

RESEARCH

Relación de genes con la aparición del papiloma como fenotipo patológico.

Eva M. Ayala Fernández, Raúl A. Castro Valderas, María J. Hidalgo Rodríguez, Francisco J. Rodríguez-cordoba Lucena and María Vida Montañez*

*Correspondence:
mariavida262001@uma.es
ETSI Informática, Universidad de
Málaga, Málaga, España
Full list of author information is
available at the end of the article

Abstract

Keywords: VPH; papiloma; papila; tumor; genes; cáncer

1 Introducción

Un papiloma es un tumor epitelial benigno que crece de manera exofítica [?], es decir, proyectándose hacia afuera en forma de proyecciones y de manera no agresiva ni propagándose por todo el cuerpo. En este contexto, "papila" se refiere a la proyección creada por el tumor, no a un tumor en una papila ya existente. Cuando se utiliza sin contexto específico, con frecuencia se refiere a infecciones causadas por el virus del papiloma humano (VPH). Existen casi 200 tipos distintos de VPH [?], y muchos de ellos son carcinogénicos. Sin embargo, también existen otras condiciones que pueden causar papilomas, así como muchos casos en los que la causa no se conoce. Las infecciones por el VPH de riesgo alto en ocasiones causan cáncer en las partes del cuerpo en las que el VPH infecta a las células. Por ejemplo, cáncer de cuello uterino, cáncer vulvar, cáncer vaginal, cáncer de pene, cáncer anal y cánceres orofaríngeos positivos para el VPH. La carga de los cánceres relacionados con el virus del papiloma humano causa cerca del 5% [?] de todos los cánceres en el mundo, se calcula que 570.000 mujeres y 60.000 hombres tienen un cáncer relacionado con el VPH cada año [?]. El cáncer de cuello uterino es el más frecuente de todos los causados por el VPH, debido a que este es la causa de casi todos los cánceres de cuello uterino del mundo [?].

1.1 Información sobre los genes a estudiar

A continuación, veremos una breve información de los genes con mayor grado de interconexión que hemos encontrado al realizar un análisis fenotípico (HPO).

AKT1: La proteína kinasa B (Akt1) tiene un papel fundamental en el crecimiento y la supervivencia celular al transducir señales en la cascada de señalización celular (PI3K)/AKT, la cual genera mensajeros que participan en la regulación de la progresión del ciclo celular, adhesión y migración. La vía PI3K/AKT es una de las que suelen estar afectadas en distintos tipos de cáncer en humanos, como el cáncer de ovario, de mama, y de Cowden. Además, se asocia los niveles altos de fosforilación de la proteína con los peores pronósticos de cáncer [?].

TP53: TP53 (tumor protein p53) es un gen supresor de tumores involucrado en procesos biológicos fundamentales para la estabilidad genética. Las mutaciones en

este gen han sido asociadas con un peor pronóstico para pacientes con carcinoma oral de células escamosas, dando lugar a un cáncer más agresivo al combinarse con el VPH [?].

HRAS: Este gen participa en la regulación de la vía de la proteína quinasa activada por mitógenos (MAPK) y mediada por la proteína quinasa Raf. En los últimos años se han definido una serie de síndromes con mutaciones en genes implicados en esta vía Ras/PAPK, entre ellos el síndrome de Costellos. Este síndrome refleja características cutáneas distintivas, como papillomas [?].

SDHD: Las mutaciones en este gen están asociadas con la formación de tumores, incluyendo el paraganglioma hereditario. La transmisión de la enfermedad ocurre casi exclusivamente a través del alelo paterno, lo que sugiere que este locus puede estar impreso maternalmente. Hay pseudogenes para este gen en los cromosomas 1, 2, 3, 7 y 18. Resultados de empalme alternativos en múltiples variantes de transcripción [?].

SDHB: Las mutaciones en este gen dan como resultado feocromocitoma y paraganglioma. Las alteraciones germinales y variaciones en este gen causan el síndrome de Cowden. Recientemente se ha reconocido el cáncer de endometrio como un componente importante de este síndrome [?].

1.2 Hipótesis del trabajo

Al mapear los genes asociados con el papiloma en una red de interacción proteína-proteína, podemos modelar este fenotipo y buscar grupos de genes que formen conglomerados dentro de la red. Al hacerlo, planteamos la hipótesis de que encontraremos grupos de genes involucrados en procesos subyacentes importantes relacionados con el crecimiento tumoral, y diversos cánceres como el de cuello uterino.

1.3 Objetivos

- 1 Obtener genes relacionados con el papiloma.
- 2 Utilizar estos genes para hacer una network para modelar los procesos subyacentes del papiloma.
- 3 Analizar esta red para encontrar grupos de genes que interactúan entre sí.
- 4 Investigar las funciones de los genes en estos grupos para obtener información sobre los subyacentes mecanismos que conducen al papiloma (aquí podemos ver si están relacionados en las mismas enfermedades).

2 Materiales y métodos

2.1 Materiales

2.2 Métodos

En primer lugar creamos una función que busca genes asociados a un fenotipo específico utilizando la API de **HPO**. La función recibe el parámetro fenotipo, que es el código HPO para el fenotipo de interés. Realizamos una solicitud GET a la API de HPO para obtener información sobre genes asociados al fenotipo. Si la solicitud es exitosa (código de estado 200), extraemos y devolvemos los genes asociados al fenotipo. Si no, imprimimos un mensaje de error y devolvemos una lista vacía.

Utilizamos la función *buscar_genes_por_fenotipo* para obtener los genes asociados al fenotipo "Papilloma" (código HPO: HP:0012740). Si encontramos genes asociados, imprimimos los nombres de los genes. Si no encontramos genes asociados, mostramos un mensaje indicando que no se encontraron genes.

Tras esto construimos una URL con los símbolos de los genes obtenidos anteriormente para realizar una solicitud a la API de StringDB, que proporciona información sobre interacciones de proteínas. Realizamos una solicitud GET a la API de StringDB para obtener una red de interacciones de proteínas. Guardamos la respuesta como una imagen llamada 'img.png'.

Imprimimos y filtramos los resultados del análisis de enriquecimiento por categorías ('Process' y 'KEGG'). Realizamos búsquedas específicas en los resultados de enfermedades y genes utilizando palabras clave y patrones de genes relevantes, respectivamente. Guardamos los resultados de las búsquedas en archivos CSV.

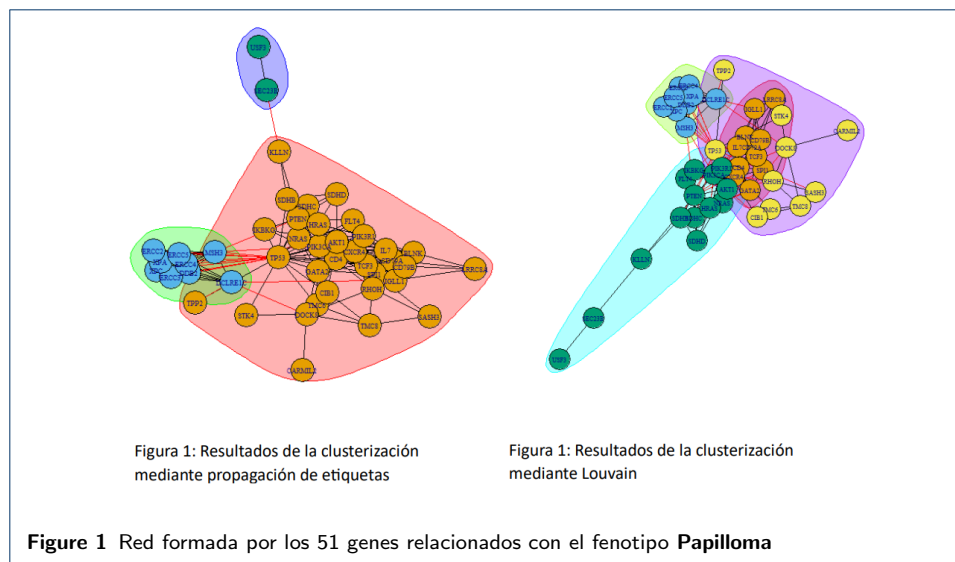
Realizamos una solicitud GET a la API de StringDB para descargar la red de interacciones en formato TSV. Si la descarga es exitosa, guardamos el archivo con el nombre proporcionado por la respuesta o como '*red_descargada.tsv*' si no se obtiene un nombre de archivo. Si hay un error en la descarga, imprimimos un mensaje de error.

Cargamos el archivo TSV descargado. Seleccionamos columnas específicas ('*preferredName_A*' y '*preferredName_B*') y eliminamos duplicados. Guardamos estas columnas en un nuevo archivo de texto llamado '*genes_graph.txt*'.

Nosotros usamos la biblioteca de **igraph** disponible en el lenguaje de programación de R para realizar el análisis y la visualización de la red. En primer lugar, importamos la librería y leímos el fichero que obtuvimos al generar la red con la API de STRINGDB y la guardamos en un grafo. A partir de este, estudiamos las **propiedades del grafo**, si todos los nodos estaban conectados con la función *is_connected()*, si era o no dirigido con la función *is.directed()*, el grado de centralidad que nos informaba del número de conexiones de cada gen con la función *degree()*, la centralidad de cercanía con la que obtuvimos la distancia promedio entre un nodo y todos los demás nodos mediante la aplicación de *closeness()*, y la conectividad que nos indicaba la fortaleza de la conexión del nodo aplicando *edge_density()*.

En segundo lugar, llevamos a cabo la **identificación de comunidades** mediante distintos **algoritmos de clusterización**: método de Girvan-Newman, algoritmo de optimización voraz, propagación de etiquetas y el algoritmo de Louvain.

- 1 El método de **Girvan Newman** detecta comunidades basándose en la centralidad de intermediación de los nodos, en otras palabras, va eliminando gradualmente las aristas más importantes para identificar las comunidades de la red [?].
- 2 El **algoritmo voraz**, busca formar grupos de datos de manera iterativa tomando en cada paso la elección más beneficiosa para fusionar o dividir clusters con el objetivo de maximizar un criterio local [?].
- 3 La **propagación de etiquetas** es un enfoque basado en la difusión de información a través de la red que agrupa los nodos que están fuertemente conectados [?].
- 4 Por último el **algoritmo de Louvain** busca organizar los nodos de una red en comunidades de manera que la modularidad global de la red sea máxima,



o en otras palabras, cómo de bien se dividen los nodos de una red en grupos o comunidades distintas [?].

Para estudiar mejor a qué comunidad pertenece cada nodo, visualizamos el resultado de la aplicación del algoritmo de Louvain mediante la aplicación de **link communities** que nos permitía identificar si alguno de los nodos se incluían en varias comunidades.

En tercer lugar, estudiamos la interacción de nuestros **genes de interés** especificados anteriormente: TP53, HRAS, AKT1, SDHB, SDHD. Para ellos creamos una función que nos devolvía una tabla con los vecinos de cada uno de estos genes para saber si entre ellos estaban relacionados.

En último lugar, tras estudiar la detección de comunidades optamos por realizar un enriquecimiento funcional para la comunidad identificada donde se encuentren nuestros genes de interés. Para ello, hemos guardado los genes que conforman la comunidad donde se encuentran los genes de interés en un archivo y hemos realizado un enriquecimiento funcional de estos genes filtrando para quedarnos con las entradas de categoría HPO [?] (para relacionar estos genes con fenotipos patológicos). Estos datos se guardan en un archivo de extensión csv y hemos consultado las entradas de este buscando por palabras clave (papilloma, genital, renal, carcinoma, ovarian, ovary, ovaries, uterine).

3 Resultados

3.1 Red de interacción entre genes

Se identificaron un total de 51 genes asociados al fenotipo **Papilloma** y obtuvimos una representación visual de la red de interacción entre los genes.

3.2 Análisis de enriquecimiento funcional

Se realizó un análisis de enriquecimiento funcional destacando las categorías **Process** y **KEGG**, obteniendo relaciones entre algunos grupos de genes con problemas del sistema inmune, carcinoma y las adipoquinas.

Gen	Degree_Centrality	Gen	Degree_Centrality
CD4	15	SDHC	5
CD79A	13	NRAS	8
BLNK	11	PTEN	15
SPI1	11	STK4	2
DDB2	9	XPA	9
LRRC8A	4	SDHB	7
FLT4	5	SDHD	5
TCF3	10	TPP2	2
IL7	10	DCLRE1C	10
PIK3CA	11	RHOH	8
MSH3	11	ERCC2	8
TP53	30	CD79B	11
XPC	9	CXCR4	12
ERCC3	9	HRAS	9
ERCC4	9	PIK3R1	12
USF3	1	AKT1	20
TMC8	5	TMC6	4
IGLL1	7	IKBKG	4
CARMIL2	1	DOCK8	7
SEC23B	2	ERCC5	9
GATA2	10		
SASH3	3		

Figure 2 Descripción de la figura.

Los términos y categorías relacionados con enfermedades como **cervix**, **ovarian**, **HPV**, **herpes**, **papillomavirus**, **costellos** y **cowden** fueron encontrados en el análisis de enriquecimiento funcional, indicando posibles vínculos entre estos términos y los genes asociados al fenotipo estudiado.

Se realizaron búsquedas específicas de genes clave como **TP53**, **AKT1**, **SDHB**, **SDHD**, **HRAS** dentro de las categorías significativas. Estos genes podrían tener una importancia particular en relación con el fenotipo de interés, evidenciando su posible relevancia funcional.

3.3 Descarga y procesamiento de la red de interacción

Se descargó la red de interacción en formato TSV. El archivo descargado se procesó para dar como resultado un archivo de texto *genes;graph.txt* formado por dos columnas de nombres de genes, en las que cada fila representa una relación entre dos genes.

3.4 Propiedades de la red y detección de comunidades

En nuestra red vimos que todos los nodos estaban conectados entre sí y que se trataba de un grafo no dirigido.

Obtuvimos una tabla con el **grado de centralidad** de cada gen. Pudimos observar como el gen de interés TP53 es el que presentaba un mayor grado de centralidad y por tanto más relaciones con otros genes.

En cuanto al **grado de conectividad** obtuvimos un valor del 19%, el cual era bastante pequeño y nos indicaba que nuestro grafo no era muy fuerte. Esto puede

Figure 3 Descripción de la figura.**Figure 4** Descripción de la figura.

deberse a que las comunidades entre si no tenían una gran dependencia y teníamos varios genes que no nos interesaban.

Empleamos distintos algoritmos para **detectar comunidades**. En el caso del método de Girvan-Newman había nodos que no pertenecían a ninguna comunidad. Esto podía deberse a la forma en que el algoritmo de betweenness identifica comunidades. Puede detectar comunidades basándose en la centralidad de intermediación de los nodos, y algunos nodos pueden no estar claramente vinculados a una comunidad en función de esta medida.

Con el algoritmo voraz todos los nodos pertenecían a alguna comunidad. al igual que con el algoritmo de propagación de etiquetas. Sin embargo, en este último obtuvimos tan solo 3 comunidades por lo que la clusterización es mínima. Con la aplicación del método de Louvain también resultaban menos comunidades que con el algoritmo voraz. Observamos 4 comunidades.

Los algoritmos anteriores tenían la desventaja de que no reflejaban con exactitud la posibilidad de que un gen perteneciera a más de una comunidad. Para estudiar esto hicimos uso de **Link Communities**, y obtuvimos una gráfica donde cada gen era representado por un diagrama de sectores. Observamos que la mayoría de los genes pertenecen a más de una comunidad. Al centrarnos en nuestros genes de interés detectamos que el gen **TP53** tenía relación con 4 clusters distintos mientras que los otros genes de interés, SDHB, SDHb, HRAS, AKT1 se encontraban en una comunidad más diferenciada de color rosa.

Para el **estudio genes de interés** obtuvimos una tabla con dos columnas, el gen en concreto y sus genes vecinos. Observamos que todos los genes estaban relacionados entre sí menos el caso de TP53 y SDHD.

3.5 Relación de los genes de interés con fenotipos patológicos.

Tras enriquecer el conjunto de genes de interés hemos obtenido 729 entradas en las cuales hemos encontrado relaciones de genes con fenotipos patológicos como estos:

Table 1 Fenotipos patológicos con palabra clave papiloma

Genes	Fenotipos patológicos asociados
FLT4, PIK3CA, TP53, SDHC, NRAS, PTEN, SDHB, SDHD, HRAS, PIK3R1, AKT1, IKBKG	Papiloma
TP53, NRAS	Papiloma del plexo coroideo

Figure 5 Descripción de la figura.

Figure 6 Descripción de la figura.

Table 2 Fenotipos patológicos con palabra clave papiloma

Genes	Fenotipos patológicos genital
PIK3CA, TP53, SDHC, NRAS, PTEN, SDHB, SDHD, AKT1	Neoplasia genital
FLT4, PIK3CA, TP53, SDHC, NRAS, PTEN, SDHB, SDHD, HRAS, AKT1	Anomalía de los genitales externos masculinos
PIK3CA, TP53, SDHC, NRAS, PTEN, SDHB, SDHD, AKT1	Morfología anormal de los genitales internos femeninos

Table 3 Fenotipos patológicos con palabra clave renal

Genes	Fenotipos patológicos asociados
PIK3CA, TP53, SDHC, NRAS, PTEN, SDHB, SDHD, AKT1	Neoplasia renal
PIK3CA, SDHC, NRAS, PTEN, SDHB, SDHD, AKT1	Carcinoma de células renales
PIK3CA, TP53, SDHC, PTEN, SDHB, SDHD, AKT1	Anomalía de las glándulas suprarrenales
TP53, SDHC, PTEN, SDHB, SDHD	Neoplasia de la glándula suprarrenal
SDHC, SDHB, SDHD	Feocromocitoma extraadrenal
SDHC, SDHB, SDHD	Feocromocitoma suprarrenal

Table 4 Fenotipos patológicos con palabra clave carcinoma

Genes	Fenotipos patológicos asociados
PIK3CA, SDHC, NRAS, PTEN, SDHB, SDHD, HRAS, AKT1	Carcinoma folicular de tiroides
PIK3CA, TP53, SDHC, NRAS, PTEN, SDHB, SDHD, HRAS, AKT1	Carcinoma de tiroides
PIK3CA, TP53, SDHC, PTEN, SDHB, SDHD, AKT1	Carcinoma de células renales
PIK3CA, SDHC, NRAS, PTEN, SDHB, SDHD, AKT1	Carcinoma de endometrio
PIK3CA, SDHC, PTEN, SDHB, SDHD, AKT1	Carcinoma de células transicionales de vejiga
PIK3CA, NRAS, PTEN, HRAS, AKT1	Carcinoma colorrectal hereditario no polipósico
PIK3CA, NRAS, AKT1	Carcinoma no medular de tiroides
NRAS, HRAS1	Adenocarcinoma papilar de ovario
PIK3CA, AKT1	Adenocarcinoma de pulmón
PIK3CA, TP53	Carcinoma papilar de tiroides
NRAS, HRAS	Carcinoma basocelular
PIK3CA, HRAS	Carcinoma hepatocelular
PIK3CA, TP53	Carcinoma
TP53, AKT1	

Table 5 Fenotipos patológicos con palabra clave ovario y derivados

Genes	Fenotipos patológicos asociados
PIK3CA, TP53, PTEN, AKT1	Neoplasia ovárica
PIK3CA, AKT1	Adenocarcinoma papilar de ovario
PIK3CA, TP53, SDHC, PTEN, SDHB, SDHD, AKT1	Anomalía del ovario
PIK3CA, SDHC, PTEN, SDHB, SDHD, AKT1	Ovarios poliquísticos agrandados

Table 6 Fenotipos patológicos con palabra clave útero y derivados

Genes	Fenotipos patológicos asociados
PIK3CA, SDHC, NRAS, PTEN, SDHB, SDHD, AKT1	Neoplasia uterina
PIK3CA, NRAS, AKT1	Leiomiomas uterinos
FLT4, SDHB, SDHD, PIK3R1	Retraso del crecimiento intrauterino

4 Discusión

5 Conclusiones

Abreviaciones

VPH: virus del papiloma humano

Disponibilidad de datos y materiales

https://github.com/mariavidaa/HPO_Papilloma

Contribución de los autores

E.M.A.F. hipótesis, R.A.C.V. introducción, M.J.H.R. introducción, F.J.R.L. objetivos, M.V.M introducción