \cdot \log_{10}(P)$, where P is the absolute probability. The lower the expectation value, the more significant the score.

phosphate dehydrogenase, phosphoglycerate mutase 1, enolase A, triosephosphate isomerase and L-lactate dehydrogenase B. We consider that the expression of this group of proteins could be due to the cancerous cells' need to divide their incoming nutrients between energy production and macromolecular biosynthesis to support cell growth and DNA replication. This increase in cell metabolism has been described in cancer cells, which exhibit a much higher intake of glucose than normal cells, as well as increased rates of glycolysis and lactate production even in the presence of oxygen [23-26].

Another protein detected by us as a member of the “central core of cervical cancer” was galectin-1. Remarkably, very recently it was reported that cells can stimulate tumor angiogenesis by secretion of this protein, and it was previously found in the vasculature of many human tumors, including colon, head and neck, lung, prostate, and oral cancers [27,28]. All of the proteins identified as members of the “central core of cervical cancer” are summarized in Table 1.

14-3-3 ζ is a Hub in our protein-protein interaction network

We performed a theoretical interactome (Figure 2) using the Cytoscape plugin, Bisogenet, to add the experimentally verified interactions between the 66 identified proteins. This plugin employs the SysBiomics database which integrates information from the INTACT [29], BIOGRID [30], MINT [31], DIP [32], BIND [33] and HPRD [34] databases.

Even though we identified 66 members of the central core of cervical cancer, we were able to integrate into a network only half of them because of a lack of experimental evidence supporting PPIs. Notably, through a topological analysis carried out on this network, we

found that the 14-3-3 ζ protein features as a highly interconnected node (Figure 2). Additionally, western blot analysis confirmed that all cervical cancer cell lines express high protein levels of 14-3-3 ζ . In contrast, no blot of this protein was found in our control (Figure 3).

Afterwards, we expanded the PPI network as a way to integrate all the proteins of the central core, in order to obtain a coherent output which could be related in its entirety to the neoplastic phenotype. It also served the purpose of broadening our perspective and strengthening the statistical significance of our results.

The network expansion was performed considering the experimentally obtained proteins as bait; the expansion criterion was to include proteins that have a direct experimental interaction in the databases, and were connected via a shortest path of 1 to our proteins. This resulted in a network of 1,321 nodes and 9,666 edges (Figure 4).

We considered that the network expansion would allow us to place the entities belonging to our central core in a broader context, specially because the amount of proteins that were identified represent a small fraction of all the entities responsible for the phenotype. We also expected to observe the same topological phenomena in the expanded network as we did in the network composed solely of cervical cancer-specific proteins.

A connectivity distribution analysis allowed us to identify those proteins that highly interact with others, defined as the hubs of the expanded network. This pointed once again to the multifunctional regulator of cell signal transduction and adaptor protein, 14-3-3 ζ , followed by 14-3-3 γ with less than half the number of interactions. Also, we calculated the betweenness of all the proteins in the network, because it is a feature

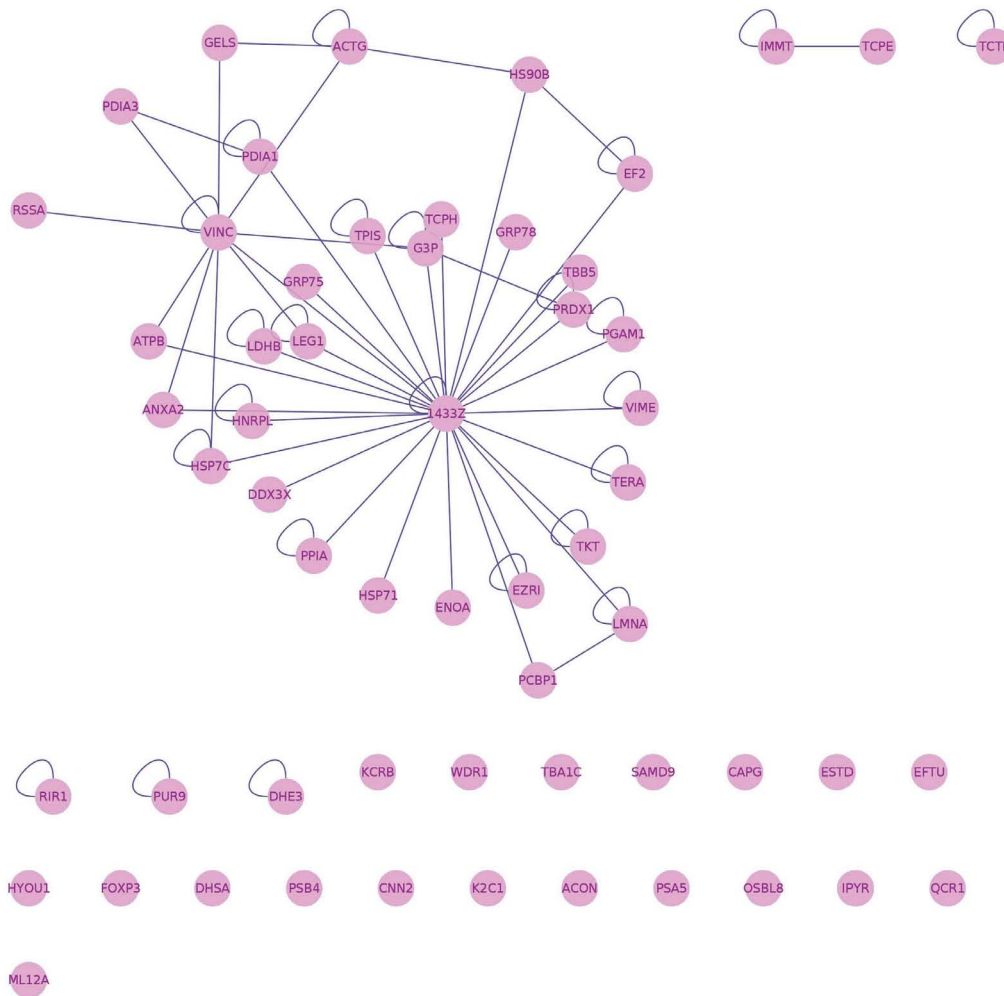


Figure 2 Network Reconstruction. Network of differentially expressed proteins in cervical cancer cell-lines compared to a non-cancerous control. The connections between proteins represent experimentally verified protein-protein interactions between 33 out of the 66 proteins identified in this study. The proteins not included in this network do not have experimentally proven interactions. The network was constructed using Cytoscape and the plugin Bisogenet.

which allows us to explore the connection between local network structure and global network topology, besides representing a measure of a node's influence over information transfer [35,36]. The proteins with the highest

betweenness were 14-3-3 ζ , vimentin and vinculin, part of our central core of cervical cancer proteins, which is consistent with their role as bait proteins. However, there were also non-bait nodes that featured among the proteins with highest betweenness, such as ubiquitin and 14-3-3 γ ; thus, supporting the notion that the betweenness metric, though biased in this case, is still a valid metric of essentiality and robustness.

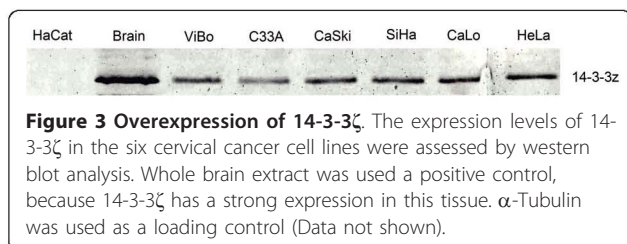
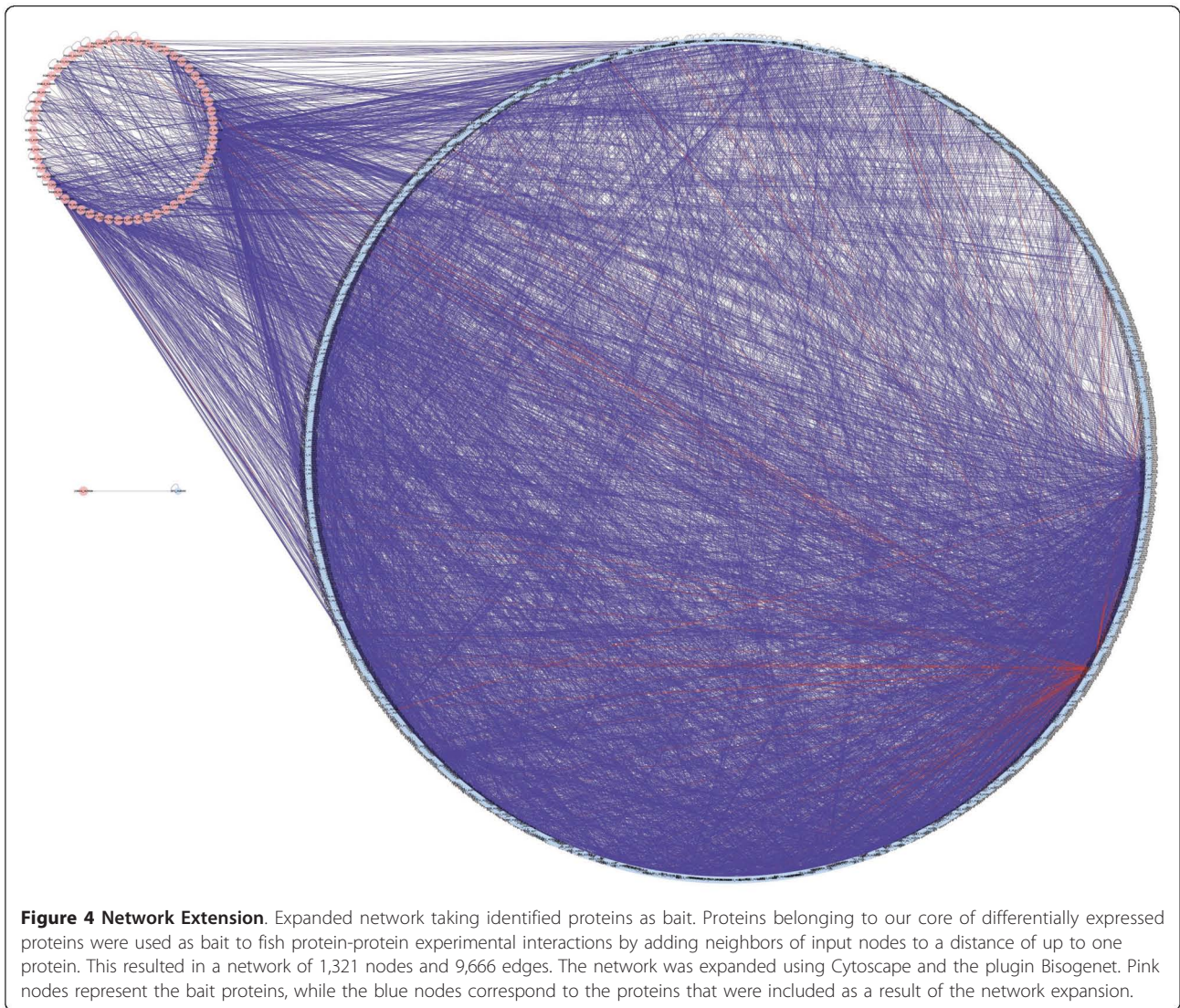


Figure 3 Overexpression of 14-3-3 ζ . The expression levels of 14-3-3 ζ in the six cervical cancer cell lines were assessed by western blot analysis. Whole brain extract was used a positive control, because 14-3-3 ζ has a strong expression in this tissue. α -Tubulin was used as a loading control (Data not shown).

The enrichment analysis of gene ontology level 3 category of biological processes shows 16 overrepresented GOs (Table 2); the enriched KEGG pathway-based analysis shows 16 significant pathways (Table 3). Although both enrichments are of great value for the analysis of



the biological significance of our proteins, these have the limitation of considering the same protein or set of proteins as part of different, and sometimes unrelated, processes or pathways (Table 3). Consequently, results should be carefully assessed before making further interpretations.

There are common regulatory features among the proteins of the expanded network

Using the information generated by the ENCODE project, we searched among the transcription factors (TF) that had chromatin immunoprecipitation-sequencing (ChIP-Seq) data available, for overrepresented binding sites of a given TF in the genes of our expanded network (Figure 1). We evaluated for enrichment by considering binding sites that were located up to 700 bp upstream or 300 bp downstream of all the transcription start sites reported in the GENCODE database and that

had a score larger than 750. We considered these 1,000 bp as the promoter sequence, because it has been reported to be the region of highest ChIP-Seq peak concentration in previous studies [37].

For the promoters of genes that belong to our expanded network, we focused again on the information provided in GENCODE. Out of the 1,321 nodes in the network we retrieved 1,431 promoters corresponding to 957 unique genes of the network. We found that the transcription factors E2F1 (p-value of 4.81E-13), TCF4 (p-value of 1.46E-12), c-Myc (p-value of 5.52E-11), Max (p-value of 9.77E-11), E2F6 (p-value of 1.17E-10) and NFkB (p-value of 5.02E-10) were the most significantly overrepresented by a hypergeometric test (Table 4). Importantly, the ChIP-Seq assays of the E2F1, c-Myc and Max transcription factors were performed on the HeLa-S3 cell line <http://genome.ucsc.edu/>. The fact that these experiments were performed in one of the cell

Table 2 Enriched gene ontology level 3 categories of biological processes

	<i>gene ontology term</i>	<i>set size</i>	<i>candidates contained</i>	<i>p-value</i>	<i>q-value</i>
GO:0006986	response to unfolded protein	73	8	3.11E-009	7.03E-007
GO:0070841	inclusion body assembly	13	4	4.26E-007	4.82E-005
GO:0006091	generation of precursor metabolites and energy	363	11	2.02E-006	1.52E-004
GO:0032507	maintenance of protein location in cell	80	6	3.14E-006	1.77E-004
GO:0016052	carbohydrate catabolic process	132	7	4.67E-006	1.93E-004
GO:0051651	maintenance of location in cell	87	6	5.13E-006	1.93E-004
GO:0012501	programmed cell death	1338	20	7.18E-006	2.32E-004
GO:0010941	regulation of cell death	1136	18	1.06E-005	2.99E-004
GO:0046907	intracellular transport	855	15	2.08E-005	5.21E-004
GO:0002349	histamine production involved in inflammatory response	2	2	2.56E-005	5.25E-004
GO:0043455	regulation of secondary metabolic process	2	2	2.56E-005	5.25E-004
GO:0002443	leukocyte mediated immunity	184	7	4.05E-005	7.64E-004
GO:0044248	cellular catabolic process	1028	16	4.56E-005	7.92E-004
GO:0050777	negative regulation of immune response	40	4	4.91E-005	7.93E-004
GO:0070727	cellular macromolecule localization	610	12	5.29E-005	7.98E-004
GO:0051235	maintenance of location	133	6	5.79E-005	8.18E-004

Set size refers to the number of entities that have a Uniprot ID as stated in the corresponding GO category at the ConsensusPathDB site. The number of candidates contained refers to amount of proteins identified in this study that appear as part of the GO category. P-values are calculated according to a hypergeometric test; q-values represent p-values corrected for multiple testing using the false discovery rate method.

lines we are investigating gives this bioinformatic analysis additional validity, because the cellular context is unlikely to be significantly different.

Also, it is noteworthy to mention that we performed a similar analysis, but using the promoter regions as reported in the Transcriptional Regulatory Element Database (TRED) [38] and employing the SwitchGear Database for the acquisition of the transcription start sites. Using this approach, we obtained a similar enrichment (Table 5).

Among the overrepresented TFs, we consider that the presence of *c-Myc* is remarkable because it is in agreement to what is currently known about cancer cells in general. Also, *c-Myc*'s primary binding partner for gene activation is Max, so that the presence of this protein is complementary to *c-Myc*'s activity. The mechanisms that make of *c-Myc* a very powerful oncogene are not very clear; however, there is evidence that indicates that the activation of *c-Myc* correlates with approximately 70% of human cancers [39].

Table 3 Enriched KEGG pathway-based sets

	<i>pathway name</i>	<i>set size</i>	<i>candidates contained</i>	<i>p-value</i>	<i>q-value</i>	<i>pathway source</i>
	Small cell lung cancer	84	37	8.67E-009	1.43E-006	KEGG
	Protein processing in endoplasmic reticular	164	57	4.21E-008	3.47E-006	KEGG
	Pathways in cancer	325	94	8.81E-008	4.05E-006	KEGG
	Apoptosis	87	36	9.82E-008	4.05E-006	KEGG
	Neurotrophin signaling pathway	125	46	1.24E-007	4.10E-006	KEGG
	T cell receptor signaling pathway	108	41	2.25E-007	6.19E-006	KEGG
	Focal adhesion	199	62	9.45E-007	2.23E-005	KEGG
	Prostate cancer	89	33	6.11E-006	1.26E-004	KEGG
	Pathogenic Escherichia coli infection	54	23	1.13E-005	2.07E-004	KEGG
	Endometrial cancer	52	22	1.99E-005	3.28E-004	KEGG
	Shigellosis	61	24	3.61E-005	5.42E-004	KEGG
	Non-small cell lung cancer	54	22	4.03E-005	5.54E-004	KEGG
	Proteasome	44	19	5.17E-005	6.56E-004	KEGG
	Bacterial invasion of epithelial cells	70	26	5.59E-005	6.59E-004	KEGG
	B cell receptor signaling pathway	75	27	7.62E-005	8.38E-004	KEGG
	Leukocyte transendothelial migration	116	37	8.27E-005	8.53E-004	KEGG

Set size refers to the number of entities that have a Uniprot ID as stated in the corresponding KEGG pathway-based set at the ConsensusPathDB site. The number of candidates contained refers to amount of proteins which are part of the extended network and appear as part of the pathway. P-values are calculated according to a hypergeometric test; q-values represent p-values corrected for multiple testing using the false discovery rate method.

Table 4 Significantly overrepresented transcription factors identified by ENCODE ChIP-Seq peaks based on the GENCODE Database

Transcription Factor	cell line	tissue of origin	peaks near TSS of the gene of the extended network	peaks near TSS of all the GENCODE genes	p-value
E2F1	HeLa-S3	cervical	187	1445	4.81E-013
TCF4	HCT-116	colorectal	244	2059	1.46E-012
Pol2	HeLa-S3	cervical	312	2878	3.93E-011
c-Myc	HeLa-S3	cervical	123	880	5.52E-011
Max	HeLa-S3	cervical	161	1266	9.77E-011
E2F6	k562	Leukemia	265	2379	1.17E-010
NFKB	GM12878	Lymphoblastoid	111	794	5.02E-010

The peaks were mapped to regions corresponding to the promoter sequences as considered for this study (-700, 300 bp), based upon transcription start sites (TSS) annotated by GENCODE. Peaks near TSS of the genes of the extended network and all GENCODE genes refers to the amount of peaks found in the promoter region. P-values were calculated by a hypergeometric test. Sample size was considered as the amount of genes that corresponded to the proteins of the extended network. Population size was the number of annotated GENCODE entries that had a complete status and were separated by a distance of up to 500 base pairs.

The first possible explanation for c-Myc's prevalence in cancer is that the genes it induces represent the primary response of nearly all signal transduction pathways known to be involved in cancer [39]. Likewise, Nilsson *et al.*, [39] have suggested that a loss in c-Myc's ability to induce apoptosis can convert c-Myc into a pure promoter of cell growth and transformation. Concordantly, the abrogation of p53 function, which is the protein responsible for executing c-Myc's apoptotic response, has been substantiated in all HPV positive cell lines, as well as the HPV negative cell line, C33-A [40].

The last factor, E2F1, controls cell-cycle progression and DNA replication and it has been shown to be induced by c-Myc's activity [41]. E2F1 has also been implicated in c-Myc mediated down regulation of p27^{KIP1}, which is a cyclin-dependent kinase inhibitor [42]. Finally, polymerase 2 featured as the DNA-binding protein with the most significant p-value, thereby providing us with a notion of the genes of the network that are being transcribed (Tables 4 and 5).

Discussion

We performed a comprehensive analysis that stems from data obtained by proteomic techniques, using perspectives of biological networks and the annotated data of the human genome and large-scale sequencing techniques available to the public as a tool for building models to infer biological behaviors (Figure 1).

As defined in this study, the "central core of cervical cancer" does not describe the immortalization process directly; rather, it represents processes involved in maintaining a viable and proliferating cell. Table 2 presents the biological processes that are enriched in our core set of proteins and as can be seen, these are suggestive of a malignant phenotype, where the main GOs are those related to UPR, metabolism, catabolic processes, maintenance of cell-location and cell death. These observations are consistent with our initial conjectures, since we considered that using the HaCaT cell line as a control would allow us to distinguish between immortalization and transformation, because HaCaT is an immortalized cell line with a stable karyotype and phenotype, which is non-tumorigenic.

Table 5 Significantly overrepresented transcription factors identified by ENCODE ChIP-Seq peaks based on the TRED Database

Transcription Factor	cell line	tissue of origin	peaks near TSS of the genes of the extended network	peaks near TSS reported on SwitchGear	p-value
Pol2	HeLa-S3	cervical	608	7766	2.15E-067
TCF4	HCT-116	colorectal	455	5543	1.93E-050
E2F6	k562	Leukemia	488	6343	1.12E-046
Max	HeLa-S3	cervical	313	3375	2.90E-042
c-Myc	HeLa-S3	cervical	250	2395	1.88E-041
NFKB	GM12878	Lymphoblastoid	195	1802	6.78E-034
E2F1-HA	HeLa-S3	cervical	315	4272	7.57E-024

ChIP-Seq peaks mapped to regions corresponding to the promoter sequences as considered for this study (-700, 300 bp), based upon transcription start sites (TSS) annotated by SwitchGear. Peaks near TSS of the genes of the extended network and reported on SwitchGear refers to the amount of peaks found in the corresponding promoter regions. P-values were calculated by a hypergeometric test. Sample size was considered as the amount of genes that corresponded to the proteins of the extended network. Population size was the number of reported TSS reported on SwitchGear with a score greater than 20.

One of the studies that uses the HaCaT cell line as a control was performed by Choi, *et al.*[14]. The main difference between the latter study and our work, is that they used cervix biopsies; this implies that there are important differences in the intracellular and extracellular contexts of the biological model. Likewise, even though the methodology they utilized only considers samples with the same pathologic diagnosis, it is likely that there is still an important amount of heterogeneity in their samples due to inter-individual differences. As expected, we only found one common protein among the results of both studies: Vimentin. However, they were able to identify a number of proteins that strengthen our findings such as Amy-1, which stimulates c-Myc's E-box dependent transactivation activity [43]; and Miz-1, which must be inhibited in order for c-Myc to mediate apoptosis [44]. Another result that supports our findings is that they observed a downregulation of 14-3-3 σ , which plays an opposite role in cell growth, compared to 14-3-3 ζ [45]. Finally, they also found a clear overexpression of chaperone proteins, different from the ones in our central core.

The acquisition of the reported interactions between the identified proteins provided a first insight of the key elements in the neoplastic dynamics, and became the meter that allowed us to judge the validity of later analysis, by seeking an overall agreement between the original and the extended network. It is noteworthy that for the small network the highest interconnected node was 14-3-3 ζ , and this was not lost when we expanded the network. Also, both networks are enriched in similar processes.

When analyzing the extended network it is apparent that 6 of the members of the 14-3-3 protein family (ζ , γ , β , τ , σ and ϵ) feature among the highest interconnected nodes. This is consistent with the biological role of the members of this protein family as signal transduction molecules [46]. The identification of 14-3-3 ζ and its role as a hub and as a bottleneck in the PPI network is particularly relevant because it is known to be involved in three important cellular processes: cell cycle regulation, signal transduction and regulation of apoptosis [47-50].

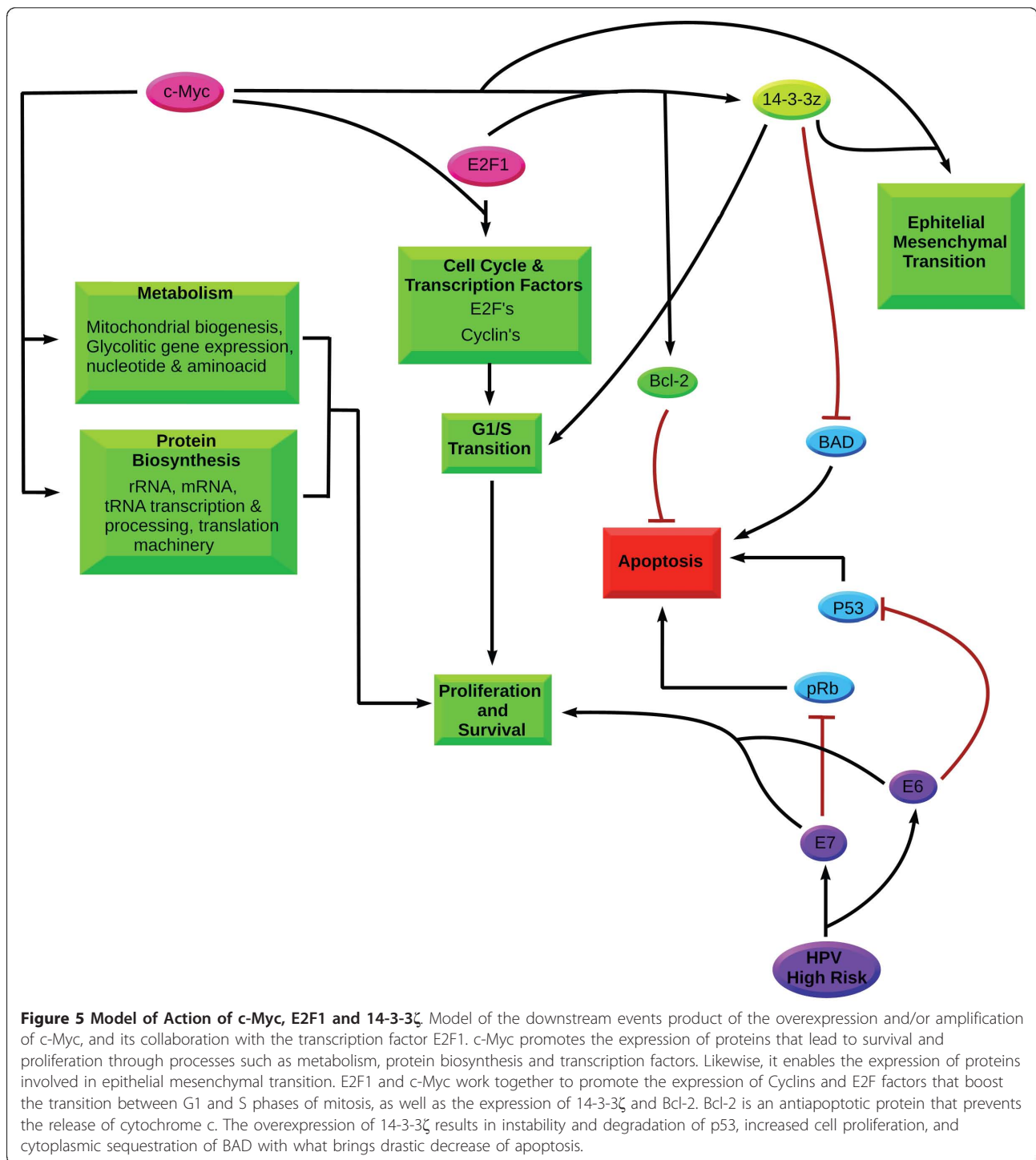
Furthermore, we noticed that the 14-3-3 ζ protein, which features as the highest interconnected node of both our networks, did not possess any ChIP-Seq peak near the transcription start site of its encoding gene. Upon more careful inspection throughout the entire length of the gene, we noticed that there appears to be a regulatory region inside the transcript. Importantly, we found binding sites of c-Myc, Max, E2F1 and Pol2 very close to each other. This has led us to believe that one of the mechanisms that lead to 14-3-3 ζ 's overexpression is mediated by c-Myc and E2F1 (Figure 5).

It was previously reported by Jonsson and Bates [13] that cancer proteins exhibit a significantly different network topology when compared to proteins unrelated to cancer; as well as a higher ratio of promiscuous structural domains. This is in agreement with what is known about the members of the 14-3-3 protein family, which have been described as signal integrators, amplifying strong signals and filtering out weaker conflicting ones to achieve a meaningful, coordinate biological output, such as cell death or survival [51].

Also, there is strong evidence that the overexpression of 14-3-3 ζ promotes p53 degradation by the proteosomal route [52], which is increased by the action of E6 in HPV-infected cells [7]. An enhanced turn-over rate of the p53 protein and the pleiotropic effects consequence of the increased number of 14-3-3 ζ proteins, which are involved in a large number of cellular processes, could be the reason why not all infections develop into cancer and most of them are subclinical: c-Myc could be activating 14-3-3 ζ in a differential manner. Therefore, we consider that we possess sufficient evidence to suggest the role of 14-3-3 ζ in cervical cancer cell lines as the determining factor that can ultimately dictate the fate of a cell, regardless of it being infected with HPV; and cause it to undergo cell cycle deregulation onto malignant transformation (Figure 5).

Likewise, when we performed a biological pathway enrichment, which we consider the best way to perform a functional analysis, we found consistent results. To do this, we used the ConsensusPathDB site to find overrepresented KEGG pathways in the proteins of our extended network. The enrichment analysis showed an overrepresentation of survival factors of apoptosis pathways and focal adhesion that is related to invasion and metastasis, two of the most important and well-documented hallmarks of cancer; as well as various established cancer pathways like small cell lung cancer, prostate cancer, endometrial cancer and pathways in cancer (Table 3). This was obviously expected and helps to strengthen our analysis, indicating that the central core of cervical cancer, and its subsequent expansion proposed in the PPI interaction network, still retain those proteins that are closely related to the neoplastic process.

However, there are other pathways that do not present such a straightforward connection to cancer, such as the neurotrophin and T-cell receptor signaling pathways, bacterial infections and proteasome pathways. We can only speculate about their meaning, as the cellular stress, chronic inflammation, turnover of proteins, cell survival and evasion of the immune system are other features related to the progression and maintenance of tumorigenesis.



As a way to understand the underlying regulatory dynamics of the system, we also performed an enrichment analysis of transcription factor binding sites among the promoters of the genes of our extended network. We focused on the three transcription factors that complied with both of the following: those which were significantly overrepresented among the promoter

regions of the genes of the proteins of our network and those in which the corresponding ChIP-Seq assays were carried out on the HeLa-S3 cell line <http://genome.ucsc.edu/>. The resulting transcription factors are: E2F1, c-Myc and Max. We believe that together, these transcription factors bring an overall feeling of coherence to the network because the overexpression of c-Myc is

widespread in human cancers, and specifically in cervical cancer [53]. Moreover, there is evidence that E2F1 is activated by the E7 protein of HPV-16 in a pRB-independent manner [54], apart from its pRB-dependent activation, which in the case of HPV infected cells is also promoted by pRB inactivation by viral oncoprotein E7. In the context of cervical cancer, it has been reported that an overexpression of E2F1 can drive quiescent cells through G1 into S-phase of the cell cycle, ultimately leading to apoptosis or neoplastic transformation.

E2F1 has been shown to be an inhibitor of *c-Myc*'s activator, β -catenin, by means of the Wnt pathway [55,56]. We also found a clear overrepresentation of the TCF4 transcription factor in the genes of our network. This contradicts our previous finding because TCF4 functions in the activation of the Wnt pathway, and therefore, of *c-Myc*. However, we had to consider that all the published works that point to E2F1 as an inhibitor of *c-Myc* have been performed in colorectal cancer. Likewise, the evidence of enrichment of TCF4 binding sites was performed in the HCT-116 cell line and so we do not possess information about the function of this transcription factor in cervical cancer.

It has been observed in human foreskin keratinocyte and fibroblast (HFK and HFF) cells that the genetic background of the cell type can be determinant for the outcome of altered gene expression. This is the case of HFF which cannot be immortalized by the mere addition of oncoproteins E6 and E7, unlike the HFK cells, due to a differential turnover of *c-Myc* [57]. Moreover, as pointed by Bernards [58] in the context of colorectal cancer pRB is more likely to be acting as an oncoprotein than as a tumor suppressor, which is clearly not the case in cervical cancer. Considering all the previous evidence, we can only infer that the proteins pRB and E2F1 are acting with opposing roles in cervical cancer, as compared to colorectal cancer.

On the other hand, there are reports of E2F1 and *c-Myc* overexpression in cervical cancer, which also correlate their expression with advanced states of the disease [59]. Also, genetic studies of the E2F promoters have shown that these genes are induced by *c-Myc*, dependent on the E box sites. Coordinately, all data point to a model where *c-Myc* is activated by means different from the canonical pathways and is working with E2F1 to promote the neoplastic phenotype.

The possible scenario of molecular events in cervical cancer suggested by our study portrays the evasion of apoptosis mediated in two ways, first, the sequestration of the pro-apoptotic protein BAD by 14-3-3 ζ [60] and the overexpression of the anti-apoptotic protein Bcl-2 managed by the interaction of *c-Myc* and E2F1 [61]. Evasion of apoptosis by these routes can be supported by the interaction of viral oncoproteins E6 and E7 with

tumor suppressor proteins p53 and pRb, when there is persistent infection with high-risk HPV. Simultaneously, these oncoproteins are able to promote significant increase in proliferation, and immortalize cells through the activation of hTERT [62].

The expression of E2F factors and cyclins facilitated by *c-Myc*, together E2F1 leads to a quick transition from the G1 to the S phase of the cell cycle, and the consequent boost in growth and cell proliferation [63].

Finally, *c-Myc* overexpression allows a significant increase in cell proliferation by key roads, such as protein biosynthesis, central metabolism, the expression of transcription and cell cycle factors [64] and expression of 14-3-3 ζ . All of these contribute to facilitate the transition from G1 to S phase [65]. *c-Myc*'s mediated expression of Ezrin [66] also promotes epithelial mesenchyma transition, facilitated by the overexpression of vimentin [67] (Figure 5).

Conclusions

Our results suggest a phenotype shared by the six cervical cancer cell lines as a result of the overexpression of *c-Myc*, helped by E2F1, which in turn allows the overexpression of 14-3-3 ζ and other proteins of the "central core of cervical cancer". This signal transduction protein has been reported in other models of cancer as being responsible for malignant transformation and the decision between life and death of cells (Figure 5).

Methods

Cell culture

The CaSki, HeLa, SiHa, C-33A, ViBo and CaLo cell lines were provided by the oncology laboratory of the Centro Medico Siglo XXI which belongs to the Instituto Mexicano del Seguro Social. The HaCaT cell line was donated by the Centro de Investigación Sobre Enfermedades Infecciosas, which belongs to the Instituto Nacional de Salud Pública. All cell lines were cultured in RPMI-advanced 1640 serum-free media (Gibco BRL, USA) with red phenol and antibiotic-antimycotic solution (10,000 units penicillin, 10 mg streptomycin, and 25 μ g amphotericin B per mL), supplemented with 1% fetal bovine serum (Invitrogen, Carlsbad, CA) and 200 mM of GlutaMAX (Invitrogen). The cells were incubated in 5% of CO₂ and humidity saturation at 37°C in culture flasks of 75 cm² (Nalge Nunc International, Rochester, NY). Cells were harvested at 70% confluence with Verseno solution (Tris base 25 mM, NaCl 136.8 mM, KCl 5.36 mM, EDTA 1 mM pH7.7) and washed 3 times in phosphate buffer saline (0.1 M sodium phosphate and 0.15 M NaCl in one liter, pH 7.2).

Proteomic Analysis

Protein extraction and two dimensional gel electrophoresis were done as previously described [68]. Gels were

died in colloidal coomassie [69] and scanned in a GS-800 densitometer (Bio-Rad, Hercules, CA). Digital images were analyzed and compared using the PDQuest 8.0.1 software (Bio-Rad). Each experiment was done in triplicate. Once the digital image of each gel was compared against the rest, the electrophoretic entities of interest were cut, alkylated, reduced, digested and automatically transferred to a MALDI analysis target by a Proteomeer SP II and SP robot using the SPcontrol 3.1.48.0 v software (Bruker Daltonics, Bremen, Germany), with the aid of a DP Chemicals 96 gel digestion kit (Bruker Daltonics) and processed in a MALDI-TOF Autoflex (Bruker Daltonics) to obtain a mass fingerprint. We performed 100 satisfactory shots in 20 shotsteps, the peak resolution threshold was set at 1,500, the signal/noise ratio of tolerance was 6, and contaminants were not excluded. The spectrum was annotated by the flexAnalysis 1.2 v SD1 Patch 2 (Bruker Daltonics). The search engine MASCOT [70] was used to compare the fingerprints against the UNIPROT [71] release 2010_09 database with the following parameters: Taxon-Human, mass tolerance of up to 500 ppm, one miss-cleavage allowed, and as the fixed modification Carbamidomethyl and oxidation of methionine as the variable modification.

Western blot analysis

Antibodies for immunoblotting were as follows: anti-14-3-3 ζ human polyclonal (Imgenex, San Diego, CA) and anti- α -tubulin mouse monoclonal (Zymed Laboratories, South San Francisco, CA). Equal amounts of protein samples were subjected to SDS-PAGE, and transferred to a polyvinylidene difluoride membrane (Immobilon-P, Millipore Corp., Billerica, MA, USA). After blocking with 5% skimmed milk, the membrane was washed in TBS-tween 20%, and incubated with a primary antibody, followed by a AP-conjugated second antibody anti-rabbit or anti-mouse (Zymed). Blots were detected by chromogenic substrate BCIP/NBT (Zymed).

Network Reconstruction

The network reconstruction was performed with the aid of the Cytoscape [72] Plugin, BisoGenet [73], using the identified proteins as bait nodes and adding edges with the following parameters: Organism> Homo sapiens, protein identifiers only; Data Settings>protein-protein interactions; all data sources and all experimental methods; method> By adding edges connecting input nodes and as Output> Proteins.

Network Extension

The primary network was extended by performing the bioinformatic bait technique, using the Cytoscape Plugin BisoGenet with the expand network option and the previous parameters, except that the method used was By

adding neighbors of input nodes to a distance of one. The resulting network was subjected to a centrality measure analysis with the Plugin CentiScaPe1.1 [74].

Pathway and GO enrichment analysis

We performed an enrichment analysis of pathway-based sets of proteins considering all the nodes of our extended network. Enrichment was done employing ConsensusPathDB [75], of the Max Planck Institute for Molecular Genetics, by using the overrepresentation analysis online tool. As input, we uploaded the UNIPROT protein identifiers of all the elements of the extended network. We searched against pathways as defined by KEGG [76], with a minimal overlap with the input list of 5 and a p-value cutoff of 0.0001.

Also, employing the same website and the same analysis tool, we performed an enrichment analysis based on Gene Ontology [77] level 3 category of biological processes. For this analysis, we considered only the identified core proteins and set the p-value cutoff on 0.00001.

Promoter Analysis

To help us uncover some of the regulatory dynamics underlying cervical cancer we downloaded the Gencode Genes-ENCODE Gencode Gene Annotations table (wgEncodeGencodeManual V3) [78,79] from the University of California Santa Cruz Genome Browser website [80]. From this table we retrieved the transcription start site of all GENCODE genes with a score greater than 750, as well as its associated ENSEMBL Transcript ID and gene name. Because some transcripts present several nearby start sites, we also set a filter that only considers start sites located at a distance greater than 500 bp, this reduced the amount of entries considered to 18,534. Promoter sequences were defined as 700 bp upstream and 300 bp downstream of the transcription start site. The promoter locations corresponding to the genes of the proteins of our network were identified based on the ENSEMBL transcript ID as well as the gene names and synonyms stated in the UNIPROT data of those proteins.

Peak information was downloaded from the Yale transcription factor binding site (TFBS) track of the ENCODE Project [80] with a preference for the assays from the HeLa-S3 cell line. A peak was considered to be inside the promoter if the mid-point of the peak was inside the 1,000 bp reported as the promoter sequence. Hypergeometric tests were performed in order to assess the significance of our findings. Sample size was considered as the number of promoters that contain TFBS. All methodology is summarized in Figure 1.

The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche using the following hash:

ixDTEeYyZfAfXrh8HgM4rHTY0BDvNFtaU/GHBp
QW/ldOSRKlStNTsd5aKw9dla
58 iAmEm8L7H8jAZ7A+TR2SNpTg+EUAAAAAAA
AB8Q==

Abbreviations

ChIP: Chromatin immunoprecipitation; GO: Gene Ontology; HPV: Human Papillomavirus; HR-HPV: High Risk Human Papillomavirus; HFF: Human foreskin fibroblast; HFk: Human foreskin keratinocyte; PCa: Prostate cancer; PPI: Protein-protein interaction; TF: Transcription factor; TFBS: Transcription factor binding site; UPR: Unfolded protein response; OD: Optical Density; EMT: Epithelial Mesenchymal Transition; KEGG: Kyoto Encyclopedia of Genes and Genomes; UniProt: Universal Protein Resource; ENCODE: Encyclopedia of DNA Elements; ConsensusPathDB: ConsensusPath Database; EBI: European Bioinformatics Institute; EMBL: European Molecular Biology Laboratory.

Acknowledgements

We would like to thank Mauricio Salcedo-Vargas to provide the cervical cancer cell lines, Gabriel Martínez-Batallar for his advice in PDQuest handling, Alberto Checa-Rojas for his help with cell culture, and Alberto Ramírez-Torres during the elaboration of 2D SDS-PAGE. Part of this work was supported by CONACyT grant 152167 and DGAPA-PAPIIT grant IN-216260. Juan Carlos Higareda-Almaraz is a recipient of a PhD Studentship from the CONACyT, CVU 176426.

Author details

¹Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México. Apdo, Postal 565-A, Cuernavaca, Morelos, CP 62210, México. ²Licenciatura en Ciencias Genómicas, Universidad Nacional Autónoma de México. Apdo. Postal 565-A, Cuernavaca, Morelos, CP 62210, México.

Authors' contributions

JCHA and MREG carried out the experiments, participated in its design and drafted the manuscript. ORA reviews the draft and was part of the discussion. MHO realized the mass spectrometry analysis and SEG conceived the study, participated in its design and coordination, and proof-read the manuscript. All authors read and approved the final manuscript.

Received: 4 February 2011 Accepted: 22 June 2011
Published: 22 June 2011

References

1. Kreeger PK, Lauffenburger DA: **Cancer systems biology: a network modeling perspective.** *Carcinogenesis* 2010, **31**:2-8.
2. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100**:57-70.
3. Hanahan D, Weinberg RA: **Hallmarks of cancer: The Next Generation.** *Cell* 2011, **144**:646-674.
4. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJ, Meijer CJ: **Epidemiologic classification of human papillomavirus types associated with cervical cancer.** *N Engl J Med* 2003, **348**:518-527.
5. von Knebel Doeberitz M, Rittmüller C, zur Hausen H, Dürst M: **Inhibition of tumorigenicity of cervical cancer cells in nude mice by HPV E6-E7 anti-sense RNA.** *Int J Cancer* 1992, **51**:831-834.
6. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM: **The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53.** *Cell* 1990, **63**:1129-1136.
7. Boyer SN, Wazer DE, Band V: **E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway.** *Cancer Res* 1996, **56**:4620-4624.
8. Bernard HU, Burk RD, Chen Z, van Doorslaer K, Hausen H, de Villiers EM: **Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments.** *Virology* 2010, **401**:70-79.
9. Masters JR: **HeLa cells 50 years on: the good, the bad and the ugly.** *Nat Rev Cancer* 2002, **2**:315-319.
10. Scarinci IC, Garcia FA, Kobetz E, Partridge EE, Brandt HM, Bell MC, Dignan M, Ma GX, Daye JL, Castle PE: **Cervical cancer prevention: new tools and old barriers.** *Cancer* 2010, **116**:2531-2542.
11. Molano M, Van Den Brule A, Plummer M, Weiderpass E, Posso H, Arslan A, Meijer CJ, Muñoz N, Franceschi S: **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**:486-494.
12. Perez-Plasencia C, Dueñas-Gonzalez A, Alatorre-Tavera B: **Second hit in cervical carcinogenesis process: involvement of wnt/beta catenin pathway.** *Int Arch Med* 2008, **1**:10-18.
13. Jonsson PF, Bates PA: **Global topological features of cancer proteins in the human interactome.** *Bioinformatics* 2006, **22**:2291-2297.
14. Choi YP, Kang S, Hong S, Xie X, Cho NH: **Proteomic analysis of progressive factors in uterine cervical cancer.** *Proteomics* 2005, **5**:1481-1493.
15. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R: **The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology.** *Nucl Acids Res* 2004, **32**:D262-D266.
16. Shiozawa Y, Havens AM, Jung Y, Ziegler AM, Pedersen EA, Wang J, Wang J, Lu G, Roodman GD, Loberg RD, Pienta KJ, Taichman RS: **Annexin II/annexin II receptor axis regulates adhesion, migration, homing, and growth of prostate cancer.** *J Cell Biochem* 2008, **105**:370-380.
17. Goplen D, Wang J, Enger PØ, Tysnes BB, Terzis AJ, Laerum OD, Bjerkvig R: **Protein disulfide isomerase expression is related to the invasive properties of malignant glioma.** *Cancer Res* 2006, **66**:9895-9902.
18. Wei J, Xu G, Wu M, Zhang Y, Li Q, Liu P, Zhu T, Song A, Zhao L, Han Z, Chen G, Wang S, Meng L, Zhou J, Lu Y, Wang S, Ma D: **Overexpression of vimentin contributes to prostate cancer invasion and metastasis via src regulation.** *Anticancer Res* 2008, **28**:327-334.
19. Elliott BE, Meens JA, SenGupta SK, Louvard D, Arpin M: **The membrane cytoskeletal crosslinker ezrin is required for metastasis of breast carcinoma cells.** *Breast Cancer Res* 2005, **7**:R365-R373.
20. Mierke CT, Kollmannsberger P, Zitterbart DP, Diez G, Koch TM, Marg S, Ziegler WH, Goldmann WH, Fabry B: **Vinculin facilitates cell invasion into three-dimensional collagen matrices.** *J Biol Chem* 2010, **285**:13121-13130.
21. Dong D, Ni M, Li J, Xiong S, Ye W, Virrey JJ, Mao C, Ye R, Wang M, Pen L, Dubeau L, Groshen S, Hofman FM, Lee AS: **Critical role of the stress chaperone GRP78/BiP in tumor proliferation, survival, and tumor angiogenesis in transgene-induced mammary tumor development.** *Cancer Res* 2008, **68**:498-505.
22. Reddy RK, Mao C, Baumeister P, Austin RC, Kaufman RJ, Lee AS: **Endoplasmic reticulum chaperone protein GRP78 protects cells from apoptosis induced by topoisomerase inhibitors: role of ATP binding site in suppression of caspase-7 activation.** *J Biol Chem* 2003, **278**:20915-20924.
23. Warburg O: **On the Origin of Cancer Cells.** *Science* 1956, **123**:309-314.
24. Ferguson EC, Rathmell JC: **New roles for pyruvate kinase M2: working out the Warburg effect.** *Trends Biochem Sci* 2008, **33**:359-362.
25. Resendis-Antonio O, Checa A, Encarnación S: **Modeling core metabolism in cancer cells: surveying the topology underlying the warburg effect.** *PLoS one* 2010, **5**:e12383.
26. Vander Heiden MG, Cantley LC, Thompson CB: **Understanding the Warburg effect: the metabolic requirements of cell proliferation.** *Science* 2009, **324**:1029-1033.
27. Thijssen VL, Barkan B, Shoji H, Aries IM, Mathieu V, Deltour L, Hackeng TM, Kiss R, Kloog Y, Poirier F, Griffioen AW: **Tumor cells secrete galectin-1 to enhance endothelial cell activity.** *Cancer Res* 2010, **70**:6216-6224.
28. Jung EJ, Moon HG, Cho BI, Jeong CY, Joo YT, Lee YJ, Hong SC, Choi SK, Ha WS, Kim JW, Lee JS, Park ST: **Galectin-1 expression in cancer-associated stromal cells correlates tumor invasiveness and tumor progression in breast cancer.** *Int J Cancer* 2007, **120**:2331-2338.
29. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Liefink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roehert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H: **IntAct—open source resource for molecular interaction data.** *Nucl Acids Res* 2007, **35**:D561-D565.
30. Stark C, Breitkreutz B, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets.** *Nucl Acids Res* 2006, **34**:D535-D539.
31. Ceol A, Aryamontri CA, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: **MINT, the molecular interaction database: 2009 update.** *Nucl Acids Res* 2010, **38**:D532-D539.

32. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim S, Eisenberg D: **DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions.** *Nucl Acids Res* 2002, **30**:303-305.
33. Bader GD, Betel D, Hogue CW: **BIND: the Biomolecular Interaction Network Database.** *Nucl Acids Res* 2003, **31**:248-250.
34. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Chandrika KN, Deshpande N, Suresh S, Rashmi BP, Shanker K, Padma N, Niranjani V, Harsha HC, Talreja N, Vrushabendra BM, Ramya MA, Yathish AJ, Joy M, Shivashankar HN, Kavitha MP, Menezes M, Choudhury DR, Ghosh N, Saravana R, Chandran S, Mohan S, Jonnalagadda CK, Prasad CK, Kumar-Sinha C, Deshpande KS, Pandey A: **Human protein reference database as a discovery resource for proteomics.** *Nucl Acids Res* 2004, **32**:D497-D501.
35. Joy MP, Brock A, Ingber DE, Huang S: **High-betweenness proteins in the yeast protein interaction network.** *J Biomed Biotechnol* 2005, **2**:96-103.
36. Hao P, Zheng S, Ping J, Tu K, Gieger C, Wang-Sattler R, Zhong Y, Li Y: **Human gene expression sensitivity according to large scale meta-analysis.** *BMC Bioinformatics* 2009, **10**(Suppl 1):S56-S63.
37. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S: **Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.** *Nat Methods* 2007, **4**:651-657.
38. Jiang C, Xuan Z, Zhao F, Zhang MQ: **TRED: a transcriptional regulatory element database, new entries and other development.** *Nucl Acids Res* 2007, **35**:D137-D140.
39. Nilsson JA, Cleveland JL: **Myc pathways provoking cell suicide and cancer.** *Oncogene* 2003, **22**:9007-9021.
40. Scheffner M, Münger K, Byrne JC, Howley PM: **The State of the p53 and Retinoblastoma Genes in Human Cervical Carcinoma Cell Lines.** *Proc Natl Acad Sci USA* 1991, **88**:5523-5527.
41. Zeller K, Zhao X, Lee CW, Chiu KP, Yao F, Yustein JT, Ooi HS, Orlov YL, Shahab A, Yong HC, Fu Y, Weng Z, Kuznetsov VA, Sung WK, Ruan Y, Dang CV, Wei CL: **Global mapping of c-Myc binding sites and target gene networks in human B cells.** *Proc Natl Acad Sci USA* 2006, **103**:17834-17839.
42. Baudino TA, Maclean KH, Brennan J, Parganas E, Yang C, Aslanian A, Lees JA, Sherr CJ, Rousset MF, Cleveland JL: **Myc-mediated proliferation and lymphomagenesis, but not apoptosis, are compromised by E2f1 loss.** *Mol Cells* 2003, **11**:905-914.
43. Sakamuro D, Prendergast GC: **New Myc-interacting proteins: a second Myc network emerges.** *Oncogene* 1999, **18**:2942-2954.
44. Patel JH, McMahon SB: **Targeting of Miz-1 Is Essential for Myc-mediated Apoptosis.** *J Biol Chem* 2006, **281**:3283-3289.
45. Hong HY, Jeon WK, Bae EJ, Kim ST, Lee HJ, Kim SJ, Kim BC: **14-3-3 sigma and 14-3-3 zeta plays an opposite role in cell growth inhibition mediated by transforming growth factor-beta 1.** *Mol Cells* 2010, **29**:305-309.
46. Tzivion G, Shen YH, Zhu J: **14-3-3 proteins; bringing new definitions to scaffolding.** *Oncogene* 2001, **20**:6331-6338.
47. Morrison DK: **The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development.** *Trends Cell Biol* 2009, **19**:16-23.
48. Zhang P, Chan SL, Fu W, Mendoza M, Mattson MP: **TERT suppresses apoptosis at a premitochondrial step by a mechanism requiring reverse transcriptase activity and 14-3-3 protein-binding ability.** *FASEB J* 2003, **17**:767-779.
49. Avruch J, Khokhlatchev A, Kyriakis JM, Luo Z, Tzivion G, Vavvas D, Zhang XF: **Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade.** *Recent Prog Horm Res* 2001, **56**:127-155.
50. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R: **Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF.** *Cell* 2004, **116**:855-867.
51. Niemantsverdriet M, Wagner K, Visser M, Backendorf C: **Cellular functions of 14-3-3 zeta in apoptosis and cell adhesion emphasize its oncogenic character.** *Oncogene* 2008, **27**:1315-1319.
52. Danes CG, Wyszomierski SL, Lu J, Neal CL, Yang W, Yu D: **14-3-3 Zeta Down-Regulates P53 in Mammary Epithelial Cells and Confers Luminal Filling.** *Cancer Res* 2008, **68**:1760-1767.
53. Sagawa Y, Nishi H, Isaka K, Fujito A, Takayama M: **The correlation of TERT expression with c-myc expression in cervical cancer.** *Cancer Lett* 2001, **168**:45-50.
54. Hwang SG, Lee D, Kim J, Seo T, Choe J: **Human papillomavirus type 16 E7 binds to E2F1 and activates E2F1-driven transcription in a retinoblastoma protein-independent manner.** *J Biol Chem* 2002, **277**:2923-2930.
55. Abramova MV, Zatulovskiy EA, Svetlikova SB, Pospelov VA: **e2f1 Gene is a new member of Wnt/beta-catenin/Tcf-regulated genes.** *Biochem Biophys Res Commun* 2010, **391**:142-146.
56. Morris EJ, Ji JY, Yang F, Di Stefano L, Herr A, Moon NS, Kwon EJ, Haigis KM, Näär AM, Dyson NJ: **E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8.** *Nature* 2008, **455**:552-556.
57. Liu X, Dakic A, Chen R, Disbrow GL, Zhang Y, Dai Y, Schlegel R: **Cell-restricted immortalization by human papillomavirus correlates with telomerase activation and engagement of the hTERT promoter by Myc.** *J Virol* 2008, **82**:11568-11576.
58. Bernards R: **Cancer: Entangled pathways.** *Nature* 2008, **455**:479-480.
59. Arvanitis DA, Spandidos DA: **Deregulation of the G1/S phase transition in cancer and squamous intraepithelial lesions of the uterine cervix: a case control study.** *Oncol Rep* 2008, **20**:751-760.
60. Yang H, Masters SC, Wang H, Fu H: **The proapoptotic protein Bad binds the amphipathic groove of 14-3-3zeta.** *Biochim Biophys Acta* 2001, **1547**:313-319.
61. Li Q, Dashwood WM, Zhong X, Nakagama H, Dashwood RH: **Bcl-2 overexpression in PhIP-induced colon tumors: cloning of the rat Bcl-2 promoter and characterization of a pathway involving b-catenin, c-Myc and E2F1.** *Oncogene* 2007, **26**:6194-6202.
62. Moody CA, Laimins LA: **Human papillomavirus oncoproteins: pathways to transformation.** *Nat Rev Cancer* 2010, **10**:550-560.
63. Leung JY, Ehmann GL, Giangrande PH, Nevins JR: **A role for Myc in facilitating transcription activation by E2F1.** *Oncogene* 2008, **27**:4172-4179.
64. Eilers M, Eisenman RN: **Myc's broad reach.** *Genes Dev* 2008, **22**:2755-2766.
65. Hong HY, Jeon WK, Bae EJ, Kim ST, Lee HJ, Kim SJ, Kim BC: **14-3-3 sigma and 14-3-3 zeta plays an opposite role in cell growth inhibition mediated by transforming growth factor-beta 1.** *Mol Cells* 2010, **29**:305-309.
66. Chuan YC, Iglesias-Gato D, Fernandez-Perez L, Cedazo-Minguez A, Pang ST, Norstedt G, Pousette A, Flores-Morales A: **Ezrin mediates c-Myc actions in prostate cancer cell invasion.** *Oncogene* 2010, **29**:1531-1542.
67. Lu J, Guo H, Treekitarmongkol W, Li P, Zhang J, Shi B, Ling C, Zhou X, Chen T, Chiao PJ, Feng X, Seewaldt VL, Muller WJ, Sahin A, Hung MC, Yu D: **14-3-3zeta Cooperates with ErbB2 to promote ductal carcinoma in situ progression to invasive breast cancer by inducing epithelial-mesenchymal transition.** *Cancer Cell* 2009, **16**:195-207.
68. Salazar E, Diaz-Mejia JJ, Moreno-Hagelsieb G, Martinez-Batallar G, Mora Y, Mora J, Encarnación S: **Characterization of the NifA-RpoN Regulon in *Rhizobium etli* in Free Life and in Symbiosis with *Phaseolus vulgaris*.** *Appl Environ Microbiol* 2010, **76**:4510-4520.
69. Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L, Righetti PG: **Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis.** *Electrophoresis* 2004, **25**:1327-1333.
70. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *Electrophoresis* 1999, **20**:3551-3567.
71. The UniProt Consortium: **The universal protein resource (UniProt).** *Nucl Acids Res* 2008, **36**:D190-D195.
72. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498-2504.
73. Martin A, Ochagavia ME, Rabasa LC, Miranda J, Fernandez-de-Cossio J, Bringas R: **BisoGenet: a new tool for gene network building, visualization and analysis.** *BMC Bioinformatics* 2010, **11**:91-100.
74. Scardoni G, Pitterlini M, Laudanna C: **Analyzing biological network parameters with CentiScaPe.** *Bioinformatics* 2009, **25**:2857-2859.
75. Kamburov A, Wierling C, Lehrach H, Herwig R: **ConsensusPathDB—a database for integrating human functional interaction networks.** *Nucl Acids Res* 2009, **37**:D623-D628.

76. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucl Acids Res* 2000, **28**:27-30.
77. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology.** *Nat Genet* 2000, **25**:25-29.
78. Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Meyer LR, Pohl A, Raney BJ, Wang T, Hinrichs AS, Zweig AS, Fujita PA, Learned K, Rhead B, Smith KE, Kuhn RM, Karolchik D, Haussler D, Kent WJ: **ENCODE whole-genome data in the UCSC Genome Browser.** *Nucl Acids Res* 2010, **38**: D620-D625.
79. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, Learned K, Hsu F, Hillman-Jackson J, Harte RA, Giardine B, Dreszer TR, Clawson H, Barber GP, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2010.** *Nucl Acids Res* 2010, **38**:D613-D619.
80. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The Human Genome Browser at UCSC.** *Genome Res* 2002, **12**:996-1006.

doi:10.1186/1752-0509-5-96

Cite this article as: Higareda-Almaraz *et al.*: Proteomic patterns of cervical cancer cell lines, a network perspective. *BMC Systems Biology* 2011 5:96.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Analysis and Prediction of Pathways in HeLa Cells by Integrating Biological Levels of Organization with Systems-Biology Approaches.

Juan Carlos Higareda-Almaraz¹, Ilse A Valtierra-Gutiérrez^{1,2}, Magdalena Hernández-Ortiz¹, Sandra Contreras¹, Erika Hernández², Sergio Encarnación-Guevara^{1§}

¹ Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México. Apdo, Postal 565-A, Cuernavaca, Morelos, CP 62210, México.

² Licenciatura en Ciencias Genómicas, Universidad Nacional Autónoma de México. Apdo. Postal 565-A, Cuernavaca, Morelos, CP 62210, México.

[§]Corresponding author

Email addresses:

JCHA: ixca@ccg.unam.mx

IAVG: ivaltier@lcg.unam.mx

MHO: magda@ccg.unam.mx

SC: sandra@ccg.unam.mx

EH: erikah@lcg.unam.mx

SEG: encarnac@ccg.unam.mx

Abstract

Cancer has recently begun to be considered as a systemic disease that must be studied by many approaches, including "omics" technologies. To advance biological understanding of the disease and discover comprehensive treatments, the currently available tools must be used to understand the behavior of cancer cells at different levels of complexity. A total expression analysis of the HeLa cell line showed that 19974 genes were transcribed. 3360 of these were over-expressed, and 2129 under-expressed, according to a differential expression analysis of HeLa versus the NHEK cell line. A protein-protein interaction network was recovered from the over-expressed genes in order to identify central elements and, together with the analysis of over-represented transcription factor motifs, to predict active signaling and regulatory pathways. 87 of these were validated by the identification of 271 phosphoproteins by tandem mass spectrometry. In the HeLa cell line 87 signaling and regulatory pathways were over-expressed, and 56 pathways were under-expressed. Therefore, this cellular system is highly robust. The 14-3-3 family members emerge as some of the most important regulators in carcinogenesis in different models and as possible clinical targets.

Certain signaling pathways, transcription networks and metabolic events could be maintained in different types of cancer, regardless of mutations or genomic rearrangements; this would favor the appearance of the hallmarks of cancer and the evasion of tissue control. This could explain why they are not eliminated by selective pressure and why therapy trials directed against molecular targets are not as effective as expected.

Author Summary

Cervical cancer is a major cause of death, ranked just lower than breast cancer, in developing countries. Despite having a well-described etiologic agent, the high risk human papillomavirus, there are still molecular events that occur in transformation and maintenance of the malignant state that are not understood. Because knowledge of their biology is the most efficient way to prevent, predict and treat a disease, it is necessary to have thorough understanding of these events. In this paper, we perform the prediction and validation of the state of various pathways that allow cell communication and govern gene expression and cell metabolism. Our systems-level analysis allowed us to find a highly organized meta-pathway network that governs the malignant state in the HeLa cell line. Furthermore, we found a great robustness in the form of redundancy in the pathways responsible for maintaining malignant cells viable and constantly growing, allowing us to explain the efficiency of replicative capacity and the resilience to targeted therapies and drugs.

Introduction

Understanding the complexity at different levels of biological organization requires combining the results obtained from various experiments to recreate the system's behavior [1, 2]. The molecular profiling methods known as “omics” (e.g., transcriptomics, proteomics, and metabolomics) allow a global search of the characteristics that define the system under study and the integration of this knowledge into simple models with great explanatory and predictive power. However, these models can and must be contrasted against new experimental data. Currently, the cellular signaling system represents the biggest challenge for systems biology [3–5].

A signaling pathway consists of multiple sequential events, including covalent modifications, recruitment, allosteric activation or inhibition and protein binding [6]. However, as our understanding of the interactions between signaling pathways increases, it becomes more apparent that the signals do not necessarily occur independently through parallel linear pathways, but rather, through a large and complex network of interconnected signaling pathways [7, 8].

The complex architecture of signaling networks can be understood as a set of interacting network motifs, which can provide specific network properties and add a new level of complexity to that which already exists within the spatio-temporal organization and compartmentalization of signals [9, 10].

Cells are complex, dynamic systems, which use molecular signaling circuits that govern basic cellular activities and coordinate their actions [11]. The ability of cells to perceive and respond in an appropriate manner to the microenvironment is the basis for homeostasis, development, tissue repair and immunity. Errors in information management are responsible for different cell-derived conditions, such as autoimmune diseases, metabolic syndromes and cancer [12–15].

Cancer requires a very complex set of conditions. It is driven by a Darwinian model of evolution at the cellular level [16], comprising all levels of cellular information (i.e., genetics, epigenetics, transcriptional and translational regulation and translational modifications). Accordingly, it involves communication between different cell types and interactions between the tumoral microenvironment and the whole organism [17, 18]. Our understanding of cancer has evolved because of this context and acquired knowledge.

Hanahan and Weinberg suggested that all cancers have certain essential alterations in cell physiology that coordinate the malignant phenotype, which is characterized by self-sufficiency in growth signals, insensitivity to growth inhibitors, evasion of programmed cell death, increased replicative potential, sustained angiogenesis, tissue invasiveness and metastasis, reprogramming of energy metabolism and evasion of immune destruction. Moreover, these hallmarks are accompanied by additional enabling features, including mutations and genomic instability and the promotion of inflammation by tumors [19, 20].

Cervical cancer represents an interesting opportunity for the study of malignant transformation, mainly due to our understanding of its etiologic agent, High-Risk Human Papilloma Viruses (HR-HPVs), which are found in 90.7% of cases [21]. The HR-HPV oncoproteins E6 and E7 are able to interact with the p53 and pRb tumor suppressors in addition to more than 300 other known proteins. Of the more than 120 types of HPVs that infect humans, only a few high-risk types are associated with carcinogenesis. HPV16 and HPV18 are the most prevalent high-risk HPVs, and they are present in 54.6% and 11% of cervical squamous cell carcinomas, respectively [22–24]. Patients with cancer caused by these HPV types are the most widely studied. The first established cervical carcinoma cell line, HeLa, is positive for HPV18 and has served as the basis for most of our knowledge regarding the underlying cell biology of cancer.

However, due to spontaneous elimination of the virus, not all patients infected with HR-HPV develop cervical cancer. Most HPV infections are subclinical, with only a small fraction producing epithelial lesions, with an even smaller fraction of these lesions developing into cancer [25]. Consequently, HR-HPV infection is necessary but not sufficient for the development of cervical cancer [26]. Thus, the conditions that allow the development of cervical cancer both following HR-HPV infection and in its absence are not thoroughly known.

In a previous study, we proposed that the delicate balance between the life and death decisions of cells, as well as the neoplastic phenotype, might be due to the overregulation of the transcription factors c-Myc and E2F1. This can apparently result from both viral infection and the overexpression of the protein 14-3-3Z. 14-3-3Z has been shown to deregulate apoptosis and promote the G1 to S phase transition. Furthermore, it has been suggested to play an important role in the epithelial-mesenchymal transition (EMT) [27].

The aim of this study was to both predict the behavior of signaling pathways and regulatory networks and determine the molecular signature of cervical cancer in a HeLa cell line model. We used data generated via sequencing, performed a differential expression analysis and incorporated microarray data to predict the response of transcription factors. This information allowed the reconstruction of the signaling, metabolic and transcriptional regulation pathways. Finally, we enriched phosphorylated proteins using Metal-Oxide Affinity Chromatography (MOAC) and identified these proteins using tandem mass spectrometry (MS/MS) to validate and build a model based on these different levels of biological information.

Results

We integrated different layers of information within the dynamics of a biological cell that track the flow of information. Using this procedure, we were able to obtain information regarding the maintenance of the malignant state and the differences between cancerous and normal cells (Fig.1).

There are different gene expression profiles and Gene Ontologies in the HeLa and NHEK cell lines

First, to assure the certainty of the gene expression profile of the HeLa cells, we used the RNA-Seq data set generated by [28]. This total expression analysis yielded a set of 19,974 transcripts that are expressed in HeLa cells. The principal difference between these analyses and those from the Nagaraj group was that we performed quartile normalization, which improved the accuracy of the differential expression calls for low-abundance transcripts (Fig. 2a). With the resulting full expression data, we developed a metric to elucidate the representation of cellular processes by means of Gene Ontology (GO) [29] using the domain of cellular components in level 3 (Fig. 2b). Importantly, the expression of each GO term fulfills at least 97% of the total reported for that term (Table A1). We used this metric to establish the biological quality of our analysis and to build our predictions from these data, as well as to validate the differential expression analysis.

To understand the changes in gene expression in the HeLa cell line compared with a normal cell, we next performed a differential expression analysis via RNA-Seq, using the epithelial keratinocyte cell line NHEK as our expression control. We identified a total of 3,360 over-expressed genes and 2,129 under-expressed genes (Fig. 2c). Using this information, a GO enrichment analysis was built with the web tool ConsensusPathDB [30] using level 3 of the

“Biological Process” domain. Even at this level of resolution, it was clear that the differential expression of genes in HeLa cells strongly favored cell proliferation over tissue organization (Table A2). The categories with a clear over-representation were “Cell cycle”, “Gene expression”, “Metabolism building blocks” and “Cytoskeletal reorganization” (Fig. 2d). The categories that were under-represented included “Tissue development”, “Organs and systems”, “Signaling”, “Cell adhesion”, “Lipid metabolism” and “Programmed cell death” (Fig. 2e).

Based on these results, we propose that there is a strong tendency for HeLa cells to express genes that assist in the evasion of tissue control, which affords a clear adaptive advantage to proliferation without barriers.

The regulatory network of transcription factors that are differentially expressed in HeLa cells controls fundamental processes maintaining the neoplastic state

To understand the transcriptional network that governs gene expression in HeLa cells, we used the Affymetrix microarray data from HeLa cells and normal cervical epithelia generated by the Scotto group [31] and deposited in the GEO database. These data were loaded into the MARA web tool [32], which reports the transcription factors with altered expression in malignant cell lines. The transcriptional targets of E2F, ZNF143, YY1, ELKA, GABP, NRF1, MYB, NFY, HIF1A, TFDP1 and ELF were over-represented, whereas the transcriptional targets of ETS, NFATc, NR1H4, SMAD, TFCP2, HIC1, AR, TBP, SRF and KLF12 were under-represented. Next, we used the database generated by the MARA transcriptional target analysis to establish the network of regulation obtained from the differential expression analysis. Targets were obtained for each TF, and their regulatory networks were reconstructed. It should be noted that c-Myc, hepatocyte nuclear factor 4-alpha, BRCA1, VHL and NEMO were all involved in more than one transcription factor network and were overexpressed.

After obtaining information from the TFs and their targets, we performed a GO enrichment analysis using the web tool ConsensusPathDB web tool and level 3 of the “Biological Process” domain. In the overexpressed-TF networks, we identified 103 GO terms (Table A3), including “Cell proliferation”, “Metabolism of building blocks”, “Cellular organization”, “Angiogenesis”, “Central metabolism” and “Signaling” (Fig. 3a). In the under-expressed-TF networks, we identified 159 GO terms (Table A4), which were largely related to “Tissue homeostasis” or “Miscellaneous”.

With the information generated by the TF network analysis, interaction networks were built using Cytoscape software [33] and the Hubba plug-in [34]. Hubs were defined using the node degree (Fig. 3b) and betweenness centralities (Fig. 3c). Surprisingly, c-Myc, HNF4A, BRCA1, VHL and NEMO were the nodes that had the highest values for these measures. These results suggest that a set of overexpressed genes in HeLa cells has control over a particular regulatory network that is not present in normal cervical epithelium. They also suggest reduced expression of some other regulatory networks. This feature may be an important source of the complexity that allows the strengthening of the cellular system by selection pressure.

Identification of under- or over-represented pathways that govern the neoplastic phenotype in HeLa cells

To understand which pathways and processes are responsible for maintaining the neoplastic phenotype, the complete list of previously identified over- and underexpressed transcripts was transformed into non-redundant UniProt identifiers [35]. A pathway enrichment analysis was then performed on this list using the ConsensusPathDB online tool. From the overexpressed transcripts, 83 over-represented pathways were recovered from the Pathway Interaction Database (PID) (Table A5). Among these pathways, there were many remarkable signaling pathways, including those governed by ATR and Aurora A and B. Some transcriptional regulatory networks were identified as well, including “E2F”, “MYB”, “Targets of c-Myc transcriptional activation” and “Direct p53 effectors”. For the underexpressed transcripts, 15 over-represented pathways were detected via the PID [36] (Table A6). Some important pathways were the “Transcriptional targets of deltaNp63”, “TAp63”, “AP1 family members Fra1 and Fra2”, and “Direct p53 effectors”. Notably, both over- and underexpressed transcripts were found within the “Direct p53 effectors” (Table A7). Importantly, the identification of under-expressed and over-expressed transcripts from the same data set illustrates the power of our analysis, as it enabled us to distinguish, at the genomic level, components of the same network that promoted the neoplastic state either by actual participation or by omission.

Another pathway enrichment analysis was performed with a focus on the Kyoto Encyclopedia of Genes and Genomes (KEGG) [37] for both sets of transcripts; with this approach, we identified 17 pathways for the over-expressed genes (Table 1) and 25 pathways for the under-expressed genes (Table A8). These results offer a perspective on the diverse events that occur in HeLa cells compared with non-malignant cells. Some of the highly represented processes were involved in cell proliferation (e.g., “DNA replication”, “Cell cycle” and “Homologous

recombination”). Yet, other significant pathways included the “Fanconi Anemia pathway”, “Transcriptional misregulation in cancer” and “Small cell lung cancer”. This finding provided direct evidence of the conservation of malignant processes in different cancer types. Surprisingly, within the under-expressed genes, there was an over-representation of lipid metabolism pathways (particularly pathways involved in the metabolism of steroids, linoleic acid, arachidonic acid and the synthesis of unsaturated fatty acids), and pathways involved in cellular adhesion and several pathogen infections.

These data strengthened the results that were obtained from the MARA microarray analysis. However, even though the pathway enrichment analyses were important to our analysis of the biological significance of the changes in gene expression, there are limits to these types of analyses. For example, the same set of transcripts can be considered part of different pathways. Consequently, the results must be carefully evaluated and validated before any definitive interpretations are made. Therefore, similar to the current work, studies that involve different layers of information and validation are strongly needed.

There is a defined pattern in the activation of cellular circuits in HeLa cells that involves interconnections and crosstalk that could be identified based on the differentially expressed genes, regulatory networks and pathways

To visualize the integration of the different layers of biological information that were obtained in the present work, the PID Batch Query tool was used. As an input, we used the list of transcripts obtained in the differential expression analysis, as well as the list of super-active TFs that was obtained from the MARA analysis. The PID batch query yielded 64 pathways curated in the Biopax level 3 format [38], which displayed an increased amount of interconnections and crosstalk among their cellular circuits. Such interconnection and crosstalk allows constant proliferation, immortalization and cell migration by means of both the over- and under-expression of different genes and a pattern of transcriptional control that differed from that observed in normal cells.

To explore this hypothesis, we performed a reconstruction of every one of the aforementioned networks using the Bisogenet plug-in [39] in Cytoscape. Afterwards, we searched for the nodes that served as interconnectors between the explored pathways, and with the resulting data, we reconstructed two models. In the first model, we built a network by using every component of each pathway (Fig. 4). Node-degree and betweenness centralities were analyzed, and with these values, we could determine the presence of hubs in all the pathways.

This led to the observation that c-Myc, BRCA1, VEGFA and E2F1 were the most interconnected and influential nodes in all of the networks.

In our second model, we followed the hypothesis that there are well-defined cellular circuits that can be extrapolated to other malignant cells. Such circuits can be adapted to the concept of "Hallmarks of Cancer" that was originally proposed by Hanahan and Weinberg, which we reshaped in what we called "meta-pathways" (Fig. 5). Each one of the pathways that appear in the figure possesses a certain degree of connection with other pathways that was determined either by transcriptional regulation, signaling, metabolism or a combination of these mechanisms. This meta-network would be capable of maintaining the neoplastic phenotype. In Figure 6, each of the pathways and meta-pathways is related to its possible role in the generation of the hallmarks of cancer.

To validate the data that were obtained from the transcripts and TFs, we conducted a phosphoproteomic analysis of the HeLa cell line via enrichment with Metal-Oxide Affinity Chromatography followed by identification of the proteins via LC/MS-MS. We identified a total of 271 phosphorylated proteins (Table A9, reported in 40 level 3 GO terms (Table A10), 21 PID pathways (Table A11) and 16 KEGG pathways (Table A12). As expected, due to the well-documented low correlation between transcript expression and protein expression, only 17% of the proteins had their equivalence as a transcript. Nonetheless, we found there was a high correlation between the pathways based on transcript/TF expression and those based on phosphorylated proteins. Significantly, among these groups were "Validated targets of c-Myc transcriptional activation", "Signaling events mediated by HDAC Class III", "LKB1 signaling events", "Class I P3K signaling events" and "FoxO family signaling". Other important pathways included those from the identified proteins, but were actually a consequence of pathways derived from the transcripts. For example, the "c-Myb transcription factor network" and the "E2F transcription factor network", whose targets were found within "c-Myc transcriptional activation" and, simultaneously, in pathways of central metabolism.

Finally, the entire data set that was extracted from the three layers of information was used to reconstruct the signaling, regulatory and metabolic networks that govern the HeLa cell line and are different from the networks present in both normal cervical epithelium and the NHEK keratinocyte cell line. For this purpose, the KEGG Mapper tool [40] was used to assemble the general metabolic map (Fig. 6), the cell cycle map (Fig. 7a) and the adhesion molecules and focal adhesion maps (Fig. 7b). Based on these maps, we were able to identify a clear pattern that could be divided into at least three groups with the following characteristics:

1. An increase in cell proliferation, resulting from the overexpression of the MCM-complex genes and diverse cell cycle-associated proteins, which permitted sustained proliferative signaling.
2. Over-activation of carbohydrate metabolism (e.g., glycolysis and gluconeogenesis), as well as an increase in the expression of enzymes in the pentose-phosphate pathway, which generates carbon skeletons for the synthesis of nitrogenous bases and histidine. In addition, malonyl-CoA can be obtained from pyruvate metabolism, which matches the “deregulation of cellular energetics” hallmark of cancer.
3. Activation of invasion and metastasis that was caused by the loss/gain of expression of several cell adhesion proteins, such as CLDN, OCLN and ESAM, the expression of which was increased.

Discussion

In the present work, we performed an analysis of the different layers of biological information available from HeLa cells. The purpose of this analysis was to build a biological model of the pathways and meta-pathways that would allow us to define the relationships between this model and the hallmarks of cancer while maintaining a systemic perspective.

Hanahan and Weinberg postulated the existence of an intricate cellular circuitry that can be individually aligned with the hallmarks of cancer. They also proposed that in the near future, every circuit might be segmented into specialized sub-circuits that support discrete biological properties in normal cells and are reprogrammed in cancer cells, resulting in their hallmark capabilities. The work presented here is an attempt to prove, both theoretically and experimentally, the existence of these sub-circuits and their relationships to the specific pathways that afford specific cellular functions. Previous studies have attempted to integrate different levels of information for the prediction of signaling pathways that are either a determinant of or an obstacle to the malignant state.

Minn et al. [41] designed a novel methodology that permits the definition of new signaling pathways. This methodology is based on the initial identification of a gene that works as a master regulator. Next, the available information is increased via a gene set analysis (GSEA). Finally, the regulatory interactions are functionally validated via the over-expression of some

genes that participate in the proposed pathway. This strategy can be coupled to a systemic analysis, such as our own, to validate novel or non-canonical pathways.

The validation step is critical because it has to be thoroughly representative of the complete pathways network, and the most frequently used technique for validation is the direct measurement of mRNA or proteins [42]. Therefore, we decided to validate our initial predictions using proteomics, with a special emphasis on identifying phosphoproteins. The phosphoproteomic validation was pursued because reversible phosphorylation of proteins is the most frequently used post-translational modification in mammalian cells. Approximately 1.7% of the human genome encodes protein kinases [43], and at any moment, approximately 30% of all proteins can be phosphorylated [44]. Phosphorylation suggests that the protein is in its active form (as is the case with several metabolic enzymes) and that upstream signals are active. Alternatively, phosphorylation could result in a conformational change that allows the modulation of specific activities [45].

The analysis of total transcripts in the HeLa cell line yielded a notable result. A total of 19,974 genes were found to be transcribed, and while this may appear to be an exaggerated number, it has been demonstrated via RNA-Seq analysis that approximately 16,245 genes are transcribed in several types of breast cancer, and this number can vary from 14,648 to 18,290 [46]. When we conducted the GO analysis, we observed that the transcripts covered at least 97% of the GO terms. This aided in the evaluation of the quality of the information we obtained and prepared us to make the comparisons with the differential expression data.

When analyzing the differential-expression data in a classical manner, we observed over-expression of the typical oncogenic proteins that participate in diverse types of cancer. These proteins make up three groups: 1) the c-Myc and c-Myb transcription factors, which are expressed in a great variety of tumors [47–50], 2) the DNA repair and recombination proteins BRCA1 and 2, which are amplified in breast cancer and non-small-cell lung cancer [51, 52], and 3) the proteins assigned to mitotic checkpoints, such as BUB1 and BUB3, which can result in genomic instability when overexpressed [53, 54].

Within the global perspective of this work, we found 3,360 overexpressed genes and 2,129 underexpressed genes. The GOs and pathways that resulted from the ConsensusPathDB analysis showed a clear over-expression of genes that are coadjuvant in maintaining the sustained proliferative signaling cancer hallmark at the transcriptional, signaling or metabolic levels.

Similarly, when we analyzed the data regarding all the hyperactive transcription factors, we clearly found that the E2F, c-Myc and c-Myb pathways were positively regulating a large number of genes that initiate mitosis and allow cell cycle progression [55–57]. The over-expression of the FOXM1 transcription factor network has been associated with increased cell proliferation in animal models of prostate carcinoma [58]. Through the integration and validation of these data via the phosphoproteomic analysis, we found that over 70% of the components of the “Cell Cycle” KEGG pathway (Fig. 7a) and all of the components of the Minichromosome Maintenance Complex, whose increased levels have been observed in two other cancer models [59, 60], were over-expressed.

Regarding the deregulation of cellular energetics, we observed a clear increase in glycolysis and pyruvate/lactate metabolism, which is expelled from the cell, resulting in the acidification and remodeling of the extracellular matrix [61, 62]. However, the energetic inefficiency of using glycolysis to obtain ATP is compensated by the synthesis of substrates to create cellular building blocks (i.e., lipids, nitrogenous bases and peptides) from carbon skeletons that are obtained from glucose [63, 64] (Fig 6).

One of the key enzymes in cancer is Pyruvate Kinase M2 because it possesses an extra tyrosine that is phosphorylated, which allows its detection in the phosphoproteomic analysis. This phosphorylation inhibits the positive regulation resulting from the fructose-1,6-bisphosphate level, stimulating the pathway and leaving a large amount of phosphorylated intermediates that can be used for anabolic synthesis and cell growth [65]. The huge quantity of glucose that is required by the cell to obtain energy is facilitated by the over-expression of the genes that encode the membrane glucose transporters GTR3, GTR4, GTR8 and GTR14 [66]. Glucose remains inside the cell because it is phosphorylated and converted into glucose-6-phosphate by HXK2, which is also over-expressed in HeLa cells. During this step, glucose-6-phosphate is shunted into the pentose-phosphate pathway and is used for nitrogenous base metabolism, and we found that the enzymes responsible for turning glucose-6-phosphate into riboses and deoxyriboses [67] (i.e., G6PI, K6PF, DEOC, RBSK, KPRA, KPRB and PRPS1) were all over-expressed. At the transcriptional level, most of metabolic enzymes are regulated by the action of the c-Myc and HIF-1-alpha transcriptional regulatory networks [68], which were over-expressed at the transcriptional, TF and phosphoprotein levels.

The “Activation, Invasion and Metastasis” hallmark of cancer is profoundly complex because it reflects the differential expression of diverse adhesion molecules, such as the Claudins. These molecules are involved in a highly complex interplay and have been reported to be both over- and under-expressed in other malignancies (Fig. 7b) [69]. One indication of metastatic

potential is the appearance of EMT markers. In the present study, we detected the over-expression of OCLN, VIME and 14-3-3Z [70, 71]. As for transcriptional regulation, the Endotelins pathway was over-regulated due to the expression of EDN2 and EDNRA, which resulted in the crossroads between migration and cell proliferation. Finally, the “Signaling events mediated by VEGF” pathway was over-activated because of the over-expression of VEGFA and VEGFB. This pathway permits angiogenesis, vasculogenesis and endothelial cell growth. Furthermore, it represents a crosstalk among various cancer hallmarks because it induces endothelial cell proliferation, promotes cell migration, inhibits apoptosis and induces permeabilization of blood vessels [72].

At the protein level, the family of 14-3-3 signal transducers plays an important role, given that their differential expression affects cell proliferation, evasion of apoptosis, cell adhesion, mitogenic signaling and the EMT [73–76]. The 14-3-3S isoform was found to be under-expressed in the differential expression analysis. This decreased expression supports most models that suggest that this protein functions as a tumor suppressor, and its loss or reduced expression is strongly correlated with a poor prognosis [77]. It is induced by DNA damage and is required for stable G2 arrest. It is directly regulated by p53 and has been found to be silenced or diminished in the majority proportion of carcinomas. The inactivation of 14-3-3S also leads to the immortalization of primary keratinocytes [78].

At the phosphoprotein level we identified the 14-3-3B, 14-3-3E, 14-3-3F, 14-3-3G, 14-3-3S and 14-3-3Z isoforms. The 14-3-3Z isoform antagonizes the 14-3-3S isoform; the amount of the latter diminishes because of the decrease in the expression of p53 in breast cancer, and its overexpression has been correlated with the EMT, metastasis and cell proliferation. Thus, it has been proposed that this family of signal-transducing proteins is critical for the malignant phenotype [79, 80].

After conducting the integrative analysis, we discovered that a large number of spliceosome genes were over-expressed. Constituents of the U2 component, including SRSF1, which is considered oncogenic by itself, were particularly over-expressed. It has been suggested that any change in the stoichiometry or activity of splicing factors is capable of modifying the proportions of isoforms that normally do not exist or are less abundant in normal cells (Fig. 8). This phenomenon could contribute directly or indirectly to the development, progression and maintenance of cancer. Another hypothesis proposes that diverse RNA-binding proteins possess a wide array of functions, and changes in their expression could trigger oncogenic effects that are unrelated to their original role within the spliceosome [2, 81].

Integrating the evidence we obtained at the systems level, we have created a model to show the relationship between the hallmarks of cancer and the signaling, regulatory and metabolic pathways that are differentially expressed in HeLa cells. Figure 6 illustrates what we have termed “meta-pathways.” In these meta-pathways, the reconstructed pathways show their interrelations at the gene regulation, signaling and/or metabolic levels. Each node represents a pathway and is colored according to the two most representative hallmarks. One of the first features that stood out is the large number of signals that sustain the proliferative state of HeLa cells, as well as the great redundancy that can be appreciated among the different hallmarks. These data suggest that this system is extremely robust, as every hallmark is represented more than once and is supported by different pathways. Such behavior would provide an explanation for why directed therapies have not had the expected level of success. When one gene or subset of genes collapses, a bottleneck would immediately develop. This might result in the positive selection of a phenotype that could maintain the hallmarks of cancer via other pathways [82, 83].

Using the data generated in the present work, we propose that systematic robustness is not entirely random. Enough evidence was found to suggest that there are defined patterns in the activation of cellular circuits that involve interconnections and interferences associated to the hallmarks of cancer, the latter being accurately represented by the meta-pathways we have proposed. These observations allow us to hypothesize that, when a cell expresses abnormal levels of key proteins (e.g., TFs, signal transducers, metabolic enzymes, splicing factors or whole pathways) that should normally be repressed, it is highly likely that, if any of these expression patterns are advantageous for the cell, such patterns will become fixed by selection pressure. These effects will allow the cell to possess a response arsenal that can become more sophisticated with further exposure to selective pressures over time.

Methods

Gene and transcript quantification

Data was obtained and processed as described from [28]. The processed reads were aligned to the human reference genome (hg19/GRCh37 excluding additional haplotypes) using TopHat v1.4.1 [84] and transcripts and genes of the Ensembl [85] release 59 were quantified using Cufflinks v0.8.3 [86].

Differential expression Analysis from RNA-seq data

Paired-end RNA-seq data of HeLa-S3 and NHEK cell-lines was downloaded from the ENCYclopedia Of DNA Elements (ENCODE) [87] project's webpage (UCSC accession numbers wgEncodeEH000130 and wgEncodeEH000131 respectively). Fastq files from two HeLa-S3 75x75 paired-end RNA-seq libraries (experiment numbers 10881, 10882) and two NHEK libraries (experiment numbers 10884, 11586) were aligned to the hg19 version of the human reference genome using TopHat v1.4.1 and Gencode annotation Version 12. Default parameters were used and only the read length was modified to 75 bases. The human genome index was built using bowtie v0.12.7 [88]. Differentially expressed genes between NHEK and HeLa-S3 were identified using cuffdiff v 1.3.0 from the Cufflinks package. A P-value threshold of 0.01 was set for all significant differentially expressed genes.

Transcriptional Factor Analysis

Affymetrix microarray dataset HG-U133A of HeLa and normal epithelium was downloaded from the Gen Expression Omnibus (GEO) [89] with accession numbers GSM246123 and GSM246422 respectively. The .cel files were uploaded to the website of MARA. This algorithm normalizes them together; assigns PolII promoters and binding sites within them to probe sets present on the microarray, and runs the TF activity analysis.

Association of transcriptional targets with their RNA expression level

The information from the genes that were reported as targets of each differentially activated TF, as reported by MARA, was associated with their particular levels of expression according to the RNA-Seq analysis. Individual MARA reports for each gene were parsed through ad-hoc scripts written in Perl 5.12.4. This processed information, as well as the data from the RNA-Seq results, were poured into a relational database built in MySQL Server 5.5 (Community Edition). Subsequently, a query for each relevant TF was conducted, relating the targets that corresponded to every TF with RNA-Seq data by joining the corresponding UniProt identifiers. The outputs were stored in plain text files and analyzed with spreadsheet software.

Network Reconstruction

The network reconstruction was performed with the aid of the Cytoscape plug-in BisoGenet, using the identified proteins as bait nodes and adding edges with the following parameters: Organism> Homo sapiens, protein identifiers only; Data Settings>protein-protein interactions; all data sources and all experimental methods; method> By adding edges connecting input nodes and as Output> Proteins.

Pathway and GO enrichment analysis

Enrichment was done employing ConsensusPathDB, of the Max Planck Institute for Molecular Genetics, by using the overrepresentation analysis online tool. As input, we uploaded the UNIPROT protein identifiers of all the elements of the total expression analysis, differential expression analysis, TF networks and identified phosphoproteins. We searched against pathways as defined by PID and KEGG, with a minimal overlap with the input list of 5 and a P-value cutoff of 0.0001. Also, employing the same website and the same analysis tool, we performed an enrichment analysis based on Gene Ontology level 3 categories of “Biological processes”. For this analysis, we considered only the identified core proteins and set the p-value cutoff on 0.00001.

Cell Culture

The HeLa cell line was provided by the oncology laboratory of the Centro Medico Siglo XXI which belongs to the Instituto Mexicano del Seguro Social. HeLa cell line was cultured in RPMI-advanced 1640 serum-free media (Gibco BRL, USA) with red phenol and antibiotic-antimycotic solution (10,000 units penicillin, 10 mg streptomycin, and 25 µg amphotericin B per mL), supplemented with 1% fetal bovine serum (Invitrogen, Carlsbad, CA) and 200 mM of GlutaMAX (Invitrogen). The cells were incubated in 5% of CO₂ and humidity saturation at 37°C in culture flasks of 75 cm² (Nalge Nunc International, Rochester, NY). Cells were harvested at 70% confluence with Verseno solution (Tris base 25 mM, NaCl 136.8 mM, KCl 5.36 mM, EDTA 1 mM pH7.7) and washed 3 times in phosphate buffer saline (0.1 M sodium phosphate and 0.15 M NaCl in one liter, pH 7.2).

Enrichment of Phosphoproteins

Protein phenol extraction was performed [90]. The protein extraction was resuspended in incubation buffer and then the metal-oxide affinity chromatography (MOAC) [91]. The phospho-protein fraction was eluted and precipitated [92], and then a Sodium dodecyl sulfate-polyacrylamide Gel Electrophoresis (SDS-PAGE) was conducted. Gels were stained with Coomassie Blue G-250. All bands were excised with a razor blade and the tryptic digestion was performed.

Chromatographic Separation

The tryptic peptides (8 µl) were desalted and concentrated on a Zorbax (Agilent 5065-9913) prior to analysis on a reverse-phase column (Agilent Zorbax 300SB C18, 3.5 µm, 150X0.075 mm). Separation was performed at 400 nL/min using a linear gradient. Mobile phase A was water with 0.1% formic acid by volume. Mobile phase B was acetonitrile with 0.1% formic acid by volume. The gradient conditions in the chromatographic run were set up as follow: A 95% (0 min) to 95% (14min); A 95%(14min) to 60% (54 min); A 60% (54 min) to 20% (56

min); A 20% (56min) to 20% (61 min); A 20% (61min) to 95% (62 min); and A 95% (62min) to 95% (72 min)

MS/MS Analysis

Proteins were analyzed by MS/MS using a nanoflow chromatograph (Agilent 1100 nano pump G2226A) coupled to a hybrid triple quadrupole linear ion trap (QTRAP 3200, AB Sciex) equipped with a Nanospray II source and using Information Dependent Acquisition (IDA). Precursor ion determination was carried out using an Enhanced MS scan over a mass range of 300-1600 m/z at 4,000 amu/s (with not trapping in Q0 and Dynamic fill time) with an ion spray voltage of 3300 applied to a Picotip FS360-75-15-N with ion spray gas (nitrogen). Precursor ions were collided in Q2 using rolling collision energy (maximum allowed CE=80). Enhanced product ion scans (MS/MS) were performed over a mass range of 100-1700 m/z at 1000 amu/sec, and collision voltages were determined dynamically. All precursor ion mass/charge ratios were confirmed with an Enhance Resolution scan. Protein identification was done by using the Mascot algorithm (<http://www.matrixscience.com>), with SwissProt database; the search parameters included trypsin digestion, MS/MS ion search, monoisotopic mass application, protein mass unrestricted, peptide mass tolerance of ± 1.2 Da, fragment mass tolerance of ± 0.6 Da and Max Missed Cleavages of 1.

The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche using the following hash: hAB36ZvUMqsCgDBALN0mJQ4dts+h4YiAIk5ZSasBjaG2TKhIztzBenjpxYlZaoeq41YheUt9ahhLnC2iPCGKy0SDsGwAAAAAAAAAntw==

Acknowledgements

We would like to thank Martin Del Castillo Velasco Herrera for providing the data of the differential expression analysis. We are grateful to the National Autonomous University of México, and to the Biomedical Sciences PhD Program of this university.

References

1. Mazzocchi F (2008) Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. *EMBO reports*, 9:10–4.
2. Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle WF (2010) Cell biology. Irremediable complexity? *Science*, 330:920–1.
3. Kitano H (2002) Computational systems biology. *Nature*, 420:206–10.
4. Kitano H (2002) Systems biology: a brief overview. *Science*, 295:1662–4.

5. Auffray C, Chen Z, Hood L (2009) Systems medicine: the future of medical genomics and healthcare. *Genome medicine*, 1:2.
6. Zubarev RA, Nielsen ML, Fung EM, Savitski MM, Kel-Margoulis O, et al. (2008) Identification of dominant signaling pathways from proteomics expression data. *Journal of proteomics*, 71:89–96.
7. Papin J a, Hunter T, Palsson BO, Subramaniam S (2005) Reconstruction of cellular signaling networks and analysis of their properties. *Nature reviews. Molecular cell biology*, 6:99–111.
8. Hyduke DR, Palsson BØ: (2010) Towards genome-scale signaling network reconstructions. *Nature reviews. Genetics*, 11:297–307.
9. Takahashi K, Arjunan SNV, Tomita M (2005) Space in systems biology of signaling pathways--towards intracellular molecular crowding in silico. *FEBS letters*, 579:1783–8.
10. Bauch A, Superti-Furga G (2006) Charting protein complexes, signaling pathways, and networks in the immune system. *Immunological reviews*, 210:187–207.
11. Salinas S, Gov N (2009) The complexity of living: when biology meets theory. Conference on Systems Dynamics of Intracellular Communication. *EMBO reports*, 10:953–7.
12. Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J (2006) Cancer: a Systems Biology disease. *Bio Systems*, 83:81–90.
13. Park J, Lee DS, Christakis NA, Barabási AL (2009) The impact of cellular networks on disease comorbidity. *Molecular systems biology*, 5:262.
14. Schadt EE (2009) Molecular networks as sensors and drivers of common human diseases. *Nature*, 461:218–23.
15. Nguyen TP, Liu W, Jordán F (2011) Inferring pleiotropy by network analysis: linked diseases in the human PPI network. *BMC systems biology*, 5:179.
16. Little MP (2010) Cancer models, genomic instability and somatic cellular Darwinian evolution. *Biology direct*, 5:19
17. Gentles AJ, Gallahan D (2011) Systems biology: confronting the complexity of cancer. *Cancer research*, 71:5961–4.
18. Sonnenschein C, Soto AM (2011) Why systems biology and cancer? *Seminars in cancer biology*, 21:147–9.
19. Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. *Cell*, 100:57–70.

20. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell*, 144:646–74.
21. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, et al. (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. *The New England journal of medicine*, 348:518–27.
22. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*, 63:1129–36.
23. Boyer SN, Wazer DE, Band V (1996) E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer research*, 56:4620–4.
24. Yamato K, Fen J, Kobuchi H, Nasu Y, Yamada T, et al. (2006) Induction of cell death in human papillomavirus 18-positive cervical cancer cells by E6 siRNA. *Cancer gene therapy*, 13:234–41.
25. Molano M, Van den Brule A, Plummer M, Weiderpass E, Posso H, et al. (2003) Determinants of clearance of human papillomavirus infections in Colombian women with normal cytology: a population-based, 5-year follow-up study. *American journal of epidemiology*, 158:486–94.
26. Perez-Plasencia C, Duenas-Gonzalez A, Alatorre-Tavera B (2008) Second hit in cervical carcinogenesis process: involvement of wnt/beta catenin pathway. *International archives of medicine*, 1:10.
27. Higareda-Almaraz JC, Enríquez-Gasca MDR, Hernández-Ortiz M, Resendis-Antonio O, Encarnación-Guevara S (2011) Proteomic patterns of cervical cancer cell lines, a network perspective. *BMC systems biology*, 5:96.
28. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, et al. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology*, 7:1–8.
29. Camon E, Magrane M, Barrell D, Binns D, Fleischmann W, et al. (2003) The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome research*, 13:662–72.
30. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, et al. (2011) ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic acids research*, 39:D712–7.

31. Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniam S, et al. (2008) Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. *Genes, chromosomes & cancer*, 47:755–65.
32. Suzuki H, Forrest ARR, van Nimwegen E, Daub CO, Balwierz PJ, et al. (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature genetics*, 41:553–62.
33. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13:2498–504.
34. Lin CY, Chin CH, Wu HH, Chen SH, Ho CW, et al. (2008) Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology. *Nucleic acids research*, 36:W438–43.
35. The UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic acids research*, 38:D142–8.
36. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, et al. (2009) PID: the Pathway Interaction Database. *Nucleic acids research*, 37:D674–9.
37. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research*, 38:D355–60.
38. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, et al. (2010) The BioPAX community standard for pathway data sharing. *Nature biotechnology*, 28:935–42.
39. Martin A, Ochagavia ME, Rabasa LC, Miranda J, Fernandez-de-Cossio J, et al. (2010) Bisogenet: a new tool for gene network building, visualization and analysis. *BMC bioinformatics*, 11:91.
40. Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, et al. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic acids research*, 36:W423–6.
41. Minn A, Bevilacqua E, Yun J, Rosner M (2012) Identification of novel metastasis suppressor signaling pathways for breast cancer. *Cell Cycle*, 11:2452–2457.
42. Aldridge BB, Burke JM, Lauffenburger D, Sorger PK (2006) Physicochemical modelling of cell signaling pathways. *Nature cell biology*, 8:1195–203.
43. Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends in biochemical sciences*, 27:514–20.

44. Hubbard MJ, Cohen P (1993) On target with a new mechanism for the regulation of protein phosphorylation. *Trends in biochemical sciences*, 18:172–7.
45. Mumby M, Brekken D (2005) Phosphoproteomics: new insights into cellular signaling. *Genome biology*, 6:230.
46. Eswaran J, Cyanam D, Mudvari P, Reddy SDN, Pakala SB, et al. (2012) Transcriptomic landscape of breast cancers through mRNA sequencing. *Scientific reports*, 2:264.
47. Musgrove EA, Sergio CM, Loi S, Inman CK, Anderson LR, et al. (2008) Identification of functional networks of estrogen- and c-Myc-responsive genes and their relationship to response to tamoxifen therapy in breast cancer. *PloS one*, 3:e2987.
48. Fang F, Rycyzyn MA, Clevenger CV (2009) Role of c-Myb during prolactin-induced signal transducer and activator of transcription 5a signaling in breast cancer cells. *Endocrinology*, 150:1597–606.
49. Tang Y, Simoneau AR, Liao W, Yi G, Hope C, et al. (2009) WIF1, a Wnt pathway inhibitor, regulates SKP2 and c-myc expression leading to G1 arrest and growth inhibition of human invasive urinary bladder cancer cells. *Molecular cancer therapeutics*, 8:458–68.
50. Tanno B, Sesti F, Cesi V, Bossi G, Ferrari-Amorotti G, et al. (2010) Expression of Slug is regulated by c-Myb and is required for invasion and bone marrow homing of cancer cells of different origin. *The Journal of biological chemistry*, 285:29434–45.
51. Antoniou AC, Pharoah PDP, McMullan G, Day NE, Stratton MR, et al. (2002) A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British journal of cancer*, 86:76–83.
52. Rosell R, Skrzypski M, Jassem E, Taron M, Bartolucci R, et al. (2007) BRCA1: a novel prognostic factor in resected non-small-cell lung cancer. *PloS one*, 2:e1129.
53. Pinto M, Vieira J, Ribeiro FR, Soares MJ, Henrique R, et al. (2008) Overexpression of the mitotic checkpoint genes BUB1 and BUBR1 is associated with genomic complexity in clear cell kidney carcinomas. *Cellular oncology*, 30:389–95.
54. Ricke RM, Jeganathan KB, van Deursen JM (2011) Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *The Journal of cell biology*, 193:1049–64.
55. Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, et al. (2006) NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proceedings of the National Academy of Sciences of the United States of America*, 103:18261–6.

56. Xu X, Bieda M, Jin VX, Rabinovich A, Oberley MJ, et al. (2007) A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome research*, 17:1550–61.
57. Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, et al. (2010) A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Molecular systems biology*, 6:377.
58. Kalin TV, Wang IC, Ackerson TJ, Major ML, Detrisac CJ, et al. (2006) Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice. *Cancer research*, 66:1712–20.
59. Santin AD, Zhan F, Bignotti E, Siegel ER, Cané S, et al. (2005) Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology*, 331:269–91.
60. Majid S, Dar AA, Saini S, Chen Y, Shahryari V, et al. (2010) Regulation of minichromosome maintenance gene family by microRNA-1296 and genistein in prostate cancer. *Cancer research*, 70:2809–18.
61. Yamagata M, Hasuda K, Stamato T, Tannock IF (1998) The contribution of lactic acid to acidification of tumours: studies of variant cells lacking lactate dehydrogenase. *British journal of cancer*, 77:1726–31.
62. Hirschhaeuser F, Sattler UGA, Mueller-Klieser W (2011) Lactate: a metabolic key player in cancer. *Cancer research*, 71:6921–5.
63. Dang CV (2010) Rethinking the Warburg effect with Myc micromanaging glutamine metabolism. *Cancer research*, 70:859–62.
64. Cairns RA, Harris IS, Mak TW (2011) Regulation of cancer cell metabolism. *Nature reviews. Cancer*, 11:85–95.
65. Ferguson EC, Rathmell JC (2008) New roles for pyruvate kinase M2: working out the Warburg effect. *Trends in biochemical sciences*, 33:359–62.
66. Macheda ML, Rogers S, Best JD (2005) Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. *Journal of cellular physiology*, 202:654–62.
67. Vizán P, Alcarraz-Vizán G, Díaz-Moralli S, Solovjeva ON, Frederiks WM, et al. (2009) Modulation of pentose phosphate pathway during cell cycle progression in human colon adenocarcinoma cell line HT29. *International journal of cancer* 124:2789–96.

68. Feron O (2009) Pyruvate into lactate and back: from the Warburg effect to symbiotic energy fuel exchange in cancer cells. *Radiotherapy and oncology*, 92:329–33.
69. Prat A, Parker JS, Karginova O, Fan C, Livasy C, et al. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research*, 12:R68.
70. Sarrió D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, et al. (2008) Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer research*, 68:989–97.
71. Lu J, Guo H, Treekitkarnmongkol W, Li P, Zhang J, et al. (2009) 14-3-3zeta Cooperates with ErbB2 to promote ductal carcinoma in situ progression to invasive breast cancer by inducing epithelial-mesenchymal transition. *Cancer cell*, 16:195–207.
72. Wouters BG, Koritzinsky M (2008) Hypoxia signaling through mTOR and the unfolded protein response in cancer. *Nature reviews. Cancer*, 8:851–64.
73. Seimiya H, Sawada H, Muramatsu Y, Shimizu M, Ohko K, et al. (2000) Involvement of 14-3-3 proteins in nuclear localization of telomerase. *The EMBO journal*, 19:2652–61.
74. Porter GW, Khuri FR, Fu H (2006) Dynamic 14-3-3/client protein interactions integrate survival and apoptotic pathways. *Seminars in cancer biology*, 16:193–202.
75. Li FQ, Mofunanya A, Harris K, Takemaru KI (2008) Chibby cooperates with 14-3-3 to regulate beta-catenin subcellular distribution and signaling activity. *The Journal of cell biology*, 181:1141–54.
76. Wong EWP, Sun S, Li MWM, Lee WM, Cheng CY (2009) 14-3-3 Protein regulates cell adhesion in the seminiferous epithelium of rat testes. *Endocrinology*, 150:4713–23.
77. Benzinger A, Muster N, Koch HB, Yates JR, Hermeking H (2005) Targeted proteomic analysis of 14-3-3 sigma, a p53 effector commonly silenced in cancer. *Molecular & cellular proteomics*, 4:785–95.
78. Hermeking H (2003) The 14-3-3 cancer connection. *Nature reviews. Cancer*, 3:931–43.
79. Danes CG, Wyszomierski SL, Lu J, Neal CL, Yang W, et al. (2008) 14-3-3 Zeta Down-Regulates P53 in Mammary Epithelial Cells and Confers Luminal Filling. *Cancer research*, 68:1760–7.
80. Niemantsverdriet M, Wagner K, Visser M, Backendorf C (2008) Cellular functions of 14-3-3 zeta in apoptosis and cell adhesion emphasize its oncogenic character. *Oncogene*, 27:1315–9.

81. Grosso AR, Martins S, Carmo-Fonseca M (2008) The emerging role of splicing factors in cancer. *EMBO reports*, 9:1087–93.
82. Gerlinger M, Swanton C (2010) How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British journal of cancer*, 103:1139–43.
83. Greaves M (2010) Cancer stem cells: back to Darwin? *Seminars in cancer biology*, 20:65–70.
84. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25:1105–11.
85. Hubbard TJP, Aken BL, Ayling S, Ballester B, Beal K, et al. (2009) Ensembl 2009. *Nucleic acids research*, 37:D690–7.
86. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7:562–78.
87. Raney BJ, Cline MS, Rosenbloom KR, Dreszer TR, Learned K, et al. (2011) ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic acids research*, 39:D871–5.
88. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10:R25.
89. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2011) NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic acids research*, 39:D1005–10.
90. Salazar E, Díaz-Mejía JJ, Moreno-Hagelsieb G, Martínez-Batallar G, Mora Y, et al. (2010) Characterization of the NifA-RpoN regulon in *Rhizobium etli* in free life and in symbiosis with *Phaseolus vulgaris*. *Applied and environmental microbiology*, 76:4510–20.
91. Wolschin F, Wienkoop S, Weckwerth W (2005) Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics*, 5:4389–97.
92. Wessel D, Flügge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry*, 138:141–3.

Figure Legends

Figure 1: Pipeline

We have integrated different layers of information within biological cell dynamic that tracks the flow of information. First, we performed an analysis of all transcripts in HeLa cells; this analysis provided an overview of gene expression. Subsequently, we performed a RNA-seq differential expression analysis and a query of the over-representation of the activity of TFs. This allowed reconstruction of the metabolic pathways and the signaling and cellular transcriptional regulatory pathways. Finally, we validated this reconstruction with a phosphoproteomic analysis.

Figure 2: Gene expression patterns in the HeLa cells

a) The distribution of total transcripts shows that there are two populations, one low-abundance population and a second larger, high-abundance population. This dichotomy shows that the parameters used to search for low abundance transcripts was successful. b) A graphical representation of the cellular process that was distinguished based on Gene Ontology (GO), using the domain of cellular components in level 3; the amount of retrieved elements was compared against the total size of the pathway. c) A scatter plot showing the quality of the RNA-seq differential expression analysis results, using the epithelial keratinocyte cell line NHEK as a control. There were a total of 3,360 overexpressed genes and 2,129 under-expressed genes. d) The percentage distribution of the level 3 biological process domain GO terms represented by the over-expressed transcripts. e) The percentage distribution of the level 3 biological process domain GO terms represented by the under-expressed transcripts. These charts were constructed from a summary of all the similar GO terms in a functional cellular circuit.

Figure 3: Transcription factor expression networks

a) The percentage distribution of level 3 biological process domain GO terms represented by the over-represented TF network, highlighting "Cell proliferation", "Metabolism of building blocks", "Cellular organization", "Angiogenesis", "Central metabolism" and "Signaling". This chart was constructed from a summary of all the similar GO terms in a functional cellular circuit. b) Hubs that were obtained from the node degree centrality measure of the over-represented TF network. The color indicates the score, with red being the highest and yellow the lowest. c) Hubs were obtained from the betweenness centrality measure of the over-represented TF networks. The color indicates the score, with red being the highest and yellow the lowest.

Figure 4: Network of interconnections and crosstalk among the cellular circuits

The network was made from data obtained by the analysis of signaling pathways and regulation. The nodes represent proteins that make up each of the pathways; each pathway is indicated by using different colors.

Figure 5: Meta-pathways analysis

An analysis was conducted combining the obtained signaling and transcriptional regulation pathways; the edges indicate the regulatory or hierarchical relationship, and the nodes indicate the pathway. The colors denote each of the hallmarks of cancer, with the two most representative hallmarks indicated per node.

Figure 6: Over-represented metabolic pathways

Metabolic maps were reconstructed based on the KEGG database and the over-represented TF networks and identified phosphoprotein analyses, resulting in a map of general metabolism, pyruvate metabolism and glycolysis. The over-expressed transcripts are displayed in green, the under-expressed transcripts in red and the identified phosphoproteins in purple.

Figure 7: Cell cycle and cell adhesion molecules pathways

a) A reconstruction of cell cycle and MCM complex pathways. b) A reconstruction of cell adhesion molecules. Both maps were constructed based on the KEGG database and the transcriptomics, over-represented TF networks and identified-phosphoprotein analyses. The overexpressed transcripts are displayed in green, the under-expressed transcripts in red and the identified phosphoproteins in purple.

Figure 8: Spliceosome pathway

The spliceosome pathway was reconstructed based on the KEGG database and the transcriptomics, over-represented TF networks and identified phosphoprotein analyses. The over-expressed transcripts are displayed in green, the under-expressed transcripts in red and the identified phosphoproteins in purple.

Tables

Table 1. Enriched PID pathway–based sets of overexpressed transcripts.

The set size refers to the number of transcripts that have a Uniprot ID in the corresponding PID pathway–based set at the ConsensusPathDB site. The number of candidates contained refers to the number of proteins that are part of the extended network and appear as part of the pathway. P-values were calculated using a hypergeometric test; Q-values represent a correction of the P-values for multiple testing using the false discovery rate method.

Pathway name	Set size	Candidates	P-value	Q-value
DNA replication	36	28 (77.8%)	7.24E-15	1.67E-12
Cell cycle	124	53 (42.7%)	9.42E-11	1.09E-08
Homologous recombination	28	20 (71.4%)	7.71E-10	5.94E-08
Systemic lupus erythematosus	138	51 (37.5%)	4.46E-08	2.58E-06
Fanconi anemia pathway	52	26 (50.0%)	1.32E-07	6.12E-06
Mismatch repair	23	15 (65.2%)	6.97E-07	2.68E-05
Base excision repair	33	18 (54.5%)	2.27E-06	7.49E-05
Transcriptional misregulation in cancer	180	53 (29.8%)	6.84E-05	0.00198
Lysine degradation	49	20 (40.8%)	0.000148	0.0037
Purine metabolism	166	49 (29.5%)	0.00016	0.0037
One carbon pool by folate	19	10 (52.6%)	0.000642	0.0135
RNA transport	157	44 (28.0%)	0.00113	0.0217
HTLV-I infection	263	67 (25.5%)	0.00124	0.0221
Nucleotide excision repair	46	17 (37.0%)	0.00172	0.0284
Pyrimidine metabolism	101	30 (29.7%)	0.00253	0.039
Spliceosome	127	35 (27.6%)	0.00463	0.0668
Small cell lung cancer	87	25 (28.7%)	0.00876	0.119

Additional files

Additional file 1 – Additional Tables 1-12

Table A1: Enriched gene ontology level 3 categories of biological processes of total expression analysis. Table A2: Enriched gene ontology level 3 categories of biological processes of over-expressed genes. Table A3: Enriched gene ontology level 3 categories of biological processes of over-expressed-TFs networks. Table A4: Enriched gene ontology level 3 categories of biological processes of under-expressed-TFs networks. Table A5: Enriched PID pathway-based sets of over-expressed transcripts. Table A6: Enriched PID pathway-based sets of sub-expressed transcripts. Table A7: list of over and under expressed transcripts members of the regulatory network "Direct effectors of p53". Table A8: Enriched KEGG pathway-based sets of under-expressed transcripts. Table A9: Phosphoproteins relationship according to which replicates were identified. Table A10: Enriched gene ontology level 3 categories of biological processes of identified phosphoproteins. Table A11: Enriched PID pathway-based sets of identified phosphoproteins. Table A12: Enriched KEGG pathway-based sets of identified phosphoproteins.

Anexo 3

Material Adicional

Location	Pathway size	Genes present
GO:0005737 cytoplasm	8475	8253 (97.4%)
GO:0044444 cytoplasmic part	6092	5961 (97.9%)
GO:0043231 intracellular membrane-bounded organelle	8980	8704 (97.0%)
GO:0070013 intracellular organelle lumen	2668	2621 (98.3%)
GO:0044428 nuclear part	2594	2543 (98.1%)
GO:0005626 insoluble fraction	904	896 (99.1%)
GO:0043232 intracellular non-membrane-bounded organelle	3310	3225 (97.5%)
GO:0031967 organelle envelope	751	740 (98.5%)
GO:0044459 plasma membrane part	1958	1908 (97.4%)
GO:0005625 soluble fraction	395	392 (99.2%)
GO:0044429 mitochondrial part	705	694 (98.4%)
GO:0044430 cytoskeletal part	1153	1127 (97.8%)
GO:0044432 endoplasmic reticulum part	770	756 (98.2%)
GO:0044431 Golgi apparatus part	597	587 (98.3%)
GO:0043005 neuron projection	522	514 (98.5%)
GO:0031988 membrane-bounded vesicle	758	743 (98.0%)
GO:0031410 cytoplasmic vesicle	784	768 (98.0%)
GO:0042175 nuclear outer membrane-endoplasmic reticulum membrane network	697	683 (98.0%)
GO:0044437 vacuolar part	172	171 (99.4%)
GO:0031300 intrinsic to organelle membrane	167	166 (99.4%)
GO:0044440 endosomal part	275	271 (98.5%)
GO:0016585 chromatin remodeling complex	109	109 (100.0%)

Tabla S1 Enriquecimiento de ontología genética nivel 3 categoría de “procesos biológicos” a partir de análisis de la expresión total. *El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento. Vale la pena señalar que las expresiones de cada término GO cumplieron por lo menos 97% total de cada término.*

Gene Ontology term	Set size	Candidates contained	P-value	Q-value
GO:0022403 cell cycle phase	772	309 (40.0%)	1.29E-53	7.15E-51
GO:0000278 mitotic cell cycle	687	261 (38.0%)	2.21E-40	6.12E-38
GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4958	1120 (22.6%)	8.03E-32	1.48E-29
GO:0034641 cellular nitrogen compound metabolic process	5373	1194 (22.2%)	3.38E-31	4.68E-29
GO:0071842 cellular component organization at cellular level	2857	704 (24.7%)	4.98E-29	5.52E-27
GO:0044260 cellular macromolecule metabolic process	6149	1322 (21.5%)	2.01E-28	1.85E-26
GO:0033554 cellular response to stress	957	262 (27.4%)	2.86E-16	2.26E-14
GO:0051321 meiotic cell cycle	138	62 (44.9%)	2.48E-14	1.72E-12
GO:0007050 cell cycle arrest	335	113 (33.7%)	9.40E-14	5.79E-12
GO:0009059 macromolecule biosynthetic process	3775	798 (21.2%)	4.45E-13	2.46E-11
GO:0044249 cellular biosynthetic process	4603	949 (20.6%)	5.99E-13	3.02E-11
GO:0043933 macromolecular complex subunit organization	1008	254 (25.2%)	1.66E-11	7.65E-10
GO:0010467 gene expression	3914	806 (20.6%)	1.26E-10	5.37E-09
GO:0019222 regulation of metabolic process	4485	907 (20.2%)	3.22E-10	1.28E-08
GO:0000226 microtubule cytoskeleton organization	215	73 (34.0%)	1.60E-09	5.89E-08
GO:0000086 G2/M transition of mitotic cell cycle	135	52 (38.5%)	2.72E-09	9.13E-08
GO:0022607 cellular component assembly	1282	299 (23.3%)	2.80E-09	9.13E-08
GO:0048519 negative regulation of biological process	2583	548 (21.2%)	4.57E-09	1.41E-07
GO:0050794 regulation of cellular process	7306	1393 (19.1%)	1.21E-08	3.53E-07
GO:0000819 sister chromatid segregation	45	24 (53.3%)	3.54E-08	9.82E-07
GO:0007062 sister chromatid cohesion	26	17 (65.4%)	6.04E-08	1.59E-06
GO:0007091 mitotic metaphase/anaphase transition	43	22 (51.2%)	3.45E-07	8.68E-06
GO:0000083 regulation of transcription involved in G1/S phase of mitotic cell cycle	21	14 (66.7%)	6.49E-07	1.53E-05
GO:0007131 reciprocal meiotic recombination	35	19 (54.3%)	6.62E-07	1.53E-05
GO:0043412 macromolecule modification	2212	458 (20.7%)	2.16E-06	4.79E-05
GO:0000082 G1/S transition of mitotic cell cycle	169	53 (31.4%)	4.13E-06	8.80E-05
GO:0006725 cellular aromatic compound metabolic process	223	65 (29.1%)	5.88E-06	0.000121
GO:0007098 centrosome cycle	31	16 (51.6%)	1.20E-05	0.000238
GO:0050658 RNA transport	127	41 (32.3%)	2.27E-05	0.000435
GO:0006403 RNA localization	132	42 (31.8%)	2.69E-05	0.000483
GO:0015931 nucleobase, nucleoside, nucleotide and nucleic acid transport	149	46 (30.9%)	2.70E-05	0.000483
GO:0000216 M/G1 transition of mitotic cell cycle	79	28 (35.4%)	7.04E-05	0.00122

Tabla S2 Enriquecimiento de ontología genética nivel 3 categoría de “procesos biológicos” a partir de análisis de los transcritos sobre expresados. El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Gene Ontology term	Set size	Candidates contained	P-value	Q-value
GO:0022403 cell cycle phase	772	257 (33.3%)	5.33E-49	2.86E-46
GO:0034641 cellular nitrogen compound metabolic process	5373	1002 (18.7%)	7.02E-43	1.88E-40
GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4958	936 (18.9%)	2.23E-41	3.98E-39
GO:0044260 cellular macromolecule metabolic process	6149	1098 (17.9%)	6.4E-39	7.59E-37
GO:0000278 mitotic cell cycle	687	220 (32.0%)	7.08E-39	7.59E-37
GO:0071842 cellular component organization at cellular level	2857	602 (21.1%)	1.04E-37	9.33E-36
GO:0044249 cellular biosynthetic process	4603	809 (17.6%)	3.4E-23	2.6E-21
GO:0033554 cellular response to stress	957	234 (24.5%)	2.68E-22	1.8E-20
GO:0009059 macromolecule biosynthetic process	3775	669 (17.7%)	2.83E-19	1.68E-17
GO:0019222 regulation of metabolic process	4485	766 (17.1%)	4.77E-18	2.56E-16
GO:0050794 regulation of cellular process	7306	1156 (15.8%)	1.45E-17	7.08E-16
GO:0043933 macromolecular complex subunit organization	1008	225 (22.4%)	1.81E-16	8.08E-15
GO:0010467 gene expression	3914	667 (17.1%)	4.19E-15	1.73E-13
GO:0007050 cell cycle arrest	335	98 (29.3%)	6.28E-15	2.4E-13
GO:0048519 negative regulation of biological process	2583	469 (18.2%)	8.27E-15	2.95E-13
GO:0022607 cellular component assembly	1282	258 (20.1%)	5.03E-13	1.68E-11
GO:0051321 meiotic cell cycle	138	51 (37.0%)	1.49E-12	4.69E-11
GO:0000086 G2/M transition of mitotic cell cycle	135	49 (36.3%)	8.75E-12	2.6E-10
GO:0048518 positive regulation of biological process	2857	493 (17.3%)	1.03E-11	2.92E-10
GO:0043412 macromolecule modification	2212	384 (17.4%)	1.8E-09	4.84E-08
GO:0042592 homeostatic process	996	196 (19.7%)	2.94E-09	0.000000075
GO:0048731 system development	3031	501 (16.5%)	5.11E-09	0.000000125
GO:0000819 sister chromatid segregation	45	22 (48.9%)	8.2E-09	0.000000191
GO:0007131 reciprocal meiotic recombination	35	19 (54.3%)	9.23E-09	0.000000206
GO:0019538 protein metabolic process	3414	553 (16.2%)	1.36E-08	0.000000291
GO:0048513 organ development	2079	358 (17.2%)	1.75E-08	0.00000036
GO:0000226 microtubule cytoskeleton organization	215	59 (27.4%)	2.09E-08	0.000000415
GO:0007091 mitotic metaphase/anaphase transition	43	20 (46.5%)	0.000000113	0.00000216
GO:0030154 cell differentiation	2273	379 (16.7%)	0.000000256	0.00000473
GO:0009725 response to hormone stimulus	660	132 (20.0%)	0.000000492	0.00000879
GO:0051128 regulation of cellular component organization	935	175 (18.7%)	0.000000733	0.0000127
GO:0010033 response to organic substance	1394	245 (17.6%)	0.000000895	0.000015
GO:0001666 response to hypoxia	203	52 (25.6%)	0.00000143	0.0000233
GO:0070482 response to oxygen levels	221	55 (24.9%)	0.00000191	0.0000301
GO:0006793 phosphorus metabolic process	1306	228 (17.5%)	0.00000363	0.0000556
GO:0070887 cellular response to chemical stimulus	1112	197 (17.7%)	0.00000674	0.0001
GO:0050793 regulation of developmental process	1066	189 (17.7%)	0.00000994	0.000144
GO:0070271 protein complex biogenesis	604	117 (19.4%)	0.0000106	0.000149
GO:0006725 cellular aromatic compound metabolic process	223	53 (23.8%)	0.0000122	0.000167

GO:0000082	G1/S transition of mitotic cell cycle	169	43 (25.4%)	0.0000135	0.000181
GO:0007599	hemostasis	512	101 (19.7%)	0.0000194	0.000254
GO:0009792	embryo development ending in birth or egg hatching	399	82 (20.6%)	0.000026	0.000331
GO:0007596	blood coagulation	508	99 (19.5%)	0.0000387	0.000483
GO:0010035	response to inorganic substance	332	70 (21.1%)	0.0000424	0.000507
GO:0042493	response to drug	338	71 (21.0%)	0.0000425	0.000507
GO:0007062	sister chromatid cohesion	26	12 (46.2%)	0.0000445	0.000518
GO:0012501	programmed cell death	1390	233 (16.8%)	0.0000489	0.000557
GO:0009611	response to wounding	1017	177 (17.4%)	0.0000554	0.000619
GO:0007098	centrosome cycle	31	13 (41.9%)	0.0000738	0.000808
GO:0050658	RNA transport	127	33 (26.0%)	0.0000836	0.000896
GO:0000083	regulation of transcription involved in G1/S phase of mitotic cell cycle	21	10 (47.6%)	0.000141	0.00148
GO:0000216	M/G1 transition of mitotic cell cycle	79	23 (29.1%)	0.000158	0.0016
GO:0007052	mitotic spindle organization	25	11 (44.0%)	0.000158	0.0016
GO:0006403	RNA localization	132	33 (25.0%)	0.000185	0.00184
GO:0015931	nucleobase, nucleoside, nucleotide and nucleic acid transport	149	36 (24.2%)	0.000202	0.00197
GO:0001568	blood vessel development	386	76 (19.7%)	0.000212	0.00202
GO:0048514	blood vessel morphogenesis	339	67 (19.8%)	0.000436	0.0041
GO:0005975	carbohydrate metabolic process	646	115 (17.8%)	0.000464	0.00429
GO:0051239	regulation of multicellular organismal process	1402	227 (16.2%)	0.000497	0.00451
GO:0006959	humoral immune response	112	28 (25.0%)	0.00055	0.00491
GO:0035295	tube development	331	65 (19.6%)	0.000628	0.00552
GO:0048468	cell development	1211	198 (16.4%)	0.000694	0.006
GO:0006979	response to oxidative stress	223	47 (21.1%)	0.000724	0.00616
GO:0044093	positive regulation of molecular function	982	164 (16.7%)	0.000772	0.00646
GO:0001525	angiogenesis	285	57 (20.0%)	0.000834	0.00688
GO:0044283	small molecule biosynthetic process	666	116 (17.4%)	0.000951	0.00772
GO:0006091	generation of precursor metabolites and energy	427	79 (18.5%)	0.00113	0.009
GO:0009057	macromolecule catabolic process	715	123 (17.2%)	0.00115	0.009
GO:0002237	response to molecule of bacterial origin	186	40 (21.5%)	0.00116	0.009
GO:0035821	modification of morphology or physiology of other organism	35	12 (34.3%)	0.00122	0.00936
GO:0042023	DNA endoreduplication	5	4 (80.0%)	0.00137	0.0103
GO:0048469	cell maturation	91	23 (25.3%)	0.00141	0.0105
GO:0007584	response to nutrient	225	46 (20.4%)	0.00159	0.0115
GO:0009314	response to radiation	268	53 (19.8%)	0.00161	0.0115
GO:0071495	cellular response to endogenous stimulus	381	71 (18.6%)	0.00161	0.0115
GO:0007165	signal transduction	3729	548 (14.7%)	0.00174	0.0122
GO:0040008	regulation of growth	440	80 (18.2%)	0.00176	0.0122
GO:0021915	neural tube development	99	24 (24.2%)	0.00207	0.0142
GO:0048659	smooth muscle cell proliferation	62	17 (27.4%)	0.00225	0.0152
GO:0009636	response to toxin	95	23 (24.2%)	0.00258	0.0173

GO:0051303	establishment of chromosome localization	20	8 (40.0%)	0.00266	0.0174
GO:0060323	head morphogenesis	20	8 (40.0%)	0.00266	0.0174
GO:0042558	pteridine-containing compound metabolic process	34	11 (32.4%)	0.00327	0.0211
GO:0046677	response to antibiotic	30	10 (33.3%)	0.00389	0.0248
GO:0046148	pigment biosynthetic process	45	13 (28.9%)	0.00438	0.0276
GO:0009991	response to extracellular stimulus	351	64 (18.2%)	0.00446	0.0278
GO:0009887	organ morphogenesis	660	110 (16.7%)	0.00568	0.035
GO:0051702	interaction with symbiont	18	7 (38.9%)	0.00589	0.0356
GO:0045137	development of primary sexual characteristics	203	40 (19.7%)	0.00597	0.0356
GO:0046483	heterocycle metabolic process	1019	162 (15.9%)	0.00611	0.0356
GO:0001556	oocyte maturation	14	6 (42.9%)	0.00618	0.0356
GO:0010573	vascular endothelial growth factor production	14	6 (42.9%)	0.00618	0.0356
GO:0045132	meiotic chromosome segregation	14	6 (42.9%)	0.00618	0.0356
GO:0006066	alcohol metabolic process	521	89 (17.1%)	0.00628	0.0358
GO:0055086	nucleobase, nucleoside and nucleotide metabolic process	911	146 (16.0%)	0.00674	0.038
GO:0031647	regulation of protein stability	74	18 (24.3%)	0.00682	0.0381
GO:0007548	sex differentiation	230	44 (19.1%)	0.00707	0.0391
GO:0032844	regulation of homeostatic process	181	36 (19.9%)	0.00754	0.0412
GO:0044248	cellular catabolic process	1483	227 (15.3%)	0.00838	0.0453
GO:0007099	centriole replication	11	5 (45.5%)	0.00932	0.05
GO:0048729	tissue morphogenesis	350	62 (17.7%)	0.00953	0.0506
GO:0042180	cellular ketone metabolic process	802	129 (16.1%)	0.00963	0.0506
GO:0044403	symbiosis, encompassing mutualism through parasitism	88	20 (22.7%)	0.00991	0.0516

Tabla S3 Enriquecimiento de ontología genética nivel 3 categoría de “procesos biológicos” a partir de análisis de la sobre representación de FT. El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

gene ontology term	set size	candidates contained	p-value	q-value
GO:0048731 system development	3031	415 (13.7%)	1.36E-26	7.11E-24
GO:0009888 tissue development	1010	181 (17.9%)	4.44E-23	1.16E-20
GO:0007165 signal transduction	3729	459 (12.3%)	1.16E-19	2.02E-17
GO:0048513 organ development	2079	291 (14.0%)	2.23E-19	2.92E-17
GO:0030154 cell differentiation	2273	308 (13.6%)	2.14E-18	2.23E-16
GO:0051239 regulation of multicellular organismal process	1402	205 (14.6%)	1.52E-15	1.33E-13
GO:0010033 response to organic substance	1394	201 (14.4%)	1.28E-14	9.56E-13
GO:0032989 cellular component morphogenesis	800	133 (16.6%)	1.96E-14	1.28E-12
GO:0048514 blood vessel morphogenesis	339	74 (21.8%)	2.21E-14	1.29E-12
GO:0009611 response to wounding	1017	158 (15.5%)	2.48E-14	1.3E-12
GO:0001525 angiogenesis	285	65 (22.8%)	1.01E-13	4.8E-12
GO:0051128 regulation of cellular component organization	935	146 (15.6%)	1.75E-13	7.29E-12
GO:0070887 cellular response to chemical stimulus	1112	166 (14.9%)	1.81E-13	7.29E-12
GO:0001568 blood vessel development	386	78 (20.2%)	3.31E-13	1.24E-11
GO:0048468 cell development	1211	174 (14.4%)	1.35E-12	4.69E-11
GO:0034329 cell junction assembly	149	42 (28.2%)	1.63E-12	5.34E-11
GO:0048583 regulation of response to stimulus	1870	242 (12.9%)	3.65E-12	1.12E-10
GO:0050794 regulation of cellular process	7306	746 (10.2%)	5.3E-12	1.54E-10
GO:0023051 regulation of signaling	1622	213 (13.1%)	2.39E-11	6.57E-10
GO:0032879 regulation of localization	1009	146 (14.5%)	6.23E-11	1.63E-09
GO:0050793 regulation of developmental process	1066	152 (14.3%)	7.35E-11	1.74E-09
GO:0048518 positive regulation of biological process	2857	334 (11.7%)	7.43E-11	1.74E-09
GO:0006629 lipid metabolic process	1030	148 (14.4%)	7.63E-11	1.74E-09
GO:0048519 negative regulation of biological process	2583	303 (11.7%)	5.29E-10	1.15E-08
GO:0040012 regulation of locomotion	315	61 (19.4%)	7.63E-10	0.00000016
GO:0009887 organ morphogenesis	660	103 (15.6%)	8.16E-10	1.64E-08
GO:0048729 tissue morphogenesis	350	65 (18.6%)	1.28E-09	2.49E-08
GO:0045216 cell-cell junction organization	100	28 (28.0%)	9.7E-09	0.000000181
GO:0046903 secretion	716	103 (14.4%)	6.44E-08	0.00000116
GO:0012501 programmed cell death	1390	174 (12.5%)	6.82E-08	0.00000119
GO:0022612 gland morphogenesis	90	25 (27.8%)	7.07E-08	0.00000119
GO:0030036 actin cytoskeleton organization	337	58 (17.2%)	0.000000159	0.0000026
GO:0035295 tube development	331	57 (17.2%)	0.000000199	0.00000316
GO:0002682 regulation of immune system process	748	103 (13.8%)	0.000000555	0.00000854
GO:0030168 platelet activation	246	45 (18.3%)	0.000000683	0.0000102
GO:0044255 cellular lipid metabolic process	706	98 (13.9%)	0.000000729	0.0000106
GO:0002237 response to molecule of bacterial origin	186	37 (19.9%)	0.000000822	0.0000116
GO:0010035 response to inorganic substance	332	55 (16.6%)	0.00000115	0.0000159
GO:0071842 cellular component organization at cellular level	2857	309 (10.8%)	0.00000125	0.0000168

GO:0061024	membrane organization	552	80 (14.5%)	0.00000145	0.0000185
GO:0070482	response to oxygen levels	221	41 (18.6%)	0.00000145	0.0000185
GO:0008610	lipid biosynthetic process	448	68 (15.2%)	0.00000176	0.0000219
GO:0032612	interleukin-1 production	39	14 (35.9%)	0.00000188	0.0000229
GO:0044093	positive regulation of molecular function	982	125 (12.7%)	0.00000238	0.0000282
GO:0001666	response to hypoxia	203	38 (18.7%)	0.00000278	0.0000323
GO:0044092	negative regulation of molecular function	554	79 (14.3%)	0.00000316	0.0000359
GO:0044087	regulation of cellular component biogenesis	240	42 (17.5%)	0.00000514	0.0000572
GO:0006793	phosphorus metabolic process	1306	156 (11.9%)	0.00000559	0.0000609
GO:0042633	hair cycle	73	19 (26.0%)	0.00000737	0.0000787
GO:0006897	endocytosis	303	49 (16.2%)	0.00000861	0.0000883
GO:0010324	membrane invagination	303	49 (16.2%)	0.00000861	0.0000883
GO:0022404	molting cycle process	69	18 (26.1%)	0.0000123	0.000124
GO:0007599	hemostasis	512	72 (14.1%)	0.0000139	0.000137
GO:0006952	defense response	984	121 (12.3%)	0.000018	0.000175
GO:0035239	tube morphogenesis	220	38 (17.3%)	0.0000193	0.000183
GO:0007596	blood coagulation	508	71 (14.0%)	0.0000196	0.000183
GO:0016337	cell-cell adhesion	330	51 (15.5%)	0.0000205	0.000188
GO:0006066	alcohol metabolic process	521	72 (13.8%)	0.0000251	0.000226
GO:0031589	cell-substrate adhesion	169	31 (18.3%)	0.0000336	0.000297
GO:0061061	muscle structure development	363	54 (14.9%)	0.0000352	0.000307
GO:0090132	epithelium migration	18	8 (44.4%)	0.0000534	0.000451
GO:0032844	regulation of homeostatic process	181	32 (17.7%)	0.0000535	0.000451
GO:0048608	reproductive structure development	224	37 (16.5%)	0.0000663	0.000544
GO:0009725	response to hormone stimulus	660	85 (12.9%)	0.0000665	0.000544
GO:0043412	macromolecule modification	2212	236 (10.7%)	0.0000779	0.000618
GO:0042493	response to drug	338	50 (14.8%)	0.000078	0.000618
GO:0022607	cellular component assembly	1282	146 (11.4%)	0.000118	0.000918
GO:0060560	developmental growth involved in morphogenesis	97	20 (20.6%)	0.000158	0.00122
GO:0071706	tumor necrosis factor superfamily cytokine production	62	15 (24.2%)	0.000162	0.00122
GO:0010876	lipid localization	200	33 (16.5%)	0.000164	0.00122
GO:0045087	innate immune response	440	60 (13.6%)	0.000168	0.00124
GO:0044259	multicellular organismal macromolecule metabolic process	50	13 (26.0%)	0.000203	0.00147
GO:0032943	mononuclear cell proliferation	153	27 (17.6%)	0.000205	0.00147
GO:0042592	homeostatic process	996	116 (11.6%)	0.00026	0.00184
GO:0006979	response to oxidative stress	223	35 (15.7%)	0.000286	0.00197
GO:0007620	copulation	22	8 (36.4%)	0.000286	0.00197
GO:0006887	exocytosis	241	37 (15.4%)	0.000304	0.00207
GO:0060056	mammary gland involution	13	6 (46.2%)	0.000377	0.00253
GO:0048839	inner ear development	111	21 (18.9%)	0.000384	0.00254
GO:0061138	morphogenesis of a branching epithelium	136	24 (17.6%)	0.00045	0.00294
GO:0001935	endothelial cell proliferation	68	15 (22.1%)	0.000478	0.00308

GO:0040008	regulation of growth	440	58 (13.2%)	0.000514	0.00328
GO:0019538	protein metabolic process	3414	339 (9.9%)	0.000527	0.00332
GO:0032535	regulation of cellular component size	146	25 (17.1%)	0.000552	0.00342
GO:0048598	embryonic morphogenesis	385	52 (13.5%)	0.000556	0.00342
GO:0055082	cellular chemical homeostasis	510	65 (12.7%)	0.000613	0.00373
GO:0031099	regeneration	115	21 (18.3%)	0.000627	0.00377
GO:0007584	response to nutrient	225	34 (15.1%)	0.000699	0.00415
GO:0003179	heart valve morphogenesis	10	5 (50.0%)	0.000767	0.00442
GO:0071495	cellular response to endogenous stimulus	381	51 (13.4%)	0.000768	0.00442
GO:0016486	peptide hormone processing	25	8 (32.0%)	0.00077	0.00442
GO:0090066	regulation of anatomical structure size	245	36 (14.7%)	0.000838	0.00476
GO:0030728	ovulation	20	7 (35.0%)	0.000903	0.00508
GO:0009612	response to mechanical stimulus	111	20 (18.0%)	0.000994	0.00553
GO:0031069	hair follicle morphogenesis	26	8 (30.8%)	0.00103	0.00561
GO:0008038	neuron recognition	26	8 (30.8%)	0.00103	0.00561
GO:0044282	small molecule catabolic process	800	93 (11.6%)	0.0011	0.00593
GO:0048645	organ formation	39	10 (25.6%)	0.00122	0.0065
GO:0003170	heart valve development	11	5 (45.5%)	0.00131	0.00687
GO:0060349	bone morphogenesis	46	11 (23.9%)	0.00131	0.00687
GO:0030595	leukocyte chemotaxis	90	17 (18.9%)	0.00135	0.00695
GO:0071695	anatomical structure maturation	27	8 (29.6%)	0.00136	0.00695
GO:0071496	cellular response to external stimulus	190	29 (15.3%)	0.00142	0.00722
GO:0071702	organic substance transport	498	62 (12.4%)	0.00145	0.0073
GO:0048546	digestive tract morphogenesis	40	10 (25.0%)	0.0015	0.00748
GO:0051098	regulation of binding	124	21 (16.9%)	0.0017	0.00835
GO:0009595	detection of biotic stimulus	22	7 (31.8%)	0.00171	0.00835
GO:0045137	development of primary sexual characteristics	203	30 (14.8%)	0.00199	0.00955
GO:0008104	protein localization	1252	135 (10.8%)	0.00199	0.00955
GO:0070271	protein complex biogenesis	604	72 (11.9%)	0.00203	0.00966
GO:0048286	lung alveolus development	35	9 (25.7%)	0.00207	0.00972
GO:0009991	response to extracellular stimulus	351	46 (13.1%)	0.00208	0.00972
GO:0046849	bone remodeling	42	10 (23.8%)	0.00223	0.0103
GO:0021675	nerve development	49	11 (22.4%)	0.00227	0.0104
GO:0002568	somatic diversification of T cell receptor genes	4	3 (75.0%)	0.00229	0.0104
GO:0003012	muscle system process	252	35 (13.9%)	0.00261	0.0118
GO:0001946	lymphangiogenesis	8	4 (50.0%)	0.00274	0.0122
GO:0070997	neuron death	146	23 (15.8%)	0.0028	0.0124
GO:0016042	lipid catabolic process	226	32 (14.2%)	0.00287	0.0126
GO:0045104	intermediate filament cytoskeleton organization	24	7 (29.2%)	0.00298	0.013
GO:0032602	chemokine production	37	9 (24.3%)	0.00313	0.0135
GO:0001945	lymph vessel development	13	5 (38.5%)	0.00316	0.0135
GO:0045165	cell fate commitment	174	26 (14.9%)	0.00328	0.0139
GO:0003002	regionalization	247	34 (13.8%)	0.00344	0.0145

GO:0009792	embryo development ending in birth or egg hatching	399	50 (12.5%)	0.00348	0.0146
GO:0002089	lens morphogenesis in camera-type eye	19	6 (31.6%)	0.00382	0.0159
GO:0048754	branching morphogenesis of a tube	124	20 (16.1%)	0.00388	0.016
GO:0003013	circulatory system process	278	37 (13.3%)	0.00417	0.017
GO:0044248	cellular catabolic process	1483	154 (10.4%)	0.00436	0.0177
GO:0002088	lens development in camera-type eye	46	10 (21.7%)	0.00453	0.0181
GO:0042403	thyroid hormone metabolic process	14	5 (35.7%)	0.00457	0.0181
GO:0009415	response to water	9	4 (44.4%)	0.0046	0.0181
GO:0042640	anagen	9	4 (44.4%)	0.0046	0.0181
GO:0006082	organic acid metabolic process	799	89 (11.1%)	0.00472	0.0184
GO:0010817	regulation of hormone levels	357	45 (12.6%)	0.00483	0.0187
GO:0021536	diencephalon development	54	11 (20.4%)	0.00504	0.0194
GO:0030900	forebrain development	226	31 (13.7%)	0.00533	0.0202
GO:0048659	smooth muscle cell proliferation	62	12 (19.4%)	0.00537	0.0202
GO:0016197	endosome transport	119	19 (16.0%)	0.00538	0.0202
GO:0060512	prostate gland morphogenesis	27	7 (25.9%)	0.00611	0.0228
GO:0060343	trabecula formation	15	5 (33.3%)	0.00638	0.0237
GO:0010259	multicellular organismal aging	21	6 (28.6%)	0.0066	0.0243
GO:0007548	sex differentiation	230	31 (13.5%)	0.00687	0.0251
GO:0060008	Sertoli cell differentiation	10	4 (40.0%)	0.00716	0.0251
GO:0021984	adenohypophysis development	10	4 (40.0%)	0.00716	0.0251
GO:0001550	ovarian cumulus expansion	2	2 (100.0%)	0.0072	0.0251
GO:0002238	response to molecule of fungal origin	2	2 (100.0%)	0.0072	0.0251
GO:0006788	heme oxidation	2	2 (100.0%)	0.0072	0.0251
GO:0045229	external encapsulating structure organization	2	2 (100.0%)	0.0072	0.0251
GO:0042335	cuticle development	2	2 (100.0%)	0.0072	0.0251
GO:0060711	labyrinthine layer development	28	7 (25.0%)	0.00756	0.0262
GO:0072593	reactive oxygen species metabolic process	98	16 (16.3%)	0.00818	0.0281
GO:0072006	nephron development	50	10 (20.0%)	0.00838	0.0287
GO:0061383	trabecula morphogenesis	16	5 (31.2%)	0.00864	0.0292
GO:0006805	xenobiotic metabolic process	142	21 (14.8%)	0.00865	0.0292
GO:0048588	developmental cell growth	74	13 (17.6%)	0.00891	0.0299
GO:0035019	somatic stem cell maintenance	29	7 (24.1%)	0.00926	0.0308
GO:0035148	tube formation	83	14 (16.9%)	0.00973	0.0322

Tabla S4 Enriquecimiento de ontología genética nivel 3 categoría de “procesos biológicos” a partir de análisis de los transcritos sub expresados. El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Pathway name	Set size	Candidates contained	P-value	Q-value
Fanconi anemia pathway	47	32	4.66E-10	6.80E-08
AP-1 transcription factor network	70	26	2.98E-08	1.45E-06
ATR signaling pathway	37	26	2.98E-08	1.45E-06
E2F transcription factor network	77	38	8.05E-08	2.94E-06
PLK1 signaling events	44	24	1.19E-07	3.48E-06
Regulation of Telomerase	73	23	2.38E-07	5.80E-06
Direct p53 effectors	144	21	9.54E-07	1.74E-05
FOXM1 transcription factor network	42	21	9.54E-07	1.74E-05
Beta1 integrin cell surface interactions	66	20	1.91E-06	3.09E-05
Aurora B signaling	43	19	3.81E-06	5.06E-05
Regulation of nuclear SMAD2/3 signaling	78	19	3.81E-06	5.06E-05
C-MYB transcription factor network	85	18	7.63E-06	7.43E-05
ATF-2 transcription factor network	61	18	7.63E-06	7.43E-05
Regulation of retinoblastoma protein	66	18	7.63E-06	7.43E-05
PDGFR-beta signaling pathway	130	18	7.63E-06	7.43E-05
Validated targets of C-MYC transcriptional activation	87	26	8.80E-06	8.03E-05
HIF-1-alpha transcription factor network	66	17	1.53E-05	0.000117
BARD1 signaling events	31	17	1.53E-05	0.000117
p73 transcription factor network	79	17	1.53E-05	0.000117
Validated nuclear estrogen receptor alpha network	66	16	3.05E-05	0.000171
Caspase cascade in apoptosis	53	16	3.05E-05	0.000171
ATM pathway	35	16	3.05E-05	0.000171
Validated targets of C-MYC transcriptional repression	75	16	3.05E-05	0.000171
Coregulation of Androgen receptor activity	66	16	3.05E-05	0.000171
Integrins in angiogenesis	65	16	3.05E-05	0.000171
Glucocorticoid receptor regulatory network	80	16	3.05E-05	0.000171
Signaling events mediated by HDAC Class III	40	17	0.000117	0.000575
ErbB1 downstream signaling	108	14	0.000122	0.000575
p53 pathway	59	14	0.000122	0.000575
mTOR signaling pathway	65	14	0.000122	0.000575
IL4-mediated signaling events	66	14	0.000122	0.000575
Validated transcriptional targets of TAp63 isoforms	55	13	0.000244	0.00105
Syndecan-1-mediated signaling events	44	13	0.000244	0.00105
Regulation of Androgen receptor activity	51	13	0.000244	0.00105
HIF-2-alpha transcription factor network	36	12	0.000488	0.00183
Validated transcriptional targets of AP1 family members	37	12	0.000488	0.00183

Fra1 and Fra2				
RhoA signaling pathway	47	12	0.000488	0.00183
Beta3 integrin cell surface interactions	44	12	0.000488	0.00183
RAC1 signaling pathway	54	12	0.000488	0.00183
Regulation of RhoA activity	47	11	0.000977	0.00297
Notch-mediated HES/HEY network	48	11	0.000977	0.00297
CXCR4-mediated signaling events	89	11	0.000977	0.00297
Validated transcriptional targets of deltaNp63 isoforms	47	11	0.000977	0.00297
Downstream signaling in naïve CD8+ T cells	68	11	0.000977	0.00297
IL2-mediated signaling events	56	11	0.000977	0.00297
FoxO family signaling	50	11	0.000977	0.00297
FOXA2 and FOXA3 transcription factor networks	46	11	0.000977	0.00297
Signaling events mediated by HDAC Class I	57	11	0.000977	0.00297
Endothelins	65	10	0.00195	0.00483
HIV-1 Nef: Negative effector of Fas and TNF-alpha	35	10	0.00195	0.00483
IL8- and CXCR1-mediated signaling events	28	10	0.00195	0.00483
Thromboxane A2 receptor signaling	57	10	0.00195	0.00483
Arf6 trafficking events	50	10	0.00195	0.00483
Role of Calcineurin-dependent NFAT signaling in lymphocytes	57	10	0.00195	0.00483
Calcineurin-regulated NFAT-dependent transcription in lymphocytes	50	10	0.00195	0.00483
CDC42 signaling events	72	10	0.00195	0.00483
Signaling events mediated by PTP1B	53	10	0.00195	0.00483
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	84	10	0.00195	0.00483
LKB1 signaling events	45	10	0.00195	0.00483
Regulation of nuclear beta catenin signaling and target gene transcription	81	11	0.00384	0.00803
Regulation of RAC1 activity	39	9	0.00391	0.00803
LPA receptor mediated events	63	9	0.00391	0.00803
Signaling events mediated by PRL	24	9	0.00391	0.00803
p75(NTR)-mediated signaling	71	9	0.00391	0.00803
Ceramide signaling pathway	48	9	0.00391	0.00803
IL8- and CXCR2-mediated signaling events	34	9	0.00391	0.00803
Hedgehog signaling events mediated by Gli proteins	49	9	0.00391	0.00803
FGF signaling pathway	50	9	0.00391	0.00803
Neurotrophic factor-mediated Trk receptor signaling	65	9	0.00391	0.00803
IL6-mediated signaling events	48	9	0.00391	0.00803
Class I PI3K signaling events	45	9	0.00391	0.00803
C-MYC pathway	25	8	0.00781	0.0137
FAS (CD95) signaling pathway	35	8	0.00781	0.0137

Aurora A signaling	31	8	0.00781	0.0137
PAR1-mediated thrombin signaling events	44	8	0.00781	0.0137
IL12-mediated signaling events	64	8	0.00781	0.0137
Signaling events mediated by Stem cell factor receptor (c-Kit)	53	8	0.00781	0.0137
EPO signaling pathway	34	8	0.00781	0.0137
RXR and RAR heterodimerization with other nuclear receptor	26	8	0.00781	0.0137
FOXA1 transcription factor network	45	8	0.00781	0.0137
Insulin Pathway	48	8	0.00781	0.0137
Syndecan-4-mediated signaling events	32	8	0.00781	0.0137
Signaling events mediated by VEGFR1 and VEGFR2	72	8	0.00781	0.0137

Tabla S5 Enriquecimiento de vías según PID basados el conjunto de transcritos sobre expresados. El tamaño del conjunto se refiere al número de transcritos que tienen un ID Uniprot en la vía PID correspondiente según el sitio ConsensusPathDB. El número de candidatos se refiere al número de proteínas que forman parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Pathway name	Set size	Candidates contained	P-value	Q-value
Validated transcriptional targets of deltaNp63 isoforms	47	17 (36.2%)	9.63E-06	0.00177
Validated transcriptional targets of TAp63 isoforms	55	16 (29.6%)	0.000268	0.0247
E-cadherin signaling in the nascent adherens junction	39	12 (31.6%)	0.000823	0.0406
ErbB receptor signaling network	16	7 (43.8%)	0.0012	0.0406
a6b1 and a6b4 Integrin signaling	46	13 (28.9%)	0.00129	0.0406
Posttranslational regulation of adherens junction stability and disassembly	57	15 (26.8%)	0.00132	0.0406
Endogenous TLR signaling	26	9 (34.6%)	0.00179	0.0469
Direct p53 effectors	144	28 (19.9%)	0.00277	0.0637
Calcineurin-regulated NFAT-dependent transcription in lymphocytes	50	13 (26.0%)	0.00362	0.0739
Nectin adhesion pathway	32	9 (29.0%)	0.00678	0.113
Syndecan-2-mediated signaling events	42	11 (26.2%)	0.00678	0.113
Validated transcriptional targets of AP1 family members Fra1 and Fra2	37	10 (27.0%)	0.00764	0.117
Regulation of nuclear beta catenin signaling and target gene transcription	81	17 (21.2%)	0.00894	0.118
Beta5 beta6 beta7 and beta8 integrin cell surface interactions	17	6 (35.3%)	0.00944	0.118
Cellular roles of Anthrax toxin	22	7 (31.8%)	0.00964	0.118

Tabla S6 Enriquecimiento de vías según PID basados el conjunto de transcritos sub expresados. El tamaño del conjunto se refiere al número de transcritos que tienen un ID Uniprot en la vía PID

correspondiente según el sitio ConsensusPathDB. El número de candidatos se refiere al número de proteínas que forman parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Over-expressed transcripts	Folds (log2)		Under-expressed transcripts	Folds (log2)
AIFM2_HUMAN	3.20816		COIA1_HUMAN	-6.30929
KAT2A_HUMAN	2.12567		S10A2_HUMAN	-7.42878
MSH2_HUMAN	2.86183		VDR_HUMAN	-1.50616
LIF_HUMAN	4.00723		PML_HUMAN	-0.824712
DUS5_HUMAN	1.86765		MMP2_HUMAN	-8.4294
BBC3_HUMAN	1.36481		RIR2B_HUMAN	-1.16613
JMY_HUMAN	1.20239		PLK3_HUMAN	-0.634628
PCNA_HUMAN	1.92967		P63_HUMAN	-9.87507
SP1_HUMAN	1.61246		CSPG2_HUMAN	-4.44
BCL2_HUMAN	3.49745		SNAI2_HUMAN	-3.76971
TSC2_HUMAN	0.85307		SPB5_HUMAN	-2.50339
FOXA1_HUMAN	4.50315		ASC_HUMAN	-8.65988
GDF15_HUMAN	2.69952		Z385A_HUMAN	-1.76185
EDN2_HUMAN	3.10815		1433S_HUMAN	-4.25544
DDB2_HUMAN	2.26436		PERP_HUMAN	-3.34831
XPO2_HUMAN	1.21701		DKK1_HUMAN	-3.13909
E2F1_HUMAN	5.13942		MDM2_HUMAN	-1.01875
ASPP2_HUMAN	0.965443		CD82_HUMAN	-4.86454
APAF_HUMAN	1.06935		R144B_HUMAN	-3.6513
IBP3_HUMAN	6.04069		BTG2_HUMAN	-4.20738
E2F2_HUMAN	6.42348		BKRB2_HUMAN	-2.06574
			TIGAR_HUMAN	-2.66362
			GA45A_HUMAN	-3.12129
			TGFA_HUMAN	-2.43847
			CAV1_HUMAN	-5.35133
			BAK_HUMAN	-2.04026
			IRF5_HUMAN	-3.13589
			T5311_HUMAN	-6.09755

Tabla S7: Lista de transcritos miembros de la red deregulación "Direct effectors of p53" sobre y sub expresados.

pathway name	set size	candidates contained	p-value	q-value
Steroid biosynthesis	17	10 (58.8%)	3.71E-06	0.000775
Tight junction	132	33 (25.0%)	1.17E-05	0.00122
Endocytosis	202	44 (21.8%)	2.07E-05	0.00144
Adherens junction	73	20 (27.4%)	0.00016	0.00836
alpha-Linolenic acid metabolism	21	9 (42.9%)	0.000287	0.0102
Axon guidance	129	29 (22.5%)	0.000292	0.0102
Cell adhesion molecules (CAMs)	133	29 (21.8%)	0.000502	0.0143
Arachidonic acid metabolism	63	17 (27.0%)	0.00059	0.0143
Salmonella infection	86	21 (24.4%)	0.000618	0.0143
VEGF signaling pathway	76	19 (25.0%)	0.000811	0.0161
Amoebiasis	108	24 (22.6%)	0.000847	0.0161
Fc gamma R-mediated phagocytosis	97	22 (23.2%)	0.000999	0.0168
Pathogenic Escherichia coli infection	55	15 (27.3%)	0.00108	0.0168
Regulation of actin cytoskeleton	212	40 (18.9%)	0.00112	0.0168
Terpenoid backbone biosynthesis	21	8 (38.1%)	0.00158	0.0221
Notch signaling pathway	47	13 (27.7%)	0.00199	0.026
Bladder cancer	42	12 (28.6%)	0.00217	0.0267
Phosphatidylinositol signaling system	80	18 (22.5%)	0.00387	0.045
Linoleic acid metabolism	29	9 (31.0%)	0.00416	0.0458
Focal adhesion	200	36 (18.0%)	0.0044	0.0459
Bacterial invasion of epithelial cells	70	16 (22.9%)	0.00534	0.0531
Glioma	65	15 (23.1%)	0.00626	0.0586
Leukocyte transendothelial migration	116	23 (19.8%)	0.00644	0.0586
Biosynthesis of unsaturated fatty acids	21	7 (33.3%)	0.00728	0.0634
Fc epsilon RI signaling pathway	81	17 (21.5%)	0.00787	0.0658

Tabla S8 Enriquecimiento de vías basados en KEGG utilizando los transcritos sub expresados. El tamaño del conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la base de datos KEGG correspondiente según las vías almacenadas en el sitio ConsensusPathDB. El número de candidatos contenidos refiere a la cantidad de proteínas que son parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento

Gene Ontology term	Set size	Candidates contained	P-value	Q-value
GO:0043933 macromolecular complex subunit organization	1176	68 (5.8%)	1.74E-21	3.45E-19
GO:0071842 cellular component organization at cellular level	3203	116 (3.6%)	3.25E-20	3.24E-18
GO:0022607 cellular component assembly	1520	72 (4.7%)	7.10E-18	4.71E-16
GO:0007018 microtubule-based movement	144	18 (12.5%)	1.27E-11	6.31E-10
GO:0019058 viral infectious cycle	235	21 (9.0%)	1.43E-10	5.71E-09
GO:0046907 intracellular transport	1101	45 (4.1%)	3.43E-09	1.14E-07
GO:0072594 establishment of protein localization to organelle	206	18 (8.8%)	4.35E-09	1.24E-07
GO:0035966 response to topologically incorrect protein	139	15 (10.8%)	5.31E-09	1.29E-07
GO:0045104 intermediate filament cytoskeleton organization	27	8 (29.6%)	5.86E-09	1.29E-07
GO:0022411 cellular component disassembly	291	21 (7.2%)	7.09E-09	1.41E-07
GO:0019080 viral genome expression	155	15 (9.7%)	2.17E-08	3.59E-07
GO:0019083 viral transcription	155	15 (9.7%)	2.17E-08	3.59E-07
GO:0015031 protein transport	1216	44 (3.6%)	1.82E-07	2.79E-06
GO:0044260 cellular macromolecule metabolic process	6676	144 (2.2%)	1.35E-06	1.92E-05
GO:0008104 protein localization	1523	48 (3.2%)	2.59E-06	3.44E-05
GO:0070271 protein complex biogenesis	755	30 (4.0%)	3.24E-06	4.03E-05
GO:0019047 provirus integration	8	4 (50.0%)	4.01E-06	4.43E-05
GO:0030069 lysogeny	8	4 (50.0%)	4.01E-06	4.43E-05
GO:0070727 cellular macromolecule localization	895	33 (3.7%)	4.92E-06	5.15E-05
GO:0071843 cellular component biogenesis at cellular level	245	15 (6.1%)	8.33E-06	8.29E-05
GO:0019538 protein metabolic process	3827	88 (2.3%)	6.38E-05	0.000605
GO:0044248 cellular catabolic process	1683	47 (2.8%)	7.48E-05	0.000677
GO:0016052 carbohydrate catabolic process	170	11 (6.5%)	8.20E-05	0.000709
GO:0044282 small molecule catabolic process	319	15 (4.7%)	0.000172	0.00142
GO:0048731 system development	3281	75 (2.3%)	0.000347	0.00276
GO:0006139 nucleobase-containing compound metabolic process	5195	108 (2.1%)	0.000437	0.00335
GO:0034641 cellular nitrogen compound metabolic process	5621	115 (2.0%)	0.000478	0.00353
GO:0019059 initiation of viral infection	25	4 (16.0%)	0.000586	0.00403
GO:0007596 blood coagulation	518	19 (3.7%)	0.000588	0.00403
GO:0007599 hemostasis	522	19 (3.6%)	0.000645	0.00428
GO:0009057 macromolecule catabolic process	844	26 (3.1%)	0.000859	0.00552
GO:0016192 vesicle-mediated transport	945	28 (3.0%)	0.00101	0.00625
GO:0000086 G2/M transition of mitotic cell cycle	145	8 (5.5%)	0.00212	0.0128
GO:0022614 membrane to membrane docking	5	2 (40.0%)	0.00239	0.0138
GO:0030154 cell differentiation	2562	58 (2.3%)	0.00243	0.0138
GO:0030168 platelet activation	222	10 (4.5%)	0.0028	0.0155

GO:0000226	microtubule cytoskeleton organization	272	11 (4.0%)	0.00402	0.0216
GO:0012501	programmed cell death	1551	38 (2.5%)	0.0042	0.022
GO:0051235	maintenance of location	173	8 (4.6%)	0.00619	0.0316
GO:0048519	negative regulation of biological process	2999	63 (2.1%)	0.00862	0.0429
GO:0009611	response to wounding	1105	28 (2.5%)	0.00888	0.0431

Tabla S10 Enriquecimiento de ontología genética nivel 3 categoría de “procesos biológicos” a partir de las fosfoproteínas identificadas. El tamaño de conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la categoría GO correspondiente en el sitio ConsensusPathDB. El número de candidatos contenidos se refiere a la cantidad de proteínas identificadas en este estudio que aparecen como parte de la categoría GO. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representan los valores de p corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Pathway name	Set size	Candidates contained	P-value	Q-value
Signaling events mediated by HDAC Class III	40	17 (43.6%)	5.03E-19	2.57E-17
p38 signaling mediated by MAPKAP kinases	22	7 (33.3%)	1.45E-07	3.38E-06
LKB1 signaling events	45	9 (20.9%)	1.99E-07	3.38E-06
Validated targets of C-MYC transcriptional activation	87	11 (13.6%)	8.79E-07	1.12E-05
Class I PI3K signaling events mediated by Akt	36	7 (20.0%)	6.55E-06	5.64E-05
mTOR signaling pathway	65	9 (14.1%)	6.64E-06	5.64E-05
Insulin-mediated glucose transport	31	6 (20.0%)	3.07E-05	0.000224
FoxO family signaling	50	7 (14.3%)	6.52E-05	0.000416
Trk receptor signaling mediated by PI3K and PLC-gamma	37	6 (16.7%)	9.07E-05	0.000514
Role of Calcineurin-dependent NFAT signaling in lymphocytes	57	7 (13.0%)	0.000123	0.000576
Signaling events mediated by HDAC Class II	39	6 (15.8%)	0.000124	0.000576
a6b1 and a6b4 Integrin signaling	46	6 (13.3%)	0.000324	0.00138
Regulation of nuclear beta catenin signaling and target gene transcription	81	7 (8.8%)	0.0014	0.00548
Aurora B signaling	43	5 (11.9%)	0.00174	0.00635
Integrin-linked kinase signaling	46	5 (11.1%)	0.00238	0.00802
HIF-1-alpha transcription factor network	66	6 (9.1%)	0.00252	0.00802
Regulation of Telomerase	73	6 (8.8%)	0.00293	0.00878
Caspase cascade in apoptosis	53	5 (9.4%)	0.00489	0.0138
PDGFR-beta signaling pathway	130	8 (6.3%)	0.00518	0.0139
Aurora C signaling	7	2 (33.3%)	0.0062	0.0158
ATR signaling pathway	37	4 (10.8%)	0.0072	0.0175

Tabla S11 Enriquecimiento de vías según PID basados el conjunto de fosfoproteínas identificadas. El tamaño del conjunto se refiere al número de transcritos que tienen un ID Uniprot en la vía PID correspondiente según el sitio ConsensusPathDB. El número de candidatos se refiere al número de proteínas que forman parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento.

Pathway name	Set size	Candidates contained	P-value	Q-value
Pathogenic Escherichia coli infection	55	16 (29.1%)	1.16E-14	7.33E-13
Systemic lupus erythematosus	138	19 (14.0%)	5.10E-11	1.61E-09
Ribosome	91	14 (15.6%)	4.66E-09	9.79E-08
Gap junction	89	13 (14.6%)	3.68E-08	5.80E-07
Spliceosome	127	15 (11.8%)	6.07E-08	7.65E-07
Glycolysis / Gluconeogenesis	65	11 (16.9%)	8.96E-08	9.41E-07
Alcoholism	179	17 (9.5%)	2.00E-07	1.80E-06
Protein processing in endoplasmic reticulum	166	16 (9.6%)	3.79E-07	2.99E-06
Phagosome	154	15 (9.9%)	6.56E-07	4.59E-06
Antigen processing and presentation	76	10 (13.2%)	3.85E-06	2.43E-05
Legionellosis	55	8 (14.5%)	1.73E-05	9.92E-05
Epstein-Barr virus infection	203	14 (7.0%)	8.43E-05	0.000443
Cell cycle	124	10 (8.1%)	0.000275	0.00133
Pyruvate metabolism	41	5 (12.2%)	0.00158	0.00712
Propanoate metabolism	32	4 (12.5%)	0.0043	0.0181
Phenylalanine metabolism	18	3 (16.7%)	0.00592	0.0233

Tabla S12 Enriquecimiento de vías basados en KEGG utilizando las fosfoproteínas identificadas. El tamaño del conjunto se refiere al número de entidades que tienen un ID Uniprot como se indica en la base de datos KEGG correspondiente según las vías almacenadas en el sitio ConsensusPathDB. El número de candidatos contenidos refiere a la cantidad de proteínas que son parte de la red extendida y aparecen como parte de la vía. El P-valor se calcula de acuerdo a una prueba hipergeométrica, el Q-valor representa los valores de P corregido para pruebas múltiples utilizando el método de tasas de falso descubrimiento

Swisprot ID	Rep_1	Mass_1	Score_1	Matches_1	Rep_2	Mass_2	Score_2	Matches_2	Rep_3	Mass_3	Score_3	Matches_3
1433B									1433B	28179	78	5(2)
1433E					1433E	29326	47	2(1)	1433E	29326	128	3(3)
1433F					1433F	28372	49	2(2)				
1433G					1433G	28456	47	3(1)	1433G	28456	78	4(2)
1433S					1433S	27871	49	2(2)				
1433Z					1433Z	27899	136	5(4)	1433Z	27899	140	7(4)
4F2									4F2	68180	52	1(1)
ACDSB	ACDSB	47797	34	1(1)								
ACOXL					ACOXL	62383	37	1(1)	ACOXL	62383	44	2(1)
ACTA	ACTA	42381	37	1(1)								
ACTB					ACTB	42052	76	3(2)	ACTB	42052	401	13(11)
ACTBL					ACTBL	42318	63	4(1)	ACTBL	42318	161	4(4)
ACTBM					ACTBM	42331	60	2(1)				
ACTC									ACTC	42334	223	9(6)
ACTN1									ACTN1	103563	50	1(1)
AHnk					AHnk	629213	83	3(2)				
AHnk2									AHnk2	617383	47	1(1)
AINX	AINX	55391	42	1(1)								
AKTS1									AKTS1	27595	35	1(1)
ALBU									ALBU	71317	73	3(1)
ALDOA					ALDOA	39851	77	2(1)	ALDOA	39851	187	7(5)
ANO8												
ANXA2					ANXA2	38808	54	1(1)	ANXA2	38808	108	4(4)
ANXA5	ANXA5	35971	59	2(1)								
AP3D1					AP3D1	131159	36	1(1)				
ASAH1					ASAH1	45087	33	1(1)				
ASAP2					ASAP2	112835	47	1(1)				
ASSY									ASSY	46786	44	1(1)
ATR					ATR	304764	38	2(1)				

BCL7B	BCL7B	22195	37	1(1)								
BFSP1					BFSP1	74784	44	1(1)				
BLK									BLK	58126	51	1(1)
BZW1									BZW1	48184	44	2(1)
BZW2					BZW2	48360	47	1(1)	BZW2	48360	54	2(2)
C1QBP	C1QBP	31742	92	2(2)								
CALR	CALR								CALR	48283	35	1(1)
CALU	CALU	37198	36	3(1)								
CALX									CALX	67982	49	1(1)
CAPS2					CAPS2	148895	34	1(1)				
CC110	CC110	96726	49	2(1)								
CDK1									CDK1	34131	51	1(1)
CH60												
CH60					CH60	61187	103	3(3)	CH60	61187	173	9(6)
CHM2A									CHM2A	25088	37	1(1)
CHMP5									CHMP5	24612	35	1(1)
CHRC1									CHRC1	14758	36	1(1)
CLIC1									CLIC1	27248	41	1(1)
CNTLN									CNTLN	162131	37	1(1)
CNTP4					CNTP4	147178	37	1(1)	CNTP4	147178	37	1(1)
CO039	CO039	112086	35	2(1)								
CO6A1	CO6A1	108529	39	1(1)								
CTNA1									CTNA1	100693	38	1(1)
DDX21									DDX21	87804	88	4(3)
DIRA2	DIRA2	22813	52	1(1)								
DNJA2	DNJA2	46344	47	1(1)								
DPF3	DPF3	44254	37	1(1)								
DX39A	DX39A	49611	54	4(2)								
DYH2	DYH2	510796	39	1(1)								
EBP2									EBP2	34887	36	1(1)

EEPD1	EEPD1	62820	43	1(1)								
EF1A1					EF1A1	50451	59	1(1)	EF1A1	50451	138	7(4)
EF1B					EF1B	24919	106	3(3)	EF1B	24919	139	2(2)
EF1D					EF1D	31217	39	1(1)	EF1D	31217	283	6(6)
EFTU					EFTU	49852	34	1(1)	EFTU	49852	44	4(1)
EIF3B					EIF3B	92823	59	1(1)	EIF3B	92823	224	10(8)
EIF3J					EIF3J	29159	60	1(1)	EIF3J	29159	45	2(1)
ENOA	ENOA	47481	41	1(1)	ENOA	47481	66	2(1)	ENOA	47481	135	7(6)
ENPL	ENPL	92696	71	7(2)								
EZRI									EZRI	69484	50	4(2)
F10A1					F10A1	41477	65	3(2)	F10A1	41477	200	8(4)
F16A1									F16A1	117744	36	2(1)
G3BP1					G3BP1	52189	49	4(3)	G3BP1	52189	326	11(10)
G3BP2					G3BP2	54145	47	3(2)	G3BP2	54145	60	3(2)
G3P					G3P	36201	64	3(2)	G3P	36201	223	5(3)
G3PT	G3PT	44815	62	2(2)								
GBF1					GBF1	208367	45	2(1)				
GBRA5	GBRA5	52398	33	1(0)								
GFAP					GFAP	49907	56	2(1)				
GIT2	GIT2	85117	51	6(3)								
GLU2B					GLU2B	60357	56	1(1)	GLU2B	60357	71	3(1)
GOGA1	GOGA1	88244	42	1(1)								
GP128					GP128	89992	35	1(1)				
GRP75									GRP75	73920	46	1(1)
GRP78					GRP78	72402	142	7(4)	GRP78	72402	235	10(10)
H33									H33	15376	108	11(5)
H3C					H3C	15318	45	4(2)	H3C	15318	47	2(1)
H4	H4	11367	45	2(1)								
H90B2					H90B2	44492	86	4(3)	H90B2	44492	245	12(8)
H90B3									H90B3	68624	346	14(11)
H90B4					H90B4	58855	64	5(3)	H90B4	58855	45	3(1)

HBB					HBB	16102	46	2(1)				
HDGF					HDGF	26886	68	4(3)	HDGF	26886	75	4(2)
HIP1R									HIP1R	119999	36	1(1)
HMGA1									HMGA 1	11669	64	1(1)
HMMR					HMMR	84448	34	1(1)				
HNRCL	HNRCL	32180	75	1(1)								
HNRPC									HNRPC	33707	106	3(3)
HNRPD									HNRPD	38581	38	2(1)
HNRPK					HNRPK	51230	53	2(1)	HNRPK	51230	133	4(4)
HNRPU					HNRPU	91269	116	6(5)	HNRPU	91269	253	13(6)
HPPD									HPPD	45077	34	1(1)
HS71L	HS71L	70730	72	3(1)								
HS902					HS902	39454	76	6(2)				
HS90A					HS90A	85006	254	8(7)	HS90A	85006	555	22(19)
HS90B					HS90B	83554	328	18(11)	HS90B	83554	942	36(27)
HSP71					HSP71	70294	92	4(2)	HSP71	70294	240	12(9)
HSP76					HSP76	71440	40	1(1)	HSP76	71440	195	6(6)
HSP7C					HSP7C	71082	34	1(1)	HSP7C	71082	330	13(10)
HSPB1					HSPB1	22826	59	2(2)	HSPB1	22826	79	3(2)
HTSF1					HTSF1	86371	39	1(1)				
HUWE1									HUWE 1	485523	35	2(1)
IF1AX					IF1AX	16564	41	2(1)				
IF1AY									IF1AY	16546	78	2(1)
IF4A1					IF4A1	46353	67	3(1)	IF4A1	46353	73	4(2)
IF4B									IF4B	69167	96	3(1)
ILF2									ILF2	43263	80	2(2)
JKIP1									JKIP1	73506	37	1(1)
K1C10					K1C10	59020	111	5(3)	K1C10	59020	399	16(12)
K1C12					K1C12	53592	39	2(1)				

K1C13									K1C13	49900	64	2(1)
K1C14					K1C14	51872	88	3(2)	K1C14	51872	177	6(6)
K1C15	K1C15	49409	53	3(2)								
K1C16									K1C16	51578	187	7(7)
K1C17					K1C17	48361	35	3(1)	K1C17	48361	103	5(3)
K1C18	K1C18	48029	80	3(2)								
K1C19	K1C19	44079	89	4(3)								
K1C25					K1C25	49858	59	2(1)				
K1C27					K1C27	50419	72	2(1)				
K1C28	K1C28	51163	42	3(1)								
K1C9					K1C9	62255	203	7(4)	K1C9	62255	60	2(1)
K1H1					K1H1	48633	45	2(1)				
K22E					K22E	65678	87	5(2)	K22E	65678	104	7(3)
K22O					K22O	66370	44	4(1)	K22O	66370	94	4(4)
K2C1					K2C1	66170	350	13(11)	K2C1	66170	282	11(9)
K2C1B					K2C1B	62149	40	2(1)				
K2C4					K2C4	57649	93	3(2)	K2C4	57649	65	2(1)
K2C5					K2C5	62568	33	1(1)	K2C5	62568	175	10(8)
K2C6A					K2C6A	60293	44	3(1)				
K2C6B					K2C6B	60315	67	3(2)	K2C6B	60315	88	4(2)
K2C6C									K2C6C	60273	253	15(11)
K2C7					K2C7	51411	66	2(1)				
K2C73					K2C73	59457	36	2(1)				
K2C75					K2C75	59809	57	3(1)				
K2C79					K2C79	58085	60	2(1)	K2C79	58085	118	6(5)
K2C8					K2C8	53671	86	3(1)	K2C8	53671	63	4(2)
K2C80	K2C80	51007	78	2(1)								
KCNG3					KCNG3	50359	34	1(1)				
KI20B					KI20B	211868	36	1(1)				
KIF1A					KIF1A	192541	36	2(1)				
KPYM									KPYM	58470	61	1(1)

KRT35					KRT35	51640	45	3(1)				
KRT38	KRT38	52044	56	2(1)								
KT222	KT222	34158	68	2(1)								
LA									LA	46979	51	1(1)
LC7L2									LC7L2	46942	48	1(1)
LDH6A									LDH6A	36826	91	2(1)
LDHA									LDHA	36950	110	5(3)
LDHB									LDHB	36900	125	5(5)
LEG3									LEG3	26152	39	1(1)
LIMA1					LIMA1	85630	52	3(1)	LIMA1	85630	47	2(1)
LMNA					LMNA	74380	57	1(1)	LMNA	74380	215	7(6)
LMNB1									LMNB1	66653	39	3(2)
LMNB2									LMNB2	67762	39	2(2)
MAGD4	MAGD4	81555	43	1(1)								
MATR3					MATR3	95078	78	5(3)	MATR3	95078	38	4(1)
MCM2									MCM2	102516	40	1(1)
MCM3					MCM3	91551	46	1(1)	MCM3	91551	117	4(3)
MDHM	MDHM	35937	35	1(1)								
MIF									MIF	12639	54	1(1)
MMP3									MMP3	54228	36	2(1)
MOES									MOES	67892	50	5(2)
MYH15					MYH15	225904	43	1(1)				
MYO1A	MYO1A	119238	39	1(1)								
NACA					NACA	23370	214	4(4)	NACA	23370	320	6(6)
NACA2					NACA2	23209	104	1(1)	NACA2	23209	130	2(2)
NACAD									NACAD	161971	66	1(1)
NFM												
NOP2									NOP2	89589	42	1(1)
NPM					NPM	32726	65	6(3)	NPM	32726	597	18(12)
OVCH1					OVCH1	127095	36	1(1)				
PA2G4					PA2G4	44101	35	2(1)				

PARK7									PARK7	20050	36	2(1)
PCBP1									PCBP1	37987	32	1(0)
PCNP									PCNP	18913	37	2(1)
PDIA1					PDIA1	57480	649	27(20)	PDIA1	57480	884	31(28)
PDIA4					PDIA4	73229	37	1(1)	PDIA4	73229	452	22(16)
PDIA6	PDIA6	48490	62	2(1)								
PERI	PERI	53651	90	3(1)								
PGK1									PGK1	44985	78	4(2)
PGK2									PGK2	45166	79	1(1)
PHB					PHB	29843	39	2(1)				
PHLA1	PHLA1	45501	35	1(1)								
PKHG1	PKHG1	156541	30	1(1)								
PKHG5									PKHG5	118974	36	1(1)
PLCL2									PLCL2	127268	36	1(1)
POTEE					POTEE	122882	90	2(1)	POTEE	122882	256	6(5)
POTEI									POTEI	122858	126	5(3)
PPIA					PPIA	18229	50	5(2)	PPIA	18229	210	8(6)
PPIB					PPIB	23785	80	2(1)	PPIB	23785	271	5(2)
PRDM8									PRDM8	72588	34	1(1)
PRDX1					PRDX1	22324	117	5(4)	PRDX1	22324	137	9(6)
PRDX3					PRDX3	28017	58	1(1)	PRDX3	28017	44	3(1)
PRDX4									PRDX4	30749	39	2(1)
PRDX6					PRDX6	25133	42	1(1)	PRDX6	25133	99	4(3)
PROF1					PROF1	15216	75	3(2)	PROF1	15216	163	5(5)
PSA3					PSA3	28643	78	2(1)	PSA3	28643	172	6(6)
PSIP1									PSIP1	60181	70	3(1)
PVRL4									PVRL4	55933	34	1(1)
RAB1A									RAB1A	22891	52	1(1)
RAB28					RAB28	25054	36	1(1)	RAB28	25054	39	1(1)
RAB35									RAB35	23296	52	1(1)
RALY					RALY	32501	66	4(1)	RALY	32501	111	5(5)

RALYL					RALYL	32425	66	1(1)	RALYL	32425	71	1(1)
RAN					RAN	24579	87	4(2)	RAN	24579	39	1(1)
RBBP4	RBBP4	47911	44	4(1)								
RBBP5	RBBP5	59800	33	1(0)								
RBBP7	RBBP7	48132	88	3(2)								
RL11									RL11	20468	86	2(1)
RL22					RL22	14835	181	4(4)	RL22	14835	157	2(2)
RL7									RL7	29264	52	2(1)
RL8					RL8	28235	35	1(1)				
RLA0					RLA0	34423	111	4(3)	RLA0	34423	258	6(6)
RLAOL					RLAOL	34514	48	1(1)				
RLA1					RLA1	11621	41	1(1)				
RLA2					RLA2	11658	63	2(1)	RLA2	11658	206	5(5)
ROA3									ROA3	39799	34	1(1)
RS14					RS14	16434	35	1(1)				
RS15A									RS15A	14944	36	1(1)
RS16					RS16	16549	49	1(1)				
RS28									RS28	7893	35	1(1)
RS3A									RS3A	30154	43	4(1)
RS4X					RS4X	29807	61	1(1)				
RS8					RS8	24475	85	2(2)				
RU17					RU17	51583	41	1(1)				
SAE2									SAE2	71749	56	3(2)
SAFB1									SAFB1	103036	40	2(1)
SARNP									SARNP	23713	52	3(2)
sep-02					sep-02	41689	59	1(1)	sep-02	41689	149	2(2)
SF3A1					SF3A1	88888	75	7(4)	SF3A1	88888	238	15(12)
SF3B2					SF3B2	100279	139	3(2)	SF3B2	100279	36	1(1)
SNED1					SNED1	158206	33	1(1)	SNED1	158206	36	1(1)
SPTN4	SPTN4	290005	42	1(1)								
SRSF1					SRSF1	27842	57	1(1)	SRSF1	27842	109	4(2)

SRSF3					SRSF3	19546	91	3(3)	SRSF3	19546	53	3(1)
ST134									ST134	27561	53	4(1)
TAGL2									TAGL2	22548	56	1(1)
TBA1A					TBA1A	50788	59	1(1)	TBA1A	50788	43	1(1)
TBA1B					TBA1B	50804	368	13(11)	TBA1B	50804	586	22(18)
TBA1C	TBA1C	50548	363	17(10)								
TBA3E					TBA3E	50568	105	6(4)				
TBA4A					TBA4A	50634	218	9(7)				
TBA4B									TBA4B	27819	202	4(4)
TBA8					TBA8	50746	150	7(5)	TBA8	50746	262	10(7)
TBB1					TBB1	50865	148	5(4)	TBB1	50865	124	7(4)
TBB2A					TBB2A	50274	57	1(1)	TBB2A	50274	599	21(14)
TBB2C	TBB2C	50255	98	5(3)								
TBB3					TBB3	50856	90	2(2)	TBB3	50856	105	4(3)
TBB4B					TBB4B	50255	566	19(14)	TBB4B	50255	722	25(17)
TBB5					TBB5	50095	777	20(24)	TBB5	50095	936	30(21)
TBB8	TBB8	50257	103	5(2)								
TCP4									TCP4	14386	139	4(1)
TENS1					TENS1	186499	38	1(1)				
TPIS					TPIS	31057	75	2(1)	TPIS	31057	68	4(2)
TRAP1					TRAP1	80345	96	2(1)	TRAP1	80345	54	1(1)
TSNA1	TSNA1	56655	40	1(1)								
UN13C					UN13C	252693	39	2(1)				
UTS2									UTS2	14515	39	1(1)
VIME					VIME	53676	100	1(1)	VIME	53676	344	10(8)
WNT7A					WNT7A	40405	34	1(1)				
XRCC6									XRCC6	70084	53	4(1)
ZN275	ZN275	49780	34	1(1)								
ZN524					ZN524	29262	44	1(1)				

Tabla S9 Relación de fosfoproteínas de acuerdo a la replica donde fueron identificadas.