

# INSTITUTO POLITÉCNICO NACIONAL



## “PREDICCIÓN DE GENES DE HUMANO ASOCIADOS A LA INFECCIÓN POR VIRUS DEL DENGUE”

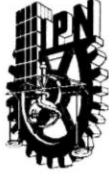
TESIS

QUE PARA OBTENER EL TÍTULO DE  
MAESTRO EN CIENCIAS DE BIOTECNOLOGÍA GENÓMICA

PRESENTA:

**Ing. AGUSTO REY LUZURIAGA NEIRA**

Reynosa, Tamaulipas, Junio 2012



# INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de Reynosa, Tam. siendo las 12:00 horas del día 11 del mes de Junio del 2012 se reunieron los miembros de la Comisión Revisora de Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de CBG para examinar la tesis titulada:  
Predicción de genes de humano asociados a la infección con virus del dengue.

Presentada por el alumno:

Luzuriaga  
Apellido paterno

Neira  
Apellido materno

Agusto Rey  
Nombre(s)

Con registro: 

B	1	0	1	9	2	2
---	---	---	---	---	---	---

aspirante de:

Maestría en Ciencias en Biotecnología Genómica

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA

Directores de tesis

M.C. Cristian Lizarazo Ortega

M.C. Aldo Segura Cabrera

Dr. Mario Alberto Rodríguez Pérez

Dra. Ana María Sifuentes Rincón

Dr. Xianwu Guo

M.C. Erick de Jesús De Luna Santillana

EL PRESIDENTE DEL

Dra. Ninfa María Rosas García



INSTITUTO POLITÉCNICO  
NACIONAL  
CENTRO DE BIOTECNOLOGÍA  
GENÓMICA



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

*CARTA CESIÓN DE DERECHOS*

En la Ciudad de Reynosa, Tam., el día 12 del mes Junio del año 2012, el (la) que suscribe Agusto Rey Luzuriaga Neira alumno (a) del Programa de Maestría en Ciencias en Biotecnología Genómica con número de registro B101922, adscrito a Centro de Biotecnología Genómica, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de M. en C. Cristian Lizarazo Ortega y el M. en C. Aldo Segura Cabrera y cede los derechos del trabajo intitulado “Predicción de genes de humano asociados a la infección con virus del dengue”, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección Bld. del Maestro esq. con Elías Piña S/N Col. Narciso Mendoza, C.P. 88710 Cd. Reynosa, Tamaulipas, México Tels. 01-899 9243627, 9251656. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.



---

Agusto Rey Luzuriaga Neira  
Nombre y firma

# ÍNDICE

<b>Contenido.....</b>	<b>Página</b>
Resumen.....	v
Abstract.....	vi
1. INTRODUCCIÓN .....	1
2. ANTECEDENTES .....	3
2.1. Historia y Características Generales del Dengue .....	3
2.2. Situación Actual del Dengue en América y México .....	4
2.3. Agente Etiológico.....	7
2.4. Mecanismo de Infección .....	8
2.5. La Enfermedad del Dengue.....	10
2.6. Redes de interacción de entre proteínas .....	12
2.6.1. Métodos de caracterización de las interacciones entre proteínas.....	14
2.6.2. Representación de las interacciones de proteínas y confiabilidad de sus predicciones .....	19
2.6.3. Utilidad de las redes de interacción entre proteínas para predecir nuevos genes asociados a enfermedades.....	22
2.7. Bases de Datos de Interacciones Entre Proteínas.....	23
2.7.1. Pfam (Protein Families Database).....	24
2.7.2. String (Search Tool for the Retrieval of Interacting Genes/Proteins).....	24
2.7.3. Intact ( <i>Intact Database</i> ) .....	24
2.7.4. InParanoid .....	25
2.8. Interactomas Reportados Previamente para Dengue.....	25
3. PLANTEAMIENTO DEL PROBLEMA .....	27
4. JUSTIFICACIÓN .....	28
5. OBJETIVOS .....	30
5.1. Objetivo General .....	30
5.2. Objetivos Específicos.....	30
6. HIPÓTESIS.....	30
7. METODOLOGÍA .....	31

7.1.	Obtención de Conjuntos de Datos de Genómica Funcional de Genes/proteínas de Humanos Asociados a la Respuesta a la Infección por el Virus del Dengue .....	31
7.2.	Modelaje de Cada uno de los Conjuntos de Datos Como una Red Individual.	32
7.3.	Integración de las Redes Individuales en una más Consistente de Acuerdo al Algoritmo Propuesto por Chua et al., (2007). .....	33
7.3.1.	Uniformización de puntaje para cada fuente de datos .....	33
7.3.2.	Estimación de la confianza de las fuentes de datos.....	35
7.3.3.	Estimación de la confianza de una interacción en la red combinada. ....	36
7.3.4.	Asignación de un puntaje de anotación a una proteína .....	37
7.5.	Estimación de la Fiabilidad de las Predicciones del Algoritmo .....	38
7.5.1.	Calculo del grado de unión a dengue de una proteína GUD.....	39
7.6.	Identificación de Asociaciones Funcionales de Proteínas Involucradas en la Respuesta a Dengue.....	40
8.	RESULTADOS.....	42
8.1.	Conjuntos de Datos de Genómica Funcional de Genes/proteínas de Humanos Asociados a la Respuesta a la Infección por el Virus del Dengue .... <b>¡Error! Marcador no definido.</b>	
8.2.	Evaluación del Desempeño del Algoritmo y Predicción de Nuevas Proteínas con Asociación a la Respuesta a Dengue .....	45
8.3.	Análisis funcional de la Red de Interacción de Proteínas de Humano en Respuesta a la Infección con Virus del Dengue .....	49
8.4.	Análisis Funcionales de Procesos Biológicos y Proteínas Candidato Relevantes.....	52
8.4.1.	Proteínas asociadas a división de megacariocitos.....	52
8.4.2.	Tirosina cinasa asociada a megacariocitos (Matk).....	55
8.4.3.	Dihidrolipoamida dehidrogenasa (Dld) .....	56
9.	DISCUSIÓN .....	58
10.	CONCLUSIONES .....	66
11.	REFERENCIAS BIBLIOGRÁFICAS.....	68
12.	APÉNDICE.....	78

## ÍNDICE DE CUADROS Y FIGURAS

Cuadro 1. Fuentes de datos y sus valores de probabilidad de acuerdo al algoritmo propuesto por Chua et al., 2007 para predecir interacciones asociadas a dengue.....	44
Cuadro 2. Asociación entre el contexto funcional de proteínas GUD y su conocida función en la respuesta a la infección con DENV.....	46
Cuadro 3. Valor predictivo positivo (VPP) para la predicción de genes asociados a la respuesta dengue basados en el GUD. ....	46
Cuadro 4. Anotaciones enriquecidas con relación anotaciones de participación en procesos biológicos de proteínas asociadas a la respuesta a DENV en términos de GO.	48
Cuadro 5. Proteínas asociadas a cascadas de señalización en la división de megacariocitos.....	53
Figura 1. (A) Método de doble híbrido en levadura. (B) Método de purificación por afinidad en tándem acoplado a espectrometría de masas.....	16
Figura 2. Métodos computacionales de predicción de las interacciones de proteínas ....	18
Figura 3. Red libre de escala. Los círculos en color gris representan a las proteínas centrales o focos de la red, en color rojo las proteínas de periferia, y las líneas representan las interacciones entre ellas .....	20
Figura 4. Calculo del GUD de una Proteína. ....	39
Figura 5. Contribución en número de proteínas e interacciones de cada fuente de datos a la red de interacciones de proteínas asociadas a la respuesta a dengue. ....	43
Figura 6. Curvas ROC construidas para evaluar la capacidad predictiva del algoritmo propuesto por Chua et al., 2007.. ....	49
Figura 7. Representación grafica de la red de interacciones de genes de humano asociados a la respuesta a la infección del DENV. ....	51
Figura 8. Representación gráfica de la red de interacciones de genes de humano asociados a la respuesta a la infección del DENV. ....	54
Figura 9. Agrupamiento de proteínas que interactúan con la Tirosina cinasa asociada a megacariocitos Matk. ....	55
Figura 10. Agrupamiento de proteínas que interactúan con la Lipoamida dehidrogenasa Dld.....	56

## **Predicción de genes de humano asociados a la infección por virus del dengue**

### **Resumen**

El dengue es una de las enfermedades virales transmitidas por vector de mayor importancia a nivel mundial. Sin embargo, algunos factores relevantes sobre la respuesta del organismo humano a esta infección no son totalmente comprendidos.

La biología de sistemas basada en el conocimiento de las interacciones entre proteínas proporciona un escenario óptimo para analizar procesos celulares, entre ellos aquellos que están involucrados en la respuesta a agentes infecciosos. Aunque los métodos experimentales proporcionan grandes cantidades de información para comprender las redes de interacción que forman las proteínas en un organismo o entre organismos, son las técnicas computacionales quienes han permitido integrar de manera sistemática dicha información y utilizarla para predecir con alta precisión nuevas proteínas asociadas a algunas enfermedades.

Con el principal propósito de integrar aproximaciones computacionales para el estudio de la respuesta del humano a la infección por el virus del dengue se tomó información de proteínas de humano provenientes de 8 bases de datos. Posteriormente se construyeron redes de interacción individuales que se integraron en una más consistente por medio del algoritmo propuesto por Chua et al. 2007, obteniendo 1794 proteínas de referencia.

Mediante un análisis de contexto funcional se encontraron 238 proteínas candidato con alta probabilidad de estar implicadas en la respuesta a la infección; estas proteínas están asociadas principalmente a rutas de señalización, respuesta a estrés y procesamiento de

antígenos. De forma individual se analizaron dos de las proteínas candidato más relevantes: la Tirosina cinasa (Matk) y la Dihidrolipoamida dehidrogenasa (Dld). La primera es una tirosina cinasa asociada a megacariocitos cuya inactivación se ha comprobado que produce trombocitopenia en ratones; mientras que Dld conforma un complejo que participa en procesos metabólicos y dentro de la red creada una varios grupos de proteínas.

En conclusión la estrategia aplicada permitió identificar proteínas candidato implicadas en procesos como transducción de señales, procesamiento de antígenos y respuesta a estrés. El análisis funcional de algunas de estas proteínas podría ayudar a explicar la trombocitopenia y el componente autoinmune en casos graves de dengue.



## **Prediction of human genes associated with dengue virus infection**

### **Abstract**

Dengue is one the most important worldwide viral vector-borne disease. However some relevant factors to the human body response to this infection are not fully understood. Systems Biology based on knowledge of the protein interactions provides an optimal environment for analyzing cellular processes, including those involved in the response to infectious agents. Although experimental methods provide large amounts of information for understanding the interaction networks of proteins in an organism or between organisms, bioinformatics methods allowed the systematic integration of this information and have been used to predict with high accuracy novel proteins associated with some diseases.

The integration of computational approaches to study human response to infection with dengue virus information was performed including information of human proteins from 8 databases. Subsequently, individual protein interaction networks were built and integrated in a more consistent network through the algorithm proposed by Chua *et al.* 2007, 1794 reference proteins were identified.

Using a functional context analysis we identified 238 candidate proteins with high probability of being involved in the response to dengue infection. These proteins are mainly associated signaling pathways, stress response and antigen processing. The particular analysis of relevant candidate proteins was performed. For example, we analyzed two candidate proteins Matk and Dld. Matk is a tyrosine kinase associated with

megakaryocytes, which inactivation of Matk gene has been found to produce thrombocytopenia in mice. Dld is the dihydrolipoamide dehydrogenase, which is involved in forming a complex that participates in several metabolic processes. Our network showed that Dld binds several groups of proteins; therefore, this protein could have an important role in the human response to dengue infection.

In conclusion the method used in present study was able to identify new proteins involved in processes such as signal transduction, antigen processing and stress response. The functional analysis of relevant candidate proteins as Dld and Matk could contribute to explain the functioning of thrombocytopenia and autoimmune component in severe cases of dengue.

## 1. INTRODUCCIÓN

Las proteínas interactúan en una forma altamente específica, dichas interacciones desempeñan un papel muy importante en varios procesos celulares, por ejemplo, la distorsión de las interfaces (zona de interacción física) de las proteínas puede conducir al desarrollo de algunas enfermedades (Shoemaker y Panchenko 2007a). La interacción entre proteínas es importante para el desarrollo de muchos procesos celulares, tales como la transducción de señales, median el paso de señales del exterior al interior de la célula.

Para una comprensión adecuada del mecanismo a nivel molecular y celular de las interacciones entre proteínas se han desarrollado varias estrategias tanto experimentales como bioinformáticas, mismas que se han utilizado previamente para comprender desde este enfoque procesos biológicos en organismos modelo, así como también las interacciones entre proteínas de humanos con las de agentes infecciosos como el virus de inmunodeficiencia humana (VIH) (Gautier *et al.* 2009) o el virus de la hepatitis C (VHC)(de Chasse *et al.* 2008).

Los virus tienen una capacidad genética limitada y dependen de los factores celulares de sus hospederos para completar su ciclo de vida; para lo cual las proteínas virales interactúan con las celulares, lo que les permite manipular rutas para crear ambientes favorables para su replicación. En respuesta a esto, las células de los hospederos contrarrestan la infección viral mediante la expresión de proteínas que permiten la inhibición o degradación de las proteínas virales. Sin embargo, para la mayoría de infecciones virales que atacan al humano se desconoce las interacciones entre proteínas

que se dan en respuesta a estas infecciones. El virus del dengue (DENV, por sus siglas en inglés) es un virus conformado por una envoltura y una molécula de RNA positivo, que pertenece a una importante familia de virus patógenos (*Flaviviridae*) la cual incluye otros virus como el de la hepatitis C y el virus del oeste del Nilo. El virus del dengue se transmite por su vector *Aedes aegypti*, un mosquito que desarrolla su ciclo biológico cerca de zonas urbanas lo cual ha contribuido al incremento de las infecciones humanas y a su distribución geográfica (Mackenzie et al., 2004). Cada año se presentan más de 50 millones de infecciones por dengue y cerca de dos billones de personas viven en zonas donde el dengue se transmite (Guha-Sapir et al., 2005). La enfermedad del dengue ocasiona un espectro clínico que va desde asintomático hasta los casos severos los cuales son potencialmente mortales. Actualmente no existen vacunas para prevenir la infección del dengue, tampoco existen tratamientos específicos para los pacientes infectados.

A pesar que hay reportes de redes de interacción entre proteínas virales y proteínas de humano, las interacciones entre proteínas de humanos en respuesta al ataque viral han sido poco caracterizadas. La red generada en la presente investigación integra información proveniente de diversas fuentes por medio de la aplicación de un algoritmo computacional, el cual permitió predecir proteínas que pueden estar implicadas en la respuesta a la infección por el virus del dengue. Se demuestra que el análisis del contexto funcional de las proteínas que integran una red permite identificar proteínas candidato con altas posibilidades de estar implicadas en la respuesta al proceso infeccioso. El trabajo aporta un listado de 238 proteínas candidato cuyo análisis podría contribuir a explicar algunos procesos de los cuales se tiene poco conocimiento.

## 2. ANTECEDENTES

### 2.1. Historia y Características Generales del Dengue

El dengue es una enfermedad viral transmitida por mosquitos del género *Aedes* que fue reconocida desde épocas tan remotas como 1635 y 1699 en el Caribe y Centroamérica, respectivamente. La primera gran epidemia de una enfermedad parecida al dengue fue descrita entre 1779 y 1780 en tres continentes: Asia, África y América del Norte. Sin embargo, cuadros clínicos semejantes a los de dengue se encuentran reportados en la Enciclopedia China de la Dinastía Chin (265-420 d. C.); los antiguos chinos ya pensaban que la enfermedad estaba relacionada a insectos voladores asociados al agua. El término "dengue" se originó en América entre 1827 y 1828, a raíz de una epidemia en el Caribe que cursaba con fiebre, artralgias y exantema. Los esclavos provenientes de África identificaron a esta entidad patológica como dinga o dyenga, homónimo del swahili "Ki denga pepo" que significa ataque repentino (calambre o estremecimiento) provocado por un "espíritu malo" (Halstead 1988)

El agente causal de la enfermedad es el virus del dengue, que en realidad son un grupo 4 de virus cada uno conocido como serotipo (DENV1-DENV4). El origen del virus ha sido objeto de muchas discusiones. En un inicio (1883) se especulaba que había surgido en África y distribuido alrededor del mundo con la trata de esclavos. Sin embargo, hacia 1956 se propone que se originan en un ciclo forestal entre primates y mosquitos en la península de Malasia. *Aedes aegypti*, es considerado el principal vector de los virus del dengue. Este díptero es de origen africano, se piensa convivía en un ambiente peri-doméstico en las aguas almacenadas de las aldeas africanas. La especie se adaptó a los humanos y la trata de esclavos que tuvo lugar durante los siglos XVII, XVIII y XIX,

contribuyó a la diseminación del mosquito en todo el mundo, fundamentalmente en las Américas, donde se adaptó al ambiente urbano. El virus dengue, así transportado, tanto en los esclavos como en su vector, se asentó del otro lado del Atlántico (Kroeger y Nathan 2006).

El dengue, a nivel global, es considerado la enfermedad viral transmitida por mosquitos de mayor importancia médica. Su espectro clínico puede variar desde individuos asintomáticos hasta otros en los que la infección tiene un desarrollo más severo, con sangramiento y choque (shock). En la mayoría de los casos, la infección evoluciona en forma asintomática, aunque la fiebre indiferenciada es observada con frecuencia en niños pequeños. El período de incubación después de la picadura del mosquito se ha estimado en 3 a 8 días como promedio. Dos cuadros clínicos principales se reconocen, la fiebre del dengue (FD), y su forma severa, la fiebre hemorrágica del dengue/síndrome de choque del dengue (FHD/SCD). Ambas formas clínicas se presentan en niños y adultos aunque la enfermedad severa se ha asociado principalmente a la niñez, sobre todo en el Sudeste asiático y Pacífico occidental (Guha-Sapir et al., 2005).

## **2.2. Situación Actual del Dengue en América y México**

La primera epidemia de dengue reportada en América ocurrió en Cuba en 1981. Se reportaron más de 344000 casos de los cuales alrededor de 10000 fueron severos y muy severos, reportándose 158 fallecidos. La segunda epidemia reportada ocurrió en 1989 en Venezuela (Kouri *et al.* 1989). Desde entonces se ha reportado un incremento de los casos de FHD, al igual que se ha reportado epidemias en otros países como Brasil (Guzman 2005).

El incremento en el número de casos de DH según Simmons y Farrar (2009), obedece a los cambios demográficos y las tendencias económicas de la explosión de la población global; con el aumento de la densidad poblacional, la urbanización no planificada, el incremento de las migraciones, se ha favorecido la introducción del virus del dengue; otros de los problemas para el incremento del DH y la FHD es el deterioro de los programas de control de vectores.

Después de las campañas de erradicación de la fiebre amarilla durante 1950 y 1960 se logró la desaparición del mosquito que actúa como vector de transmisión en la mayor parte del continente. Tras el abandono de dichas actividades, paso poco tiempo para que el vector reapareciera consecuentemente también reaparecieron los casos de FD.

México no fue la excepción y se empezaron a comunicar casos a partir de 1978, desde entonces y hasta el 2007 se ha observado un comportamiento cíclico en la aparición de casos de casi hasta 50000 casos, seguidos por años en los casos han disminuido notablemente. A partir del año 2000 ocurre un incremento notorio de casos desde menos de 2000 hasta más de 48000 en el año 2007 (DGE/SS, 2008).

En cuanto a la circulación de serotipos, la información disponible desde 1995 muestra el tránsito de los 4 serotipos de dengue existentes, en este período; no obstante, la frecuencia de los aislamientos por serotipo ha variado y se han registrado aislamientos asincrónicos. Desafortunadamente, la escasa información virológica disponible limita el panorama sobre lo que sucede fuera de los brotes, o bien lo ocurrido en años anteriores.

Para tener una idea mas completa sobre la dinámica de circulación de los serotipos virales se necesitaría caracterizar los anticuerpos neutralizantes específicos de cada serotipo y conocer el estado de la memoria inmunológica en las diferentes regiones del país, incluidos no sólo los casos sintomáticos. En este sentido, un estudio reciente en estudiantes universitarios en Tabasco realizado y por Sánchez y colaboradores (2008), reveló la sero-prevalencia del 9% de anticuerpos IgG; esta población se habría inmunizado por todos los serotipos debido a que la totalidad de los sujetos positivos tenía anticuerpos contra el DENV-2, seguidos en menor proporción por el DENV-4 (68%), DENV-1 (20%) y DENV-3 (4%). Lo anterior sugiere una notable memoria inmunológica contra el DENV-2 en la población joven de Tabasco (DGE/SS, 2008). Falcon *et al.*, (2009) señalan que, si bien el número de casos desde 1978 a 2007 indica un patrón cíclico con brotes separados por casi 20 años entre sí (1980 y 1997), el brote de 2007 acorta este periodo, lo cual podría significar que a medida que se establece la circulación de todos los serotipos, se podrían observar episodios epidémicos con una mayor frecuencia, tal y como al parecer ocurre en el sureste asiático.

Durante el 2011 hasta el mes de noviembre se han reportado 12738 casos probables de dengue, comprobándose 998 para FD y 181 para FHD, a esa fecha se presentaron 5 defunciones. El estado donde existieron el mayor número de casos de FD fue el estado de Nuevo León con 191 casos confirmados; en lo referente al estado con mayor incidencia de FHD, fue el estado de Guerrero donde se presentaron 42 casos; los estados de Chihuahua, Tlaxcala y Distrito Federal fueron los de menor incidencia y no se han



reportado casos. Para Tamaulipas se reportaron 20 casos, de los cuales 18 fueron de FD y 2 de FHD (DGPI/SSP, 2011).

### **2.3. Agente Etiológico**

La enfermedad del dengue es causada por 4 serotipos virales estrechamente relacionados, estos serotipos virales tienen una similitud en sus secuencias del 70%, dichos serotipos pertenecen a la familia *Flaviviridae* en general se habla de ellos como virus del dengue (DENV) la familia *Flaviviridae* incluye a más de 70 virus de los cuales algunos se transmiten por vectores, principalmente artrópodos (Holden *et al.* 2006)

El DENV es un virus de ARN simple positivo que codifica para tres proteínas estructurales; la cápside (C) que recubre y protege al ARN; la proteína precursor de la proteína la membrana (prM) y la envoltura (E). La proteína E es una glicoproteína la cual es el factor principal en la relación con la biología viral y la inmunidad humoral, debido a que en ellas radican la principales funciones biológicas del Virón; también media la fusión de la membrana e induce a la formación de anticuerpos neutralizantes, inhibidores de la hemaglutinación e inmunoamplificadores; adicionalmente en ella se encuentra el receptor viral (Kuhn *et al.* 2002). La proteína prM protege a la proteína E del pH ácido del medio durante la maduración conformacional evitando su cambio estructural irreversible.

El genoma de este virus también codifica para 7 proteínas no estructurales (NS) 5'-C-prM-E-NS1--NS2A-NS2B-NS3-NS4A-NS4B-NS5-3' (Xu *et al.*, 2009). La función de la

proteína NS1, no está bien determinada pero se cree que participa en la maduración viral, también se sugiere que pudiera interactuar con la NS4, en su función de replicasa (Wu *et al.* 2003; Zuo *et al.* 2009). La NS2 está constituida por dos proteínas NS2A y NS2B, se conoce que la NS2, está asociada a la membrana y forma un complejo con la NS3, siendo un cofactor requerido para la función de serina proteasa de NS3 (Zuo *et al.* 2009). La NS3 es una proteína que tiene funciones de serina proteasa, helicasa, y nucleótido trifosfatasa; es la fuente principal de epítomos citotóxicos. Lescar y cols. (2008), demostraron que algunas mutaciones pueden disminuir las funciones de esta proteína lo cual abrió el camino para el diseño de inhibidores alostéricos. La NS4 está conformada por 2 proteínas NS4A y NS4B, las dos se ubican en sitios de replicación del ARN (Bartenschlager y Miller 2008). En la proteína NS5 radica la función polimerasa del virus (Xu *et al.* 2005).

#### **2.4. Mecanismo de Infección**

Los humanos son infectados con el virus del dengue a través de la picadura del mosquito *Aedes* spp. que actúa como vector (Gubler 1987). Las hembras de los mosquitos pueden alimentarse varias veces, picando a personas diferentes en un corto período de tiempo. Este comportamiento convierte a *Aedes* spp. en un vector epidémico eficiente.

Los monocitos y otras células de origen se reconocen como células blancas para la infección (Marianneau *et al.* 1999). Sin embargo estudios más actuales sugieren que son las células dendríticas (CD) los principales blancos; entre estas células tenemos: Langerhans, CD intersticiales y CD plasmocitoides (Ho *et al.* 2001; Wu *et al.* 2000). En

la infección natural el virus parece ser infectado en el espacio subcutáneo; 24 horas post inoculación el virus se ha podido detectar en los ganglios linfáticos regionales; las células de Kupffer y los linfocitos B pueden ser otros blancos para la infección; también se han podido infectar *in-vitro* células de origen endotelial y epitelial (McBride y Bielefeldt-Ohmann 2000).

El establecimiento de la infección requiere de la entrada del virus dentro de la célula, seguido por la liberación de la nucleocápside. Esto se logra mediante la fusión de la membrana viral con la membrana celular lo cual puede ocurrir en las membranas plasmáticas o en las membranas endosomales, después que el virus se introduce por medio de endocitosis. Los cambios conformacionales provocados por la membrana plasmática debido a los receptores virales o por el pH ácido de los endosomas dejan expuesto el péptido de fusión, el cual interactúa con la membrana blanco e inicia el evento de fusión (Kielian 1995; Lu *et al.* 1999; Smit *et al.* 1999).

La proteína viral E media la unión a las células, pero la naturaleza de los receptores virales aun es elusiva. Los candidatos a receptores virales frecuentemente investigados incluyen proteínas como los receptores Fc, glicosaminoglicanos (GAGS) y moléculas de lipopolisácaridos asociados con CD14 (Bielefeldt-Ohmann 2000). Estudios recientes también sugieren que la primera interacción del virus se da con elementos de baja afinidad presentes en la superficie celular, para después ser transferido aun receptor protéico capaz de unir e internalizar el virus. Entre los elementos de baja afinidad tenemos: el sulfato de heparina, una lectina tipo C y un homólogo de este, la mas caracterizada es la lectina tipo C conocida con el nombre de CD20 (Clyde *et al.* 2006),

un receptor de baja afinidad. Cabe señalar que el mecanismo de internalización depende del tipo célula; así en algunos tipos celulares la internación del virus se da mediante la unión a complejos de anticuerpos creados en respuesta a la infección (Burke y Kliks 2006).

Luego de la internación y liberación del genoma viral, el RNA se dirige al ARN hacia los ribosomas donde se traduce el marco de lectura para producir la poliproteína. El ensamblaje de la cápside y la proteína se realiza dentro de la célula. Las partículas celulares son llevadas en vesículas de transporte que salen del retículo endoplasmático y por un mecanismo específico son llevadas al aparato de Golgi donde existe un mecanismo exocítico de membrana a la superficie celular. La liberación del virus puede ocurrir por fusión de las membranas de la vesícula exocítica, que contienen al mismo, con la membrana plasmática; o por efecto citopático por rupturas puntuales en la membrana que lo separa del exterior (Martina *et al.* 2009).

## **2.5. La Enfermedad del Dengue**

La infección por el virus del dengue puede causar un espectro de enfermedad que incluye: la fiebre indiferenciada que suele durar hasta siete días, el cuadro clínico que se presenta además de la fiebre elevada va acompañado de dolor retro-orbital, dolor de las articulaciones y sarpullido, el pronóstico es positivo y raramente causa la muerte. La fiebre hemorrágica del dengue (FHD), es una complicación potencialmente mortal que incluye tendencias hemorrágicas, trombocitopenia y escape de plasma; y el síndrome de shock de dengue (SSD) además de los factores previamente señalados incluye fallas

circulatorias y afección de órganos, principalmente el hígado. La FHD y el SSD son potencialmente mortales pero los pacientes con diagnóstico temprano y una terapia apropiada, se pueden recuperar sin ninguna secuela.

En el caso de la FD se debe hacer un seguimiento de los síntomas; para el caso de FHD aparte del seguimiento de los síntomas se debe hacer un monitoreo continuo de los signos vitales. El SSD se considera como una emergencia médica que requiere hospitalización con cuidados intensivos (Dutta *et al.* 2011).

El incremento de la mortalidad del dengue se debe considerar como el reflejo de la proporción de pacientes con FD que desarrollan FHD/SSD. La patogénesis de la FHD/SSD es ampliamente considerada como una potenciación dependiente de anticuerpos, misma que se da en una infección secundaria con un serotipo heterólogo del virus (Thein *et al.* 1997). Sin embargo, se ha observado la ausencia de una asociación significativa entre una infección secundaria, la co-circulación de diferentes serotipos y FHD/SSD (Harris *et al.* 2000).

Existen algunos otros factores como los altos niveles de viremia (niveles de circulación del virus dentro del humano) que se han asociado al surgimiento de FHD/SSD; se cree que los niveles de viremia están asociados a la capacidad virulenta de cada serotipo; sin embargo esto aun no se ha demostrado (Guha-Sapir y Schimmer 2005). La virulencia, la respuesta inmunológica y el incremento de la patogenicidad de ciertos serotipos han sido implicados como críticos para el surgimiento de FHD/SSD, esto se ha demostrado para

los serotipos DENV1, DENV2 y DENV3, pero no para DENV4 (Rico-Hesse *et al.* 1997). Las evidencias de diferentes estudios han demostrado que la patogénesis de la FHD/SSD, puede ser multifactorial, sin embargo, sus mecanismos son aun desconocidos.

La viremia es una característica que diferencia al virus del dengue de entre los demás Flavivirus, niveles altos de viremia se han observado principalmente durante el desarrollo de FHD/SSD. Se han propuesto varios mecanismos para el desarrollo de la fase clínica de FHD/SSD; la falta de un modelo animal ha obstaculizado el conocimiento de la etiología de la enfermedad en esta fase. Se ha correlacionado positivamente los niveles de viremia con la severidad de la enfermedad del dengue; sin embargo, el o los linajes celulares que sirven como blancos primarios para la fuente de viremia son aun desconocidos. Resultados de estudios *in-vivo* e *in-vitro* usando técnicas avanzadas han revelado que el virus del dengue parece estar asociado con las plaquetas y el linaje megacariocítico. Estas evidencias podrían parcialmente explicar la disfunción de las plaquetas observada en pacientes afectados con FHD/SSD (Noisakran *et al.* 2009).

## **2.6. Redes de interacción de entre proteínas**

El reduccionismo que ha dominado la investigación biológica en los últimos años, ha proporcionado una riqueza de conocimientos acerca de los componentes celulares individuales y sus funciones. A pesar de su enorme éxito, cada vez es más claro que una función biológica discreta rara vez puede atribuirse a una sola molécula. En cambio, la mayoría de características biológicas surgen de las interacciones complejas entre los

numerosos componentes de la célula, tales como proteínas y ácidos nucleicos (Barabasi y Oltvai 2004). Este nuevo enfoque, que considera que un sistema biológico es más que sólo la suma de sus partes es el objeto de estudio de la llamada biología de sistemas. Así, la biología de sistemas permite analizar de manera global a los elementos de un sistema y como, en dicho contexto global, aparecen características y/o propiedades distintas a las de las partes que lo componen y que “emergen” e éstas cuando entran en acción. El concepto de biología de sistemas no es nuevo y tuvo su inicio a mediados del siglo XX (Oyama *et al.* 2002). El espectro de la aplicación de la biología de sistemas es amplio, pudiéndose aplicar a fenómenos complejos como la circulación sanguínea, interacción entre dos células bacterianas, interacciones proteína-proteína, entre otros. Actualmente, en la era post-genómica, la principal aplicación de la biología de sistemas consiste, en integrar información que permita comprender las interacciones entre los componentes que integran a la célula, y su posible relación con las enfermedades multifactoriales tales como el cáncer, diabetes, obesidad y otras.

Las proteínas son los actores principales de los procesos biológicos que ocurren en la célula y en esta podemos encontrar miles de proteínas, cada una con su propia función. Así mismo, es bien conocido que cuando alguna proteína no desempeña su función correctamente (por ejemplo, debido a mutaciones), esto se puede traducir en la aparición de una enfermedad. Con la disponibilidad de la secuencia del genoma humano y de otros organismos, los nuevos estudios se enfocan a caracterizar los proteomas, es decir, la determinación y análisis de las proteínas que expresa un organismo bajo determinadas condiciones ambientales. Sin embargo, el análisis de los proteomas bajo

los principios de la biología de sistemas ha revelado que las proteínas no trabajan de manera aislada; por el contrario, ellas interactúan y forman redes complejas (interactoma).

### **2.6.1. Métodos de caracterización de las interacciones entre proteínas**

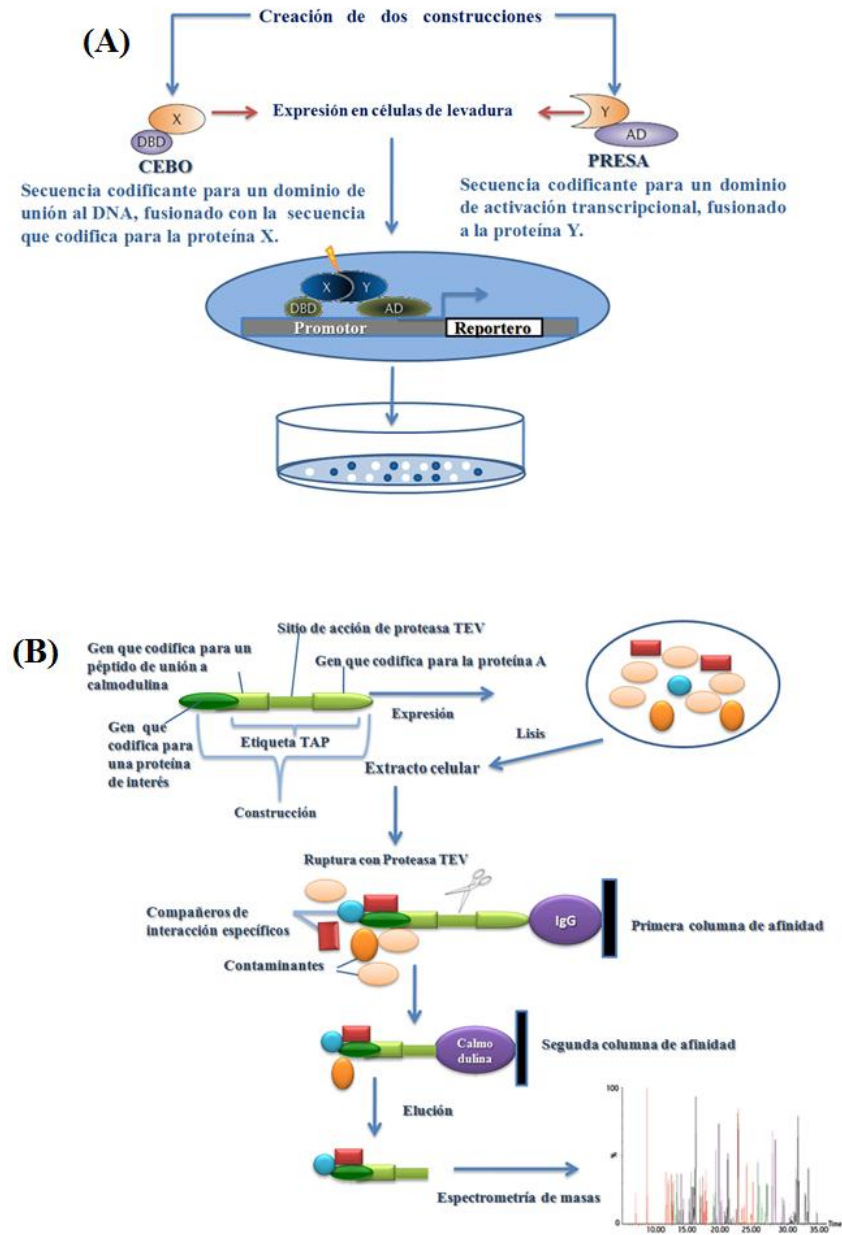
Tradicionalmente las interacciones entre proteínas se han caracterizado individualmente por métodos bioquímicos, genéticos o fisicoquímicos. Sin embargo, la velocidad con la que actualmente nuevos genomas se secuencian y el desarrollo de técnicas experimentales de gran cobertura, hacen posible analizar globalmente las interacciones entre proteínas. Consecuentemente en los últimos años, la aplicación de estas técnicas ha arrojado una gran cantidad de datos de interacciones. No obstante para un uso adecuado de estos datos se deben considerar factores como su confiabilidad, sesgo, traslape y complementariedad (von Mering *et al.* 2002).

Los métodos para caracterización de las interacciones proteína-proteína pueden ser de dos tipos: los métodos experimentales y los computacionales. Dentro de los métodos experimentales los utilizados más frecuentemente son: el doble híbrido de levadura (Y2H) (Uetz *et al.* 2000) y la purificación por afinidad acoplada a espectrometría de masas (TAP-MS) (Rigaut *et al.* 1999). El Y2H, permite identificar interacciones entre dos proteínas; y se basa en el hecho de que los factores de transcripción de los eucariotas tienen una estructura modular constituida de dos dominios separables: el dominio de unión al DNA (BD) y el dominio de activación (AD). Los factores de transcripción más utilizados en esta técnica son los de las levaduras (Gal-4). De acuerdo con el método se elaboran dos quimeras que incluyen las regiones codificantes de las



proteínas de estudio y los dominios de unión y activación al DNA. Posteriormente, se permite la expresión de ambas quimeras y si las proteínas que producen son interactuantes harán que el factor transcripcional se re-ensamble y que se active la transcripción de un gen utilizado como reportero (Figura 1A).

El método TAP-MS permite analizar las interacciones de las proteínas dentro de un complejo. La técnica TAP-MS consiste en la generación de una quimera que incluye un segmento génico que codifica para una proteína de interés con un segmento que codifica para la etiqueta TAP. Una etiqueta TAP está formada por regiones codificantes para la Proteína A de *Staphylococcus aureus* la cual funciona como sitio de unión para inmunoglobulina G (IgG), un péptido de unión a calmodulina separado y para un sitio activo para la proteasa del virus del tabaco (Puig *et al.* 2001; Rigaut *et al.* 1999). Posteriormente, la quimera se expresa en una célula de levadura donde esta puede formar complejos con otras proteínas. Para eliminar la etiqueta y purificar el complejo se utilizan procedimientos bioquímicos, basados en la actividad de algunas proteasas (Figura 1B). Finalmente los componentes de los complejos se identifican mediante espectrometría de masas (MS); esta técnica permite identificar los componentes del complejo mediante la medición de la masa de iones de las moléculas que lo constituyen (Figura 1B).



**Figura 1.** (A) Método de doble híbrido en levadura. (B) Método de purificación por afinidad en tándem acoplado a espectrometría de masas. Adaptado de Segura-Cabrera, 2008.

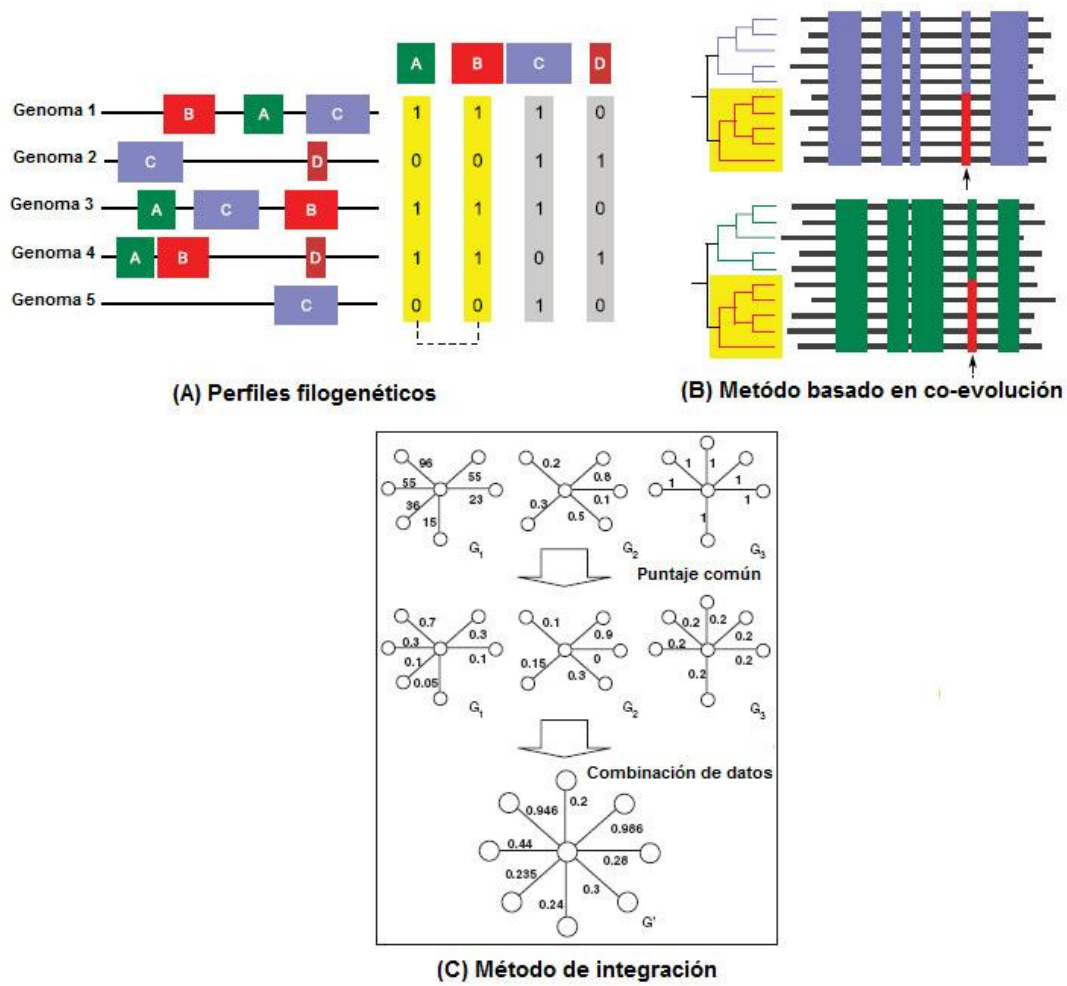
Los dos métodos presentan limitaciones, debido a ello, de forma experimental se puede obtener una mejor información cuando se combinan o cuando los datos de interacción son complementados con datos de perfiles de expresión o perfiles filogenéticos (Ge *et al.* 2003).

Los métodos computacionales de predicción de las interacciones de proteínas surgen como una alternativa a la determinación experimental, para lo cual se utiliza información disponible en algunas bases de datos públicas. Así mismo, se consideran como una alternativa para tratar con el problema de la baja cobertura de las técnicas experimentales, aunque algunos de ellos son dependientes de datos obtenidos por dichas técnicas. Se pueden utilizar de forma complementaria con los métodos experimentales para obtener una mayor consistencia de los resultados o se pueden utilizar de forma independiente.

Los métodos computacionales de predicción de interacciones entre proteínas a su vez pueden subdividirse en tres clases: los basados en el contexto genómico, los basados en co-evolución y los llamados métodos de integración de datos. Por ejemplo, el método de perfiles filogenéticos (Pellegrini *et al.* 1999) está dentro de los que se basan en el contexto genómico, consiste en la identificación de la presencia o ausencia de genes ortólogos (genes provienen de un ancestro común y que son producto de eventos de especiación). Si el patrón de proteínas ortólogas se conserva entre organismos, se debe a que el par de proteínas trabajan de manera conjunta tal que si una de las proteínas estuviera ausente la otra no podría realizar su función (Figura 2a).

Los métodos de co-evolución se basan en la teoría de que especies ecológicamente interactuantes evolucionan de forma conjunta, lo cual implica la evolución de una especie en respuesta a la selección impuesta por otra. La co-evolución entonces requiere de una presión selectiva en dos o más especies. De acuerdo a estos métodos una de las formas de predicción es la comparación de árboles filogenéticos (Figura 2b) o historias evolutivas, bajo la premisa que la proteínas interactuantes son sujeto de presiones

evolutivas similares, resultando en topologías similares para el árbol correspondiente (Jeong *et al.* 2001).



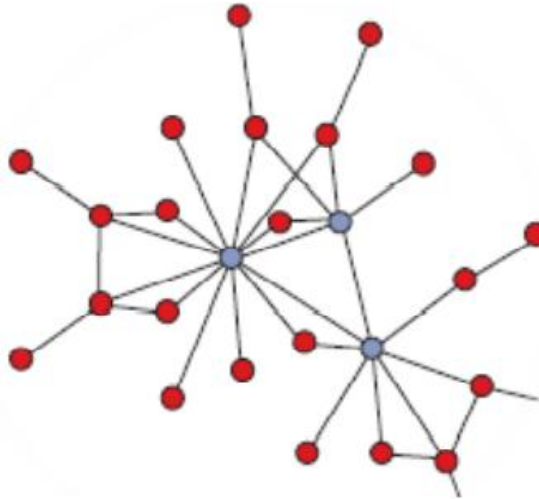
**Figura 2.** Métodos computacionales de predicción de las interacciones de proteínas. Adaptado de Rama, 2010 y Chua, et al. 2007

Con el incremento de la disponibilidad de diversa información biológica, los métodos de predicción de interacciones de proteínas convergen en la integración de datos heteróneos. Un método reciente propuesto es el “prometeora” (IWA) (Chua *et al.* 2007), este método combina datos de diferentes fuentes utilizando los puntajes asociados a las mismas, como por ejemplo los

valores  $E$  del sistema de alineamiento local (BLAST). El método incluye tres pasos básicos (Figura 2C). Primero: cada fuente de datos se evalúa calculando un puntaje de confiabilidad, basado en su habilidad para predecir funciones para proteínas respecto a datos de referencia. Segundo: cada fuente de datos se modela como una red o grafo ( $G_1$ ,  $G_2$  y  $G_3$ ); donde las proteínas corresponden a los vértices del grafo y las relaciones entre las proteínas son los enlaces entre ellas. Finalmente las redes se combinan para formar una red integrada y por ende, más consistente ( $G'$ ), encuyen un puntaje final calculado a partir de la confiabilidad de cada fuente de datos.

### **2.6.2. Representación de las interacciones de proteínas y confiabilidad de sus predicciones**

El análisis de datos experimentales reveló que en conjunto, las interacciones entre proteínas forman redes altamente complejas. Como se ha señalado, existen principios que se han aplicado a la caracterización de redes, por ejemplo, a nivel tecnológico y social. Una de las teorías que ha sido de gran utilidad para explicar la organización de las redes antes mencionadas es la teoría de grafos, en la cual cada elemento de un grafo es representado como un nodo y las líneas entre ellos representan sus asociaciones (Figura 3).



**Figura 3.** Red libre de escala. Los círculos en color gris representan a las proteínas centrales o focos de la red, en color rojo las proteínas de periferia, y las líneas representan las interacciones entre ellas. Adaptado de Barabási et al., 2004

Así mismo, algunos de los principios de la teoría de grafos han sido aplicados para la comprensión de las redes de interacción de proteínas, permitiendo hacer inferencias sobre las funciones de las mismas. Uno de los principios más utilizados para predecir función de las proteínas y derivado del análisis de las redes de interacción es el principio de la "cuerpo a cuerpo". Por ejemplo, caracterizamos a las relaciones entre proteínas cuyo mal funcionamiento ha sido asociado a cáncer y encontramos participando en dicha red a una proteína de la cual se desconoce su función, podemos asumir que un mal funcionamiento de esa proteína podría asociarse a cáncer, tal como lo demuestran diversos estudios (Gmeiner *et al.* 2008; Wachi *et al.* 2005).

Una de las redes de interacción entre proteínas mayormente descrita y utilizada para la predicción de funciones de genes o proteínas es la del gusano *Caenorhabditis elegans*. Lee et al. (2008), lograron crear una red que predice los efectos fenotípicos de la perturbación de genes. Para crear la red, integraron varios conjuntos de datos,

incluyendo mediciones de expresión de microarreglos de mRNAs, bases de datos de interacciones proteína-proteína, contenido de dominios y minería de datos. Así como también, se identificaron sus ortólogos en levaduras y humanos para integrar la información de estos organismos en la red (Interacciones físicas y asociaciones funcionales). La red de *Caenorhabditis elegans* obtenida comprende 384,700 interacciones entre 16,113 proteínas, lo cual representa aproximadamente el 82 % de su proteoma, mientras que la red determinada a nivel experimental solo comprende el 5 % (Cusick *et al.* 2005). Para demostrar que la red permite identificar efectos de la alteración del genotipo sobre el fenotipo, se utilizaron 6 genes (genes raíz), de los cuales previamente se había reportado su asociación a la supresión de tumores. Posteriormente, con estos 6 genes se realizó una búsqueda de sus parejas de interacción dentro de la red y así predecir nuevos genes involucrados en la supresión de tumores. Los 6 genes raíz interactúan con 62 genes dentro de la red de *C. elegans*. Mediante un análisis con RNAi, se probó la función de 50 de los 62 genes. Se encontró que 10 de los 50 (20%) estaban asociados a la supresión de tumores. En contraste, el análisis con RNAi a gran escala, sin considerar la predicción computacional, encontró que solamente 7 de 1748 genes (0.9 %) están asociados al fenotipo antes mencionado. Por lo tanto, la combinación de las predicciones computacionales derivadas de la red y su posterior verificación experimental son 21 veces mejores que las realizadas partiendo sólo del análisis experimental.

### **2.6.3. Utilidad de las redes de interacción entre proteínas para predecir nuevos genes asociados a enfermedades**

Las redes de interacciones entre proteínas pueden ayudar a generar listas de genes candidatos asociados a enfermedades o genes modificadores, asociados a los ya conocidos (Lim *et al.* 2006). Por ejemplo, la integración de una red con datos de co-expresión, mostró cuatro genes bien conocidos asociados a cáncer de mama. Con interacciones físicas y genéticas se elaboró una red modelo para el cáncer de mama, con el cual genes de susceptibilidad y genes modificadores fueron predichos (Pujana *et al.* 2007).

Las aplicaciones de estas redes van mucho más allá de detectar enfermedades de origen genético, también se han aplicado a la comprensión de los mecanismos de patógenos, en especial los virus, que tienen mecanismos que les permiten perturbar las redes dentro del huésped. Dado que muchas interacciones virus-huésped ocurren a nivel físico, varias redes se han obtenido por medio de métodos experimentales, como por ejemplo la del virus de la hepatitis C obtenida por el método del doble híbrido de levadura (Vidal *et al.* 2011).

La identificación de interacciones puede permitir que en nuevas investigaciones se puedan identificar blancos para combatir las manifestaciones clínicas de la enfermedad mediante el desarrollo de nuevos medicamentos. La mayoría de medicamentos actúan por unión a proteínas para modular sus actividades, afectando los procesos biológicos. Actualmente se encuentra disponible una base de datos de interacciones químico-proteicas la cual contiene datos de más de 68000 químicos conectados con 1.5 millones de genes a través de 373 genomas. Yildirim *et al.* (2007), construyeron un grafo de



medicamentos y proteínas unidas por asociaciones binarias para entender los blancos de los medicamentos en el contexto celular de las redes de enfermedades. El análisis de red mostro que los blancos de los medicamentos ocupan ciertas regiones del interactoma y sus firmas topológicas son diferentes en comparación a las proteínas esenciales. También se descubrió que la mayoría de medicamentos son paliativos ya que no perturban directamente la proteína correspondiente a la causa subyacente de la enfermedad. Un mapa a gran escala de todos los medicamentos disponibles en Estados Unidos se ha construido permitiendo una visión de estrategias posibles para el desarrollo de nuevos medicamentos.

En conclusión, la aplicación de métodos de genómica funcional tales como: el mapeo de redes de interacción de proteínas, enfocado a proteínas de interés médico puede llevar a encontrar genes modificadores, nuevos blancos para medicamentos y en consecuencia al desarrollo de nuevas terapias.

## **2.7. Bases de Datos de Interacciones Entre Proteínas**

Para el análisis de las interacciones entre proteínas se han creado varias bases de datos, que incluyen información obtenida por diversos métodos experimentales y de la información documentada en literatura científica. A continuación se exponen algunas bases de datos utilizadas para el desarrollo de la investigación.

### 2.7.1. Pfam (Protein Families Database)

<http://ipfam.sanger.ac.uk/>

Pfam es una base de datos de familias de proteínas ampliamente utilizada, actualmente (noviembre 2011) contiene información curada de 13,672 familias de proteínas (Punta *et al.* 2012). Esta base de datos es utilizada para predecir interacciones entre proteínas, basándose en la premisa que proteínas potencialmente interactuantes comparten dominios (Shoemaker y Panchenko 2007b).

### 2.7.2. String (Search Tool for the Retrieval of Interacting Genes/Proteins)

<http://string.embl.de/>

Es una base de datos para la exploración y análisis de asociaciones proteína-proteína. Las asociaciones se derivan de métodos experimentales a gran escala, minería de bases de datos y literatura, análisis de co-expresión, y también incluyen predicciones computacionales, principalmente de predicciones basadas en el contexto genómico (von Mering *et al.* 2005). La base de datos actualmente (enero 2012) tiene 5214234 proteínas de 1133 organismos (Szklarczyk *et al.* 2011).

### 2.7.3. Intact (*Intact Database*)

<http://www.ebi.ac.uk/intact/>

El objetivo principal de este proyecto es proporcionar un repositorio público con información referente a interacciones proteína-proteína derivadas experimentalmente (Chatr-aryamontri *et al.* 2008). Esta base de datos hace un uso extensivo de vocabularios controlados los cuales permiten obtener una descripción consistente con

los detalles de los experimentos que fueron utilizados para conocer la información. Para soportar búsquedas generalizadas, estos vocabularios tienen una estructura jerárquica, y en lo posible, se utilizan sistema de referencia tales como la base de datos de taxonomía del NCBI o la base de datos de Gene Ontology (GO). En la actualidad (enero 2012), tiene documentadas 290,056 interacciones.

#### **2.7.4. InParanoid**

<http://inparanoid.sbc.su.se/>

La bases de datos de InParanoid, es una base de datos de parejas de proteínas ortólogas de eucariotas, actualmente posee datos de 100 especies. InParanaoid, fue desarrollada específicamente para identificar grupos de verdaderos ortólogos y excluir proteínas estrechamente relacionadas pero no ortólogas (O'Brien *et al.* 2005).

#### **2.8. Interactomas Reportados Previamente para Dengue**

Recientemente métodos computacionales se utilizaron para predecir interacciones entre proteínas del mosquito, humano y el virus del dengue. Doolittle y cols. (2011); implementaron un método computacional para predecir interacciones entre el DENV y sus dos hospederos, *Homo sapiens* y *Aedes aegypti*. Su método se baso en las similitudes estructurales entre las proteínas del DENV y las proteínas de los hospederos incorporando reportes de interacciones encontrados en la literatura. Se predijeron mas de 4000 interacciones entre el DENV y el humano, a si como 176 interacciones entre el DENV y *A. aegypti*. Posteriormente se hizo un filtrado basándose en anotaciones de Gene Ontology y se redujo el numero, a 2000 para humanos y 18 para *A. aegypti*. Los

resultados mostraron, algunas predicciones de proteínas asociadas a la respuesta de estrés y muerte celular, y otras asociadas a rutas del sistema inmune (Doolittle y Gomez 2011).

Khadka y cols. (2011); utilizando métodos experimentales a gran escala, encontraron 139 interacciones entre proteínas del DENV2 y el humano, de las cuales la mayoría son nuevas. Estas interacciones involucran 105 proteínas, incluyendo 6 previamente reportadas como implicadas en la infección y 45 ligadas a la replicación de otros virus.

### **3. PLANTEAMIENTO DEL PROBLEMA**

El conocimiento de las interacciones entre proteínas es crucial para la comprensión del mecanismo molecular de las enfermedades. Se han hecho grandes esfuerzos para elaborar técnicas experimentales que permitan caracterizar dichas interacciones, sin embargo, la amplitud del proteoma a ocasionado que dichas técnicas sean insuficientes, además de que no se han pulido lo suficiente para evitar un alto número de falsos positivos. El contexto de las interacciones entre patógenos y humanos es aún incipiente, se han hecho esfuerzos para mediante técnicas experimentales establecer mapas de interacción de patógenos como el virus de inmunodeficiencia humana (HIV), o el virus de la hepatitis C (HCV). Para el caso del dengue existen solo dos trabajos reportados con anterioridad, sin embargo su cobertura es aún pobre y están encaminados a la búsqueda de interacciones asociadas al proceso de infección y no a la respuesta del organismo humano a dicha infección, del cual se desconocen aún muchos aspectos.

#### 4. JUSTIFICACIÓN

Las infecciones agudas del virus del dengue son uno de los mayores problemas de salud para muchos de los países de zonas tropicales y subtropicales. Las infecciones sintomáticas ocurren en espectro que va desde la fiebre de clásico (FD) hasta las formas severas que son la fiebre hemorrágica de dengue (FHD) y el síndrome de shock de dengue (SSD) (Khan *et al.* 2008). Las formas severas de dengue son una de las principales causas de hospitalización y muerte infantil, en países donde la enfermedad es endémica y con limitados recursos económicos para la salud. Mientras que el número de casos y países con incidencia de la infección reportados por la Organización Mundial de la Salud (OMS/WHO) se ha incrementado dramáticamente en las últimas cinco décadas (Gubler 2001), el conocimiento actual sobre mecanismo de infección y respuesta del organismo no ha permitido aun desarrollar una vacuna, un tratamiento específico, y se desconoce aun un modelo animal adecuado. Es notable la importancia de la infección dada la situación epidemiológica actual y sus perspectivas, por lo cual se hace necesario establecer nuevas estrategias para comprender los mecanismos moleculares de la enfermedad. Las interacciones entre proteínas son la base de las funciones biológicas, se conoce que cuando existe un proceso infeccioso estas tienden alterarse. Actualmente los mapas de interacciones entre proteínas reportados elaborados para dengue tienen ciertos inconvenientes, como la baja cobertura y el número elevado de falsos positivos y falsos negativos; otro factor relevante es que estos mapas se han desarrollado principalmente para conocer las proteínas involucradas en la entrada del virus al organismo, y no para conocer el mecanismo de respuesta del organismo en la infección. Se han descrito varios métodos para elaborar mapas de interacción, sin embargo, se conoce que se obtiene una

mayor cobertura del proteoma cuando se integra diversas fuentes, de datos. La construcción de una red de interacción de proteínas asociadas a la respuesta del humano a la infección por DENV, ayudará a revelar nueva información sobre el mecanismo de respuesta a la infección, así mismo, nos permitirá optimizar un método que nos permita identificar genes o proteínas asociados a otras infecciones virales, en futuras investigaciones.

## 5. OBJETIVOS

### 5.1. Objetivo General

- ✓ Integrar aproximaciones computacionales para el estudio de la respuesta del humano a la infección por el virus del dengue, desde una perspectiva de biología de sistemas.

### 5.2. Objetivos Específicos

- ✓ Obtener conjuntos de datos de genómica funcional de humanos asociados a la respuesta a la infección por DENV.
- ✓ Modelar cada uno de los conjuntos de datos como una red.
- ✓ Integrar las redes individuales en una más consistente, mediante el algoritmo propuesto por Chua et al., 2007
- ✓ Identificar nuevas asociaciones funcionales de proteínas involucradas con la respuesta a la infección por DENV.

## 6. HIPÓTESIS

“El análisis computacional de redes de interacción de proteínas en el contexto de la relación humano-virus del dengue, permite la identificación de proteínas de humano asociadas al mecanismo de respuesta a la infección por virus del dengue”



## 7. METODOLOGÍA

### 7.1. Obtención de Conjuntos de Datos de Genómica Funcional de Proteínas de Respuesta a la Infección del Humano por el Virus del Dengue

*Protein Corral* es una aplicación web que busca información de los resúmenes de *Medline* (Biblioteca Nacional de Medicina de Estados Unidos), de acuerdo a un criterio de búsqueda (palabras claves); con ayuda de esta herramienta se buscó proteínas de humano asociadas a dengue. Se obtuvo un grupo de 473 proteínas con las cuales se siguieron los siguientes procedimientos:

- Se buscó los dominios proteínicos en Pfam para las 473 proteínas de humano y se tomaron las proteínas que comparten por lo menos un dominio.
- Se hizo una búsqueda de homología, mediante un análisis de alineamiento local (BLAST), se seleccionaron las proteínas cuyas secuencias tengan porcentajes de similitud superiores a 80% y menores 90%.

De las bases de datos de STRING e Intact se extrajo toda la información de interacciones reportadas para humanos. También se tomó información sobre interacciones reportadas en la red. *S. cerevisiae* de la base de datos *Yeastnet* (Lee, *et al.*, 2007). Con base a esta pesquisa se extrajeron sus ortólogos de Inparanoid.

Adicionalmente se tomaron datos de tres experimentos de expresión; como criterio para la selección de las proteínas, se buscó las que habían tenido expresión diferencial; los trabajos de los cuales se seleccionó la información son:

- Fink J. et al. 2007
- Devignot S. et al. 2010

- Nascimento EJM. et al 2009

## **7.2. Modelaje de Cada uno de los Conjuntos de Datos Como una Red Individual.**

Luego de haber obtenido los conjuntos de datos, cada uno se modelo como una red individual de acuerdo a la metodología planteada por Chua *et al.*, (2007) tomando en cuenta el puntaje asociado a cada fuente de información; así, para los datos de Pfam se tomó como puntaje asociado el numero de dominios compartidos entre proteínas. Para datos de homología se tomó el logaritmo negativo del *e-value*. STRING es una base de datos que emplea un marco de puntuación único basado en un conjunto de referencia de diferentes tipos de asociaciones en contra de otro conjunto de datos común generando un puntaje de confiabilidad por predicción, el cual se utilizó como puntaje asociado a esta base de datos.

Para los datos de Intact, sin un puntaje asociado a su fuente se otorgo la puntuación de 1.

Para datos de co-expresión se determinó el coeficiente de correlación de Pearson de los genes diferencialmente expresados, siguiendo la descripción de Karaoz et al. (2004) y este se fue utilizado como puntaje asociado. Se tomaron los genes cuyo coeficiente de expresión sea mayor 0.70.

Para la información de ortólogos obtenida mediante el método de interólogos; según el cual es posible determinar la interacción de las proteínas en base a la interacción de sus ortólogas; se tomo el puntaje asociado a cada interacción, mismo que refleja la probabilidad bayesiana de que dicha interacción ocurra.

### **7.3. Integración de las Redes Individuales en una más Consistente de Acuerdo al Algoritmo Propuesto por Chua et al., (2007).**

Para integrar las redes individuales en una más consistente se siguió el método propuesto por Chua *et al.*, (2007), el cual consta de las 3 fases siguientes:

- a) Cada fuente de información fue primeramente evaluada, otorgándosele un puntaje de confiabilidad basado en su habilidad para inferir la función de una proteína en un conjunto de datos de prueba.
- b) Luego se modeló cada una como una red no direccionada con las proteínas como vértices y las relaciones entre las proteínas como enlaces.
- c) Finalmente las redes de se combinaron para formar una mas completa, con el peso de los enlaces calculado en base a la confiabilidad de cada fuente de datos. Las proteínas se predijeron con anotaciones basadas en los puntajes obtenidos por una función de voto ponderado simple que involucran solo sus vecinos en la red.

A continuación se detallan de forma más extensa las fases que comprende el método.

#### **7.3.1. Uniformización de puntaje para cada fuente de datos**

Como se señaló anteriormente, algunas fuentes de datos utilizadas tienen puntajes asociados mismos que tienen algún significado biológico. Por ejemplo, para un conjunto de datos de homología generados de un alineamiento local (BLAST), no solamente se conoce que dos proteínas con alto porcentaje de similitud forman un enlace (interaccionan) si no que también se conoce el puntaje E (*e-value*) generado por BLAST.

Las parejas de proteínas cuya similitud de secuencia tenga los menores puntajes  $E$ , tienen una mayor probabilidad de tener una función similar. Para proveer una mejor estimación del peso de cada interacción; dado un conjunto de interacciones  $F$  de una fuente de datos  $k$  donde los dos vértices de cada interacción en  $F$  tienen al menos una anotación funcional, se subdivide a  $F$  en subtipos con el siguiente procedimiento:

- Las interacciones en  $F$  son analizadas para encontrar los puntajes máximos y mínimos  $S_{k,max}$  y  $S_{k,min}$ , respectivamente.
- Las interacciones en  $F$  son ordenadas en  $n$  compartimentos  $b_1, \dots, b_n$  de intervalos iguales entre  $S_{k,max}$  y  $S_{k,min}$ .
- Cada compartimento  $b_i$  es usado como un subtipo diferente para el cual se evalúa la confiabilidad individualmente utilizando la ecuación (1);
- Para otro tipo de fuente de dato diferente a BLAST; dada una observación,  $O_{c,k,s}$ , de una interacción  $c$ , de una fuente de datos  $k$  con un puntaje su subtipo o compartimento será determinado por:

$$BinIndex_k(S) = \left\{ \min \left( n, \text{flor} \left( \left( \frac{S - S_{k,min}}{S_{k,max} - S_{k,min}} \right) \times n \right) + 1 \right) \right\} \quad (1)$$

0

si  $S \geq S_{k,min}$

si  $S < S_{k,min}$

- Si  $S \geq S_{k,min}$ , la confiabilidad de  $c$  basado en la observación  $O_{c,k,s}$  se estima por la confiabilidad del subtipo definido por  $BinIndex_k(S)$ .

- Ya que  $S_{k,min}$  se determina en datos de prueba donde ambos vértices (proteínas) están anotados, es posible que  $S$  sea menor que  $S_{k,min}$ . Si  $S < S_{k,min}$  la confianza de  $c$  basada en la observación  $O_{c,k,s}$  se toma como 0, debido a que no hay datos de prueba para estimar la confianza.

### 7.3.2. Estimación de la confianza de las fuentes de datos

Las interacciones definidas por diferentes tipos de evidencia tienen probabilidades variables de compartir una función común. Por ejemplo, las interacciones definidas por homología de secuencias pueden tener mayor probabilidad de compartir una función común que las definidas por microarreglos de expresión de genes. Aun para el mismo tipo de datos tales como los de interacciones proteína-proteína, experimentos diferentes pueden tener probabilidades distintas, debido a que están sujetos a factores como la naturaleza y procedimientos de cada experimento. La correlación con la similitud de la función no solo varía con la naturaleza de los datos, sino también con la naturaleza de la función. Por ejemplo, la similitud de secuencia indica mayormente la participación en funciones similares antes que el proceso biológico en el que actúan. Además algunas fuentes de información como se ha señalado pueden incluir puntajes. Las interacciones con diferentes puntajes pueden diferir significativamente en el grado de correlación con la similitud de funciones. Por lo tanto, las fuentes de datos pueden ser subdivididas en dos subtipos, basándose en la información disponible, tal como la fuente experimental o los puntajes asociados.

Para capturar estas variaciones, se evalúa la confianza de cada fuente, así como la de sus subconjuntos por separado para cada función. La probabilidad de que una fuente de datos  $k$  traiga la función  $f$ , se estima utilizando la ecuación (2).

$$p(k, f) = \frac{\sum_{(u,v) \in E_{kf}} S_f(u, v)}{|E_{kf}| + 1} \quad (2)$$

Donde  $E_{kf}$  es el subconjunto de interacciones de una fuente de datos  $k$  en la cual cada uno o ambos vértices tienen anotados una función  $f$ , y sus dos vértices tienen al menos una anotación funcional;

$S_f(u, v) = 1$  si  $u$  y  $v$  comparten función  $f$ , 0 en el caso contrario

Cuando  $|E_{kf}|$  es pequeño la varianza de  $p(k, f)$  es alta. Por lo tanto un pseudo-cuento de 1 se añade al denominador.

### 7.3.3. Estimación de la confianza de una interacción en la red combinada

Después de calcular la confianza de las interacciones en cada red individual que representan a cada fuente de información, estas redes pueden ser combinadas en una más completa que contiene todas las interacciones y proteínas (vértices) en los componentes de la red. Básicamente dos proteínas están conectadas en la red combinada si y solamente si esta conexión está representada en varios componentes. La fiabilidad de

cada interacción  $(u,v)$  en dicha red para cada función  $f$  se estima por los subtipos en la cual  $(u,v,k)$  es observada, como se señala en la ecuación (3).

$$r_{u,v,f} = 1 - \prod_{k \in D_{u,v}} (1 - p(k, f)) \quad (3)$$

$D_{u,v}$  es el conjunto de subtipos de fuentes de datos que contienen la interacción  $(u,v)$ .

#### 7.3.4. Asignación de un puntaje de anotación a una proteína

Para una proteína se asigna una función  $f$  en base a un puntaje  $S_f(u)$ , para lo cual se utiliza método de promedio ponderado, definido en la ecuación (4).

$$S_f(u) = \frac{\sum_{v \in N_u} (e_f(v) \times r_{u,v,f})}{1 + \sum_{v \in N_u} r_{u,v,f}} \quad (4)$$

$S_f(u)$  es el puntaje de función  $f$  para una proteína  $u$

$e_f(v) = 1$  si una proteína  $v$  tiene una función  $f$ ,  $0$  en el caso contrario

$N_u$  es el conjunto de datos que están asociadas por una interacción a una proteína  $u$

$r_{u,v,f}$  es la confianza de unión entre una proteína  $u$  y una proteína  $v$

Con base a las puntuaciones se seleccionaron las proteínas cuyo puntaje  $S_f(u)$  fue mayor 0.60; así se obtendrían proteínas con una probabilidad mayor al 60 % de estar asociadas a procesos biológicos relacionados a dengue. A este grupo de proteínas se añadieron las 473 proteínas obtenidas mediante minería de datos, creándose un nuevo grupo de proteínas de referencia al cual se denominó proteínas raíz.

#### **7.4. Construcción de la Red de Interacción de Proteínas de Humano en Respuesta a la Infección por Virus del Dengue**

Se construyó una red de interacciones únicas tomando en cuenta la información de los ocho conjuntos de datos, esta red se visualizó en la plataforma Cytoscape. Para establecer módulos funcionales se hizo un análisis de agrupamiento con el paquete MCL, para dicha plataforma; seguidamente se hizo un análisis de enriquecimiento en relación a términos de Ontologías Génicas (GO), para cada grupo establecido por el análisis con MCL; el análisis de enriquecimiento se hizo en la aplicación en línea DAVID V6.7 (Base de Datos para Aplicación, Visualización y Descubrimiento Integrado). La función para cada módulo se estableció en base a la función asociada a la mayoría de proteínas de cada grupo.

La construcción de la red específica para dengue se hizo en base a las proteínas raíz, estas se visualizaron de forma simultánea con la red combinada, permitiéndonos de esta forma establecer sus parejas interactuantes. Finalmente la red de proteínas de humano en repuesta a dengue, quedo formada por las proteínas raíz y sus parejas interactuantes; las proteínas se representaron con colores diferentes dependiendo al modulo funcional al que pertenecen.

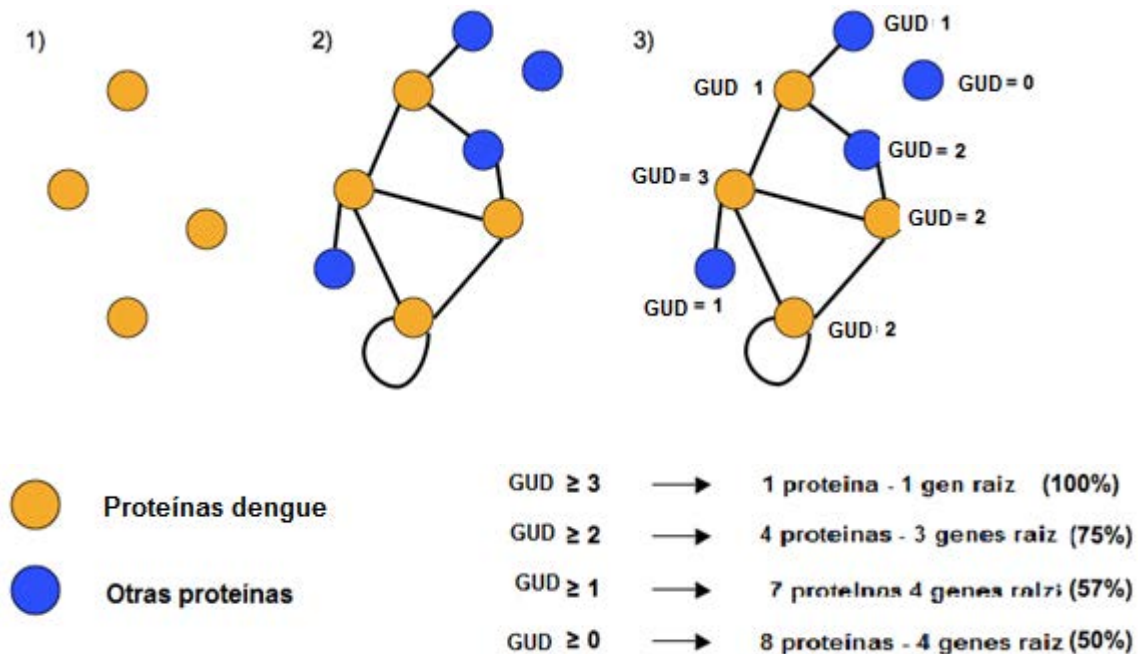
#### **7.5. Estimación de la Fiabilidad de las Predicciones del Algoritmo**

Para evaluar la confianza de las predicciones de las proteínas hechas por medio algoritmo se realizo el siguiente proceso:



### 7.5.1. Cálculo del grado de unión a dengue de una proteína GUD

Para cada una de las proteínas raíz determinadas mediante el algoritmo se calculó su grado de unión a dengue (GUD). El GUD de una proteína es igual al número de proteínas raíz con las cuales interactúan (Figura 4), excluyendo a la misma proteína (Aragues *et al.* 2008).



**Figura 4.** Cálculo del GUD de una Proteína. El GUD de una proteína es igual al número de proteínas raíz con las cuales interactúa, excluyendo a la misma proteína. El cálculo del GUD involucra tres pasos: 1) se crea una base de datos de proteínas con asociación conocida a la patogénesis; 2) se crea una red de interacción a partir de las proteínas; y 3) se calcula el GUD para cada una de las proteínas presente en la red de interacciones. En el ejemplo proporcionado, se observa que proteínas con alto valor de GUD, son más probables a estar asociadas a la respuesta a la infección en comparación a aquellas con valores bajos de GUD (Adaptado de Aragués *et al.*, 2008).

Posteriormente para evaluar la capacidad predictiva del método dado un conjunto de datos, se hizo el cálculo de las curvas ROC, estas se calcularon para cada red individual y para la red combinada. Las curvas ROC, son representaciones gráficas de la capacidad de un método de identificar, verdaderos positivos y falsos positivos en un conjunto de

datos. Para este caso se contrastaron proteínas ligadas a la respuesta a dengue versus las no ligadas; la confiabilidad se mide en relación al área bajo la curva.

El GUD, fue utilizado también como indicador de asociaciones con la respuesta a la infección por DENV, para evaluar su nivel de probabilidad estadística, se realizó el cálculo del valor predictivo positivo a diferentes umbrales de GUD ( $GUD \geq 0, \dots, GUD \geq 15$ ). El valor predictivo positivo es el número de verdaderos positivos entre los verdaderos negativos. En este sentido y teniendo como umbral de  $GUD=N$ , los verdaderos positivos serán proteínas con  $GUD \geq N$ ; por lo tanto los verdaderos positivos serán proteínas que se conoce que están involucradas a la respuesta a dengue. El análisis estadístico fue realizado mediante una tabla de contingencia y  $2\chi^2$  para la prueba estadística  $\chi^2$ . Para este análisis se utilizó el paquete RCmdr, del programa R-project.

## **7.6. Identificación de Asociaciones Funcionales de Proteínas Involucradas en la Respuesta a Dengue**

Para la identificación de asociaciones funcionales primeramente se seleccionó una lista de proteínas candidatos. Para considerar una proteína como candidato se tomó en cuenta los siguientes parámetros; que no sea una proteína raíz y que su GUD fuese mayor o igual 15; puesto que cuando se calcula el valor predictivo positivo para este umbral es superior 0.7, por lo cual puede ser considerado como un valor probabilístico óptimo.

Las proteínas que integraron la red de proteínas asociadas a la respuesta a dengue fueron sometidas a un análisis de enriquecimiento en relación a términos de Ontologías Génicas

(GO), en DAVID, de acuerdo a la función y el valor de la corrección de probabilidad de Bonferroni se seleccionaron las funciones para cada grupo de proteínas. La plataforma DAVID, al hacer un análisis de enriquecimiento, arroja un puntaje probabilístico, en base a los términos de (GO) sobre representados en un conjunto de datos, este puntaje explica la probabilidad de que la función haya podido ser asociada por azar, por lo tanto mientras mas pequeño sea el valor, esta probabilidad disminuye. Se seleccionaron las funciones asignadas para cada grupo cuyo valor probabilístico sea menor o igual  $e^{-3}$

Finalmente se seleccionaron proteínas con características importantes para ser analizadas de forma individual, teniendo en cuenta que tengan características novedosas y que puedan aportar conocimientos no documentados, sobre la respuesta a la infección.

## **8. RESULTADOS**

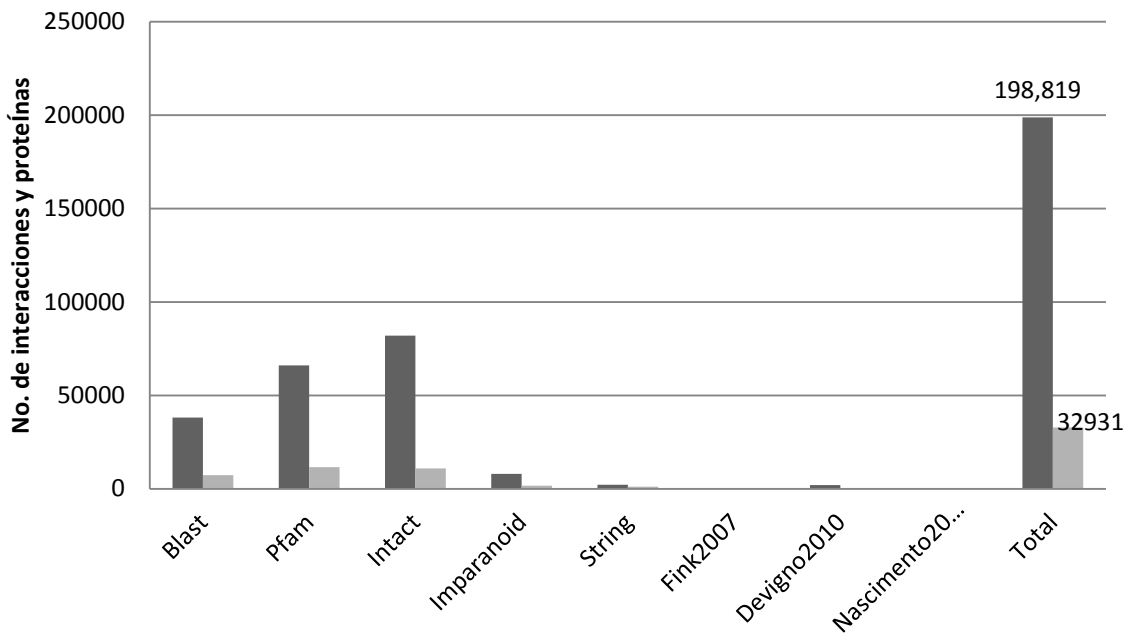
Las redes de interacciones entre proteínas, son herramientas utilizadas actualmente cada vez de forma mas frecuente para caracterizar la funcionalidad de las proteínas; debido a que en base a sus propiedades topológicas y análisis de contexto funcional podemos asignar funciones a proteínas no descritas o poco conocidas.

Diversos estudios han demostrado que proteínas asociadas aun mismo proceso biológico tienden a presentar alta conectividad entre ellas. De esta forma podemos predecir la función de una proteína, en base al principio de culpabilidad o de la mayoría; es decir en base a la función conocida de sus parejas interactuantes, o en base a la función del grupo a cual este asociado dentro de la red.

### **8.1. Proteínas de Respuesta a la Infección del Humano por el Virus del Dengue**

El trabajo desarrollado permitió la construcción una red integrada en el cual están representadas 198,819 interacciones que involucran 32,931 proteínas (Figura 5). Las bases de datos que contribuyen con un mayor número de datos de interacciones, fueron Intact y Pfam; mientras que los arreglos de expresión fueron los de menor contribución.

## Contribución de las fuentes de datos



**Figura 5.** Contribución en número de proteínas e interacciones de cada fuente de datos a la red de interacciones de proteínas asociadas a la respuesta a dengue.

Se conoce que la capacidad para predecir una interacción en una base de datos determinada, depende de la naturaleza de los datos y la naturaleza de la función a la cual esta asociada. Es decir, la probabilidad para encontrar interacciones asociadas a un proceso varía entre fuentes de datos, para capturar esta variación se calcularon los valores probabilísticos para cada base de datos y dentro de rangos.

**Cuadro 1.** Fuentes de datos y sus valores de probabilidad de acuerdo al algoritmo propuesto por Chua et al., 2007 para predecir interacciones asociadas a dengue.

<b>Intervalo</b>	<b>Probabilidad</b>
Gse17924( $0.7 \leq S < 0.85$ )	0.25
Gse17924( $0.85 \leq S < 0.95$ )	0.5
Gse18090( $0.70 \leq S < 0.85$ )	0.25
Gse18090( $0.85 \leq S < 0.95$ )	0.5
Pfam( $1 \leq S < 2$ )	0.93
Pfam( $2 \leq S < 7$ )	0.99
Imparanaoid( $1 \leq S < 3$ )	0.31
Imparanaoid( $3 \leq S < 4$ )	0.5

S representa al puntaje asociado a cada fuente de datos

En el cuadro 1 se observa que los valores predictivos para datos de Pfam se incrementan siempre que se incrementan los valores de umbral, es decir, si se hace más astringente la prueba. Para datos de expresión las probabilidades fueron bajas en relación a las otras fuentes, por lo cual, para aumentar su valor de probabilidad a 0.5 se incrementó la astringencia, tomando valores del coeficiente de correlación de Pearson mayores o iguales a 0.95.

Con estos resultados el algoritmo de Chua et al., identificó 1794 proteínas (proteínas raíz) con anotaciones de procesos biológicos asociados a la respuesta a dengue. Así, se construyó una red de alta confiabilidad que incluye a las 1794 proteínas raíz y sus parejas interactuantes, dicha red comprende 1858 interacciones y 32595 interacciones.

## **8.2. Evaluación del Desempeño del Algoritmo y Predicción de Nuevas Proteínas con Asociación a la Respuesta a Dengue**

En el presente estudio este parámetro fue calculado sobre la red integrada y utilizado con dos objetivos. El primero de ellos fue evaluar que tan bien la red integrada producida por el algoritmo de Chua et al., 2007 es capaz de identificar a los verdaderos positivos (proteínas raíz) en comparación a cada una de las redes individuales. Para ello, se construyeron una serie de curvas ROC, las cuales son ampliamente utilizadas evaluar la exactitud de una prueba diagnóstica. En la Figura 6 se observa que la red integrada es la muestra mejor desempeño en comparación a cuando se utilizaron las cada una de las redes de forma individual. Esto es soportado por su valor de área bajo la curva, el cual fue de 0.86, evidentemente mucho más alto que cada uno de los valores cuando se utilizaron las redes individuales.

Para el segundo objetivo, una vez que se demostró que la red integrada identifica mejor a las proteínas raíz el GUD fue utilizado para predecir nuevas proteínas asociadas a la respuesta a dengue. Para ello se realizó una prueba  $\chi^2$ ; de acuerdo a los resultados (Cuadro 2), se demuestra que el grado de unión a dengue es un buen indicador de la probabilidad de una proteína de estar implicada en dengue.

**Cuadro 2.** Asociación entre el contexto funcional de proteínas GUD y su conocida función en la respuesta a la infección con DENV.

Proteínas GUD>=0	Proteínas GUD>=1	p-value GUD>=0 vs. GUD>=1	Proteínas GUD>=5	p-value GUD>=1 vs. GUD>=5
1794	1607	< 2.2e-16	1232	< 2.2e-16

El valor p de la diferencia entre el conjunto de datos totales (proteínas con GUD>=0) y proteínas con GUD>=1 y GUD>=5 muestra diferencias estadísticas significativas.

Por lo tanto, utilizando el GUD, de las proteínas se calcularon los valores predictivos positivo (VPP) y con base a dichos valores predijeron nuevas proteínas involucradas en la respuesta a la infección por dengue; como criterio de selección se tomó aquellas proteínas cuyo valor predictivo fuese mayor a 0.7 (Cuadro 3.).

**Cuadro 3.** Valor predictivo positivo (VPP) para la predicción de genes asociados a la respuesta dengue basados en el GUD.

GUD	Den	NoDen	VPP
A>=0	1794	26271	0.1
B>=1	1607	15634	0.1
C>=5	1232	3798	0.2
D>=10	841	1050	0.4
E>=15	549	254	0.7
F>=20	513	120	0.8
G>=25	462	72	0.9
H>=30	441	68	0.9
I>=35	437	67	0.9
J>=40	434	67	0.9
K>=45	425	66	0.9
L>=50	422	66	0.9

GUD: Grado de unión a dengue; Den: Genes implicados en la respuesta a la infección con el DENV; NoDen: Genes no implicados en la respuesta a la infección con el DENV.



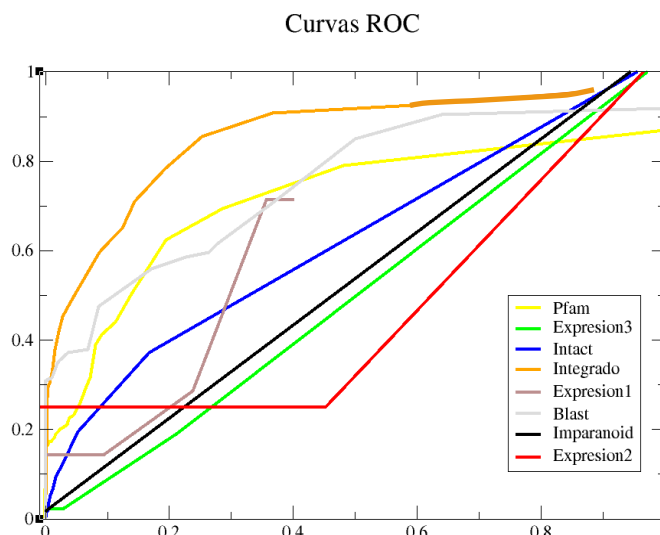
Se obtuvo un grupo final de 238 proteínas candidato. Las interacciones de estas proteínas candidato fueron localizadas en la red integrada y junto con las proteínas raíz fueron extraídas para formar la red de interacciones en respuesta a la infección por el virus del dengue. Dicha red esta conformada por 1858 proteínas y 32595 interacciones. Las proteínas candidato se sometieron a un análisis de enriquecimiento en función de términos de GO, el análisis arrojó los resultados que se muestran en el cuadro 4.

**Cuadro 4.** Anotaciones enriquecidas con relación anotaciones de participación en procesos biológicos de proteínas asociadas a la respuesta a DENV en términos de GO.

PROCESO BIOLÓGICO	TÉRMINO GEN ONTOLOGY	No Genes	%	Bonferroni
SEÑALIZACIÓN	Fosforilación de aminoácidos	41	29	2.7e-21
	Fosforilación	42	29	1.2e-19
	Proceso metabólico de fosforo	43	30	1.4e-17
	Proceso metabólico de fosfatos	43	30	1.4e-17
	Fosforilación de peptidil -tirosina	10	7.1	2.0e-8
	Modificación de peptidil-tirosina	10	7.1	2.7e-8
	Cascada proteína cinasa	16	11	5.4e-5
	Receptor transmembranal de la cascada proteína cinasa	12	8.5	2.6e-4
	Enzima ligada al receptor en la ruta de señalamiento de proteínas	14	9.9	4.4e-4
	Regulación de la señal de transducción Rho	8	5.7	1.3e-3
	Receptor de superficie celular ligado a transducción de señales	34	24	9.0e-4
	Regulador positivo de la repuesta a estímulos	11	7.8	1.7e-3
	Autofosforilación de a aminoácidos	7	5.0	3.7e-3
APOPTOSIS	Regulación de la muerte celular programada	20	14	2.2e-3
	Regulación de la muerte celular	20	14	2.2e-3
	Regulador positivo de la muerte celular programada	17	12	6.6e-5
	Regulación positiva de la muerte celular	17	12	6.4e-5
	Regulación positiva de la apoptosis	16	11	2.4e-4
ADHESIÓN CELULAR	Adhesión celular	29	20	5.7e-10
	Adhesión biológica	29	20	4.7e-10
	Adhesión célula-célula	12	8.5	1.2e-3
	Organización y proyección celular	14	9.9	8.6e-4
OTROS	Migración celular	12	8.5	1.2e-3
	Movimiento celular	15	10	2.1e-3
	Movilidad celular	12	8.5	2.6e-3
	Localización celular	12	8.5	2.6e-3
	Regulación de neurogénesis	9	6.4	3.7e-3
	Endocitosis	10	7.1	4.2e-3
	Invaginación membranal	10	7.1	4.2e-3
	Regulación de la diferenciación neuronal	8	5.7	5.5e-3
	Regulación del desarrollo del sistema nervioso	9	6.4	8.7e-3

Los principales grupos de proteínas que se formaron están asociados a tres procesos señalización, apoptosis y adhesión celular; sin embargo, también se pudieron identificar

otros procesos con menor representación, de acuerdo a los valores probabilísticos y al número de proteínas con anotaciones de GO.

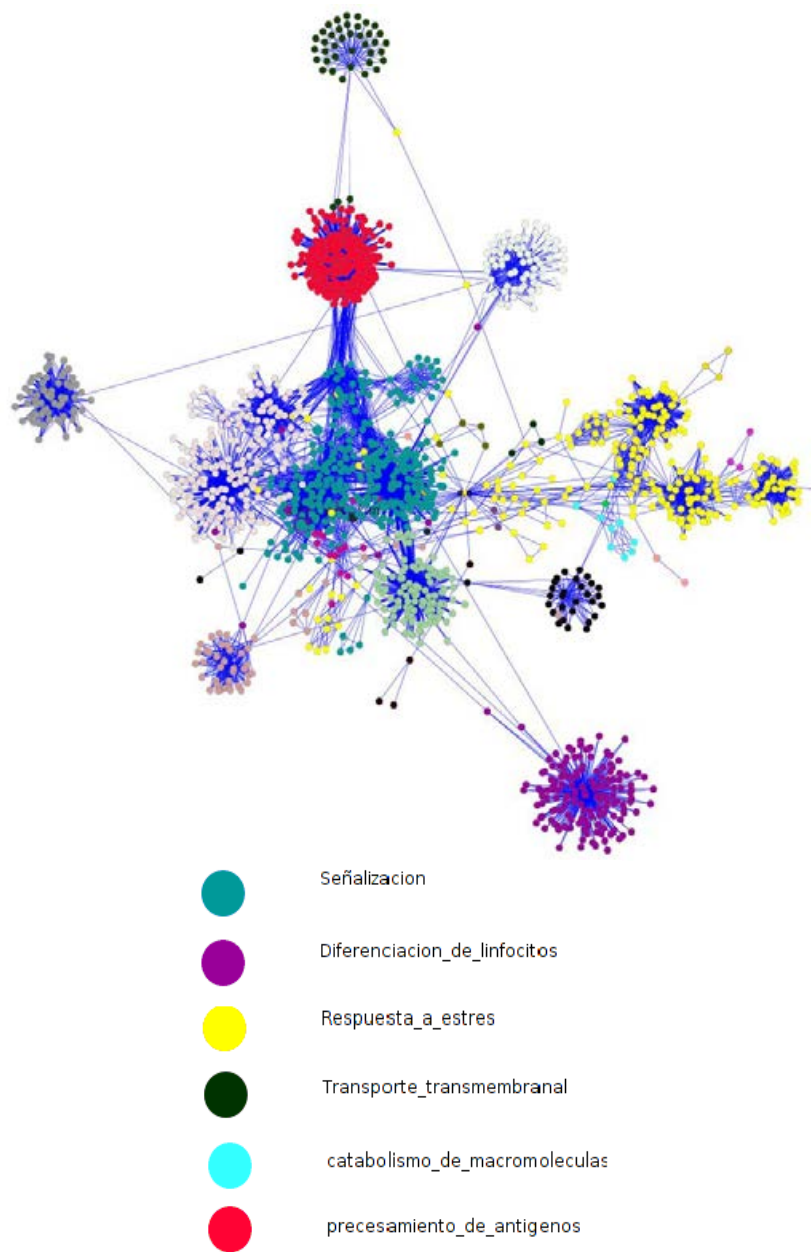


**Figura 6.** Curvas ROC construidas para evaluar la capacidad predictiva del algoritmo propuesto por Chua et al., 2007. Se presenta las curvas para cada fuente datos y para los datos integrados.

### 8.3. Análisis funcional de la Red de Interacción de Proteínas de Humano en Respuesta a la Infección con Virus del Dengue

Por medio del paquete MCL, para la plataforma Cytoscape se hizo el análisis de agrupamiento para establecer módulos funcionales, este paquete esta basado en el algoritmo de agrupamientos de Markov, este aborda la tarea de producir agrupamientos intrínsecos de datos con base en algunos índices de medida, como índices de similitud o distancia (Vlasblom y Wodak 2009). El análisis mostró la formación de varios grupos. Al hacer el análisis de estos grupos en DAVID se pudo evidenciar la tendencia de

proteínas funcionalmente relacionadas a agruparse. En la Figura 7 se muestra la red de proteínas asociadas a la respuesta a dengue, cada uno de los colores representa a una función asociada a sus anotaciones de Ontologías Génicas (GO). Los términos de GO son un conjunto de vocabularios controlados que describen un gen o atributos de un producto génico en términos de su proceso biológico, de sus componentes celulares y su función molecular (GO Consortium, 2000).



**Figura 7.** Representación gráfica de la red de interacciones de genes de humano asociados a la respuesta a la infección del DENV. Las líneas representan interacciones y los círculos a proteínas, cada uno de los colores representa una función otorgada en base a las anotaciones de GO.

Los análisis de agrupamiento y enriquecimiento muestran varios grupos de proteínas asociados a distintos procesos; los grupos funcionales que tienen una mayor

representación, dentro de la red son los asociados, a señalización, procesamiento de antígenos y anticuerpos y respuesta a estrés.

#### **8.4. Análisis Funcionales de Procesos Biológicos y Proteínas Candidato Relevantes**

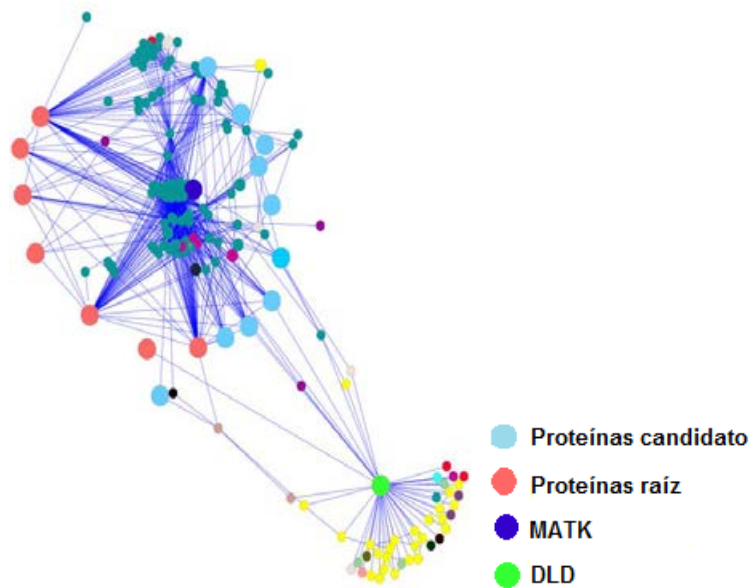
##### **8.4.1. Proteínas asociadas a división de megacariocitos**

Uno de los fenómenos particulares que se produce durante el desarrollo de la FHD y el SSD es la disminución en el conteo de plaquetas (Noisakran *et al.* 2009). González., *et al* (2011); señalan la existencia de por lo menos 30 proteínas asociadas a vías de señalización implicadas en la división de megacariocitos que son las células precursoras de plaquetas. las proteínas de dichas rutas de señalización fueron localizadas dentro de la red de proteínas de humano asociadas a la respuesta a dengue propuesta en este trabajo. Se encontraron 16 proteínas asociadas a estas vías de señalización reportadas por González., *et al* (2011), de las cuales 6 corresponden a proteínas previamente asociadas a dengue y 10 a proteínas candidatos (Tabla 5).

**Cuadro 5.** Proteínas asociadas a cascadas de señalización en la división de megacariocitos.

	<b>Código</b>	<b>Nombre</b>
Candidatos	Ffr2	Receptor 2 del factor de crecimiento de los fibroblastos
	Notch1	Proteína homóloga al locus neurogénico notch 1
	Notch2	Proteína homóloga al locus neurogénico notch 2
	Notch3	Proteína homóloga al locus neurogénico notch 3
	Akt1	Proteína Rac-alfa serina treonina cinasa
	Rasa1	Proteína activadora de la GTPasa 1
	Itk	Tirosina cinasa
	Fgr	Tirosina cinasa
Raíz	Tnfsf10	Factor de ligación de necrosis tumoral
	Pik3r1	Fosfatidilinositol cinasa 3 regulador de la subunidad alfa
	Stat3	Transductor y activador de la transcripción 3
	Stat5a	Transductor y activador de la transcripción 5
	Jak2	Janus cinasa 2
	Src	Proteína tirosina cinasa
	Epha3	Receptor 3 de efrina tipo A

De las proteínas candidato, Itk, se ha asociado se ha asociado a al síndrome linfoproliferativo causado por el virus de *Epstein-Barr* (Huck *et al.* 2009). Se puede observar en la figura 8 que estas proteínas se encuentran en un grupo asociado a señalización.



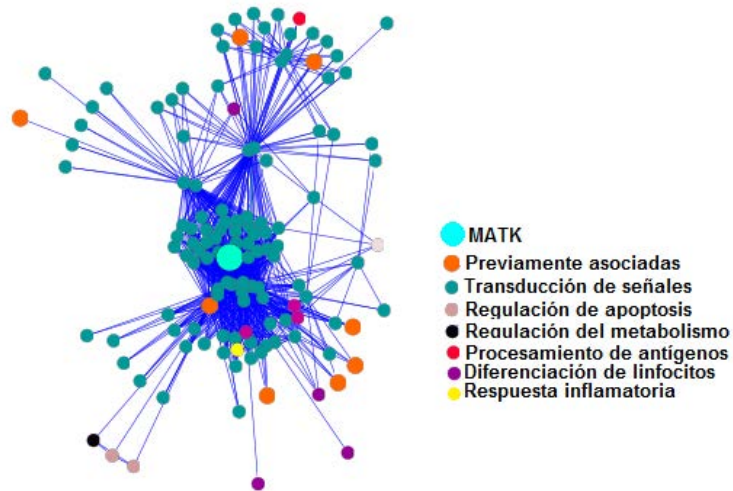
**Figura 8.** Representación gráfica de la red de interacciones de genes de humano asociados a la respuesta a la infección del DENV.

Al hacer un análisis individual con referencia en la literatura y de las propiedades topológicas de la red de proteínas asociadas a la respuesta a dengue se encontró que dos proteínas candidato con características relevantes que podrían ayudar a explicar dos fenómenos particulares en casos avanzados de dengue como lo son la trombocitopenia y eventos de autoinmunidad. Así, se realizó un análisis de las proteínas, Tirosín cinasa asociada a megacariocitos (Matk) y Dihidro lipoamida dehidrogenasa (Dld).



#### 8.4.2. Tirosina cinasa asociada a megacariocitos (Matk)

La proteína Matk se encuentra dentro del grupo de proteínas asociadas a cascadas de señalización.

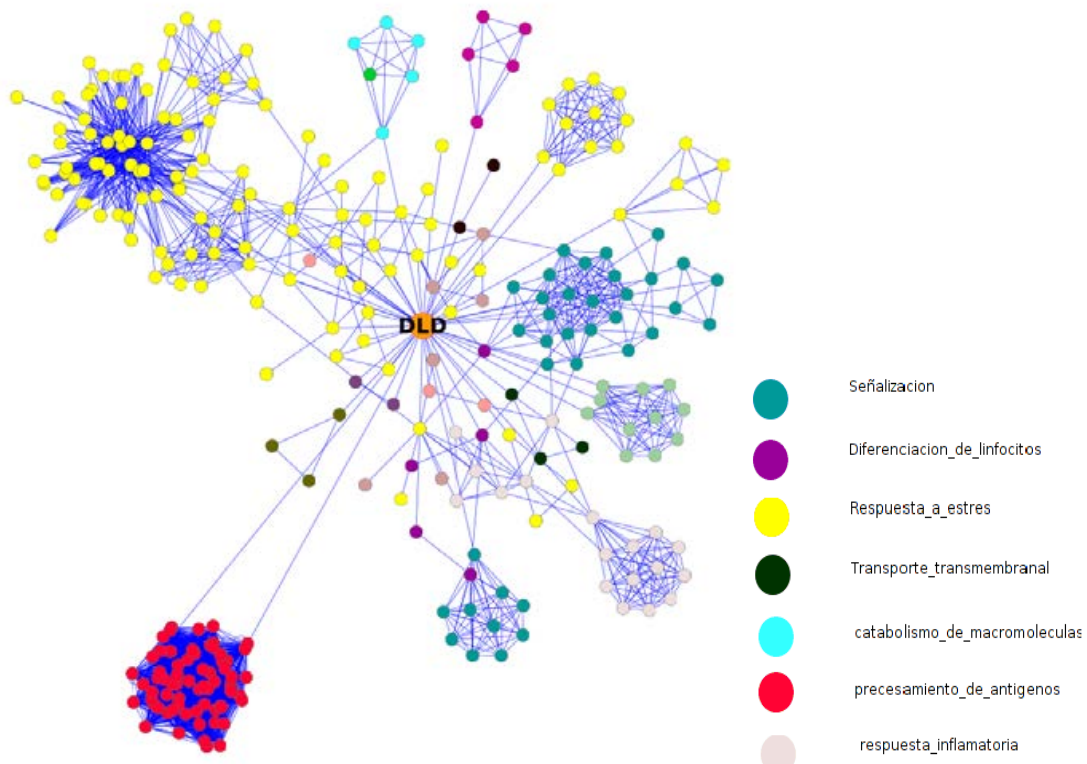


**Figura 9.** Agrupamiento de proteínas que interactúan con la Tirosina cinasa asociada a megacariocitos Matk.

MatK, fue descrita por primera vez en 1995 (Avraham *et al.* 1995); esta proteína podría jugar un rol significativo en la transducción de señales de las células hematopoyéticas. Podría regular la actividad Tirosina cinasa de los miembros de la familia SRC (Tirosina cinasas no receptoras) en el cerebro por una fosforilación de su residuo regulador tirosina el cual actúa como regulador negativo. Y también podría tener un papel como inhibidor de la proliferación de las células T (Grgurevich *et al.* 1997). En la figura 8 se puede observar que la proteína interactúa con 7 proteínas previamente asociadas a dengue de forma experimental (Khadka *et al.* 2011).

### 8.4.3. Dihidrolipoamida dehidrogenasa (Dld)

La Dihidrolipoamida dehidrogenasa es un componente del sistema de corte de las glicinas, así como también es un componente de los complejos alfa-cetoácidos.



**Figura 10.** Agrupamiento de proteínas que interactúan con la Lipoamida dehidrogenasa Dld.

A nivel de red esta proteína presenta características importantes tales como: 122 interacciones con proteínas previamente asociadas a dengue. El entrelazamiento es una de las medidas topológicas de red, esta viene dada por la tendencia de una proteína a interconectar módulos funcionales dentro de la red; Dld presenta un valor alto de entrelazamiento dentro de la red de interacción de proteínas de humano asociada a la respuesta a dengue; el valor de entrelazamiento es de 0.23; la tendencia a conectar

complejos asociados a diferentes funciones se puede observar en la figura 9. DId, dentro de la red conecta principalmente complejos asociados a respuesta estrés, procesamiento de antígenos, y transducción de señales.

## 9. DISCUSIÓN

Con el incremento de la disponibilidad de información funcional de proteínas, los métodos de predicción de sus funciones han convergido hacia la integración de fuentes de datos de diverso origen. Para este fin varios métodos han sido propuestos, los que de forma general se pueden clasificar en dos grupos; el primer grupo se enfoca en la predicción de algunas categorías funcionales, mientras que el segundo se enfocan en anotaciones funcionales más específicas basadas en los términos de GO. El principal inconveniente con estos métodos, es que emplean técnicas de aprendizaje automático, lo cual limita la normalización de cada uno de ellos para conjuntos grandes de datos, dependiendo su complejidad. La complejidad se puede referir como una combinación de: el número de proteínas a predecirse, el número de posibles anotaciones a ser predichas, el número de fuentes de datos utilizadas para las predicciones y el número de proteínas descritas para cada fuente de datos. El método aplicado para el desarrollo de la presente investigación utiliza un marco uniforme que permite la combinación de diferentes fuentes de datos. Se ha demostrado que esta combinación mejora la predicción de funciones de proteínas. Murali *et al.*, (2006) y Chua *et al.*, (2007); demostraron que este método tiene una mejor capacidad predictiva en comparación con otros métodos.

Las interacciones entre proteínas asociadas a la infección con el DENV, se han estudiado mediante métodos experimentales y métodos computacionales. Khadka y cols. (20011); analizaron mediante técnicas experimentales, proteínas asociadas al proceso de entrada del virus en células de humano. Dolittle *et al.*, 2011 predijeron mediante técnicas bioinformáticas, interacciones entre proteínas del virus del dengue y

de humano. Las predicciones se basaron en la similitud de la estructura tridimensional de las proteínas virales y las de humano. Algunas de las predicciones resultantes en este trabajo han sido reportadas previamente.; sin embargo, los resultados de los trabajos previos no son comparables debido a que la presente investigación aporta mayor información sobre las proteínas de humano, implicadas en la respuesta a la infección por el DENV.

Para el desarrollo de este trabajo se combinaron 8 fuentes de información que incluyeron: datos de expresión génica en respuesta a la infección por DENV, información de bases de datos de interacciones proteína-proteína, predicciones computacionales, contenido de dominios y similitud de secuencia. El análisis de los resultados mostro que la información relativa al contenido de dominios y similitud de secuencia fueron las que en términos probabilísticos tuvieron una mayor contribución. Mientras que los datos asociados a ensayos de co-expresión fueron los de menor contribución. Para datos de ensayos de co-expresión se utilizaron aquellos cuyo coeficiente de correlación de Pearson fuese superior 0.70; este umbral se incremento para mejorar su aporte, y al final el umbral que mejores resultados proporciono fue el de valores mayores a 0.95. Una de las principales desventajas de las diversas tecnologías de análisis a gran escala es la poca concordancia entre resultados de experimentos similares. Diversos estudios sugieren que dichas diferencias tienen que ver múltiples factores como son tal como la naturaleza y el ambiente utilizados para el experimento, y el tipo de líneas celulares utilizadas. Por lo tanto, la poca aportación encontrada de los datos de co-expresión, en comparación a las otras fuentes, podría estar relacionada a las desventajas antes citadas y al ruido inherente de los datos de expresión. Por ejemplo,

para el caso específico del dengue, la línea celular adecuada para el desarrollo de experimentos aun esta en discusión, existen experimentos que se han desarrollado en líneas de células hepáticas. Sin embargo, algunos autores sostienen que esta línea celular no es la mas adecuada, debido que existen proteínas sobre expresadas en este sistema (Khadka *et al.* 2011). Otro factor que podría afectar la contribución de los ensayos de expresión, es que a pesar de que el coeficiente de correlación de Pearson es el más aceptado para determinar la expresión de dos proteínas, se ha observado que su valor puede estar incrementado en algunos experimentos generando sesgo en los resultados (Furlotte *et al.* 2011).

Para evaluar el desempeño de la red integrada para identificar proteínas asociadas a dengue (verdaderos positivos) se elaboraron las curvas ROC (Gribskov *et al.*, 1996), dichas curvas son frecuentemente utilizadas para evaluar algunas pruebas clínicas; este tipo de gráficos se construyen en base a la capacidad de un método para distinguir los verdaderos positivos de un conjunto de datos; dicha capacidad viene representada por el área bajo la curva; como se señala en la sección de metodología, mientras el valor sea mas cercano a uno existe una mejor capacidad del método para identificar proteínas con anotaciones de procesos biológicos asociados con dengue; de acuerdo a este valor, se pueden identificar una mayor cantidad de proteínas con asociaciones a dengue cuando se combinan varias fuentes de información como se observa en la figura 6, y se corrobora con su valor de área bajo la curva (0.86). Estos resultados están en concordancia con lo previamente reportado por Chua, *et al.* 2007.

Habitualmente para caracterizar la organización de una red se realiza el cálculo de algunos parámetros topológicos como son el grado del nodo y el coeficiente de

agrupamiento; estos parámetros fueron utilizados previamente para predecir funciones (Barabasi y Oltvai 2004). Sin embargo, y a pesar de su importancia un estudio mas reciente (Aragues *et al.* 2008), demuestra que el grado del nodo no es buen predictor de función, por lo cual desarrolló el análisis de contexto funcional (GUD) como predictor de funciones de proteínas asociada a cáncer. La aplicación del análisis GUD a la red integrada permitió obtener la red de interacciones de proteínas asociadas a la respuesta a dengue; la cual quedo conformada por 1794 proteínas raíz y 238 proteínas candidatos con alta probabilidad de estar implicadas en la respuesta a dengue.

El análisis de enriquecimiento de términos de ontología génica (términos de GO ) aplicado a las proteínas que integraron la red de interacciones asociadas a la respuesta a dengue mostro términos relacionados a transducción de señales, respuesta a estrés, procesamiento de antígenos y catabolismo (Figura 7). La mayoría de proteínas candidato de acuerdo al análisis de enriquecimiento en términos de GO, son proteínas asociadas a cascadas de señalización.

La transducción de señales dentro de una célula es un complejo proceso que envuelve interacciones entre proteínas que transmiten información, así, por ejemplo, señales de la membrana plasmática pueden ser transmitidas a factores de transcripción para regular la expresión génica. De forma experimental la determinación de interacciones entre proteínas asociadas a cascadas a señalización aun tiene muchas limitaciones, lo cual se debe principalmente a su naturaleza, la mayoría de interacciones entre proteínas asociadas a señalización son transitorias, funcionan en complejos y son altamente específicas.

Dentro de las proteínas asociadas a señalización se pudo identificar varias proteínas involucradas en la división de megacariocitos que fueron reportadas previamente. González., et al., (2011); señalan la existencia de tres rutas de señalización de división megacariocítica las cuales son: megacariopoyesis dependiente de trombopoyetina (TPO); la trombopoyetina es una hormona que funciona como estimulante de la producción de trombocitos, se ha demostrado que niveles altos niveles de esta hormona incrementan los megacariocitos y plaquetas (Bartley *et al.* 1994; Kaushansky *et al.* 1994). Otra ruta de división es la que ocurre por señalización de receptores GP130, algunas citocinas de la familia de la Il6 (Iso-leucina6) que utilizan vías de señalización dependientes de la glucoproteína 130 (Gp130), como la Il11(Iso-leucina 11) y Lif (Factor inhibidor de leucemia), promueven la maduración y proliferación de megacariocitos *in vitro* y aumentan los efectos de Il3 (Iso-leucina 3) y TPO (Zheng *et al.* 2008). La tercera ruta de señalización es la asociada a las proteínas Notch, se ha demostrado que estas proteínas promueven la trombopoyesis sobre la eritropoyesis.

La importancia de las rutas de señalización asociadas al linaje megacariocítico dentro de la respuesta a dengue es evidente, debido a que algunos estudios han reportado la existencia de trombocitopenia en estados graves de la enfermedad (Noisakran *et al.* 2009). Este fenómeno puede ser ocasionado por dos factores la disminución en la producción en la producción de plaquetas y la destrucción de estas (Oishi *et al.* 2007). La trombocitopenia también se ha relacionado con la activación del sistema de complemento. Esta anomalía en las plaquetas podría incrementar el riesgo de fragilidad vascular que lleva a la hemorragia; lo cual podría contribuir a la pérdida de plasma observado durante el desarrollo de FHD/SSD.



Se pudo identificar 15 proteínas asociadas a megacariopoyesis (González *et al.* 2011) dentro de la red generada; de las cuales *Tnfr10*, *Pik3r1*, *Stat3*, *Stata5a*, *Jak2*, *Src* y *Epha3* se habían asociado previamente a dengue; mientras que, las proteínas *Fgfr2*, *Notch1*, *Notch2* y *Notch3*, *akt1*, *Rasa1*, *Itk* *Fgr*, son proteínas candidato.

*Fgr2* es el receptor del factor de crecimiento de los fibroblastos, esta proteína incrementa la producción de *Tpo*, que como se ha señalado estimula producción de plaquetas. *Notch1*, *Notch2* y *Notch3* están asociadas a una ruta de señalización para la división megacariopoyética (Zheng *et al.* 2008). *Akt1* promueve la sobrevivencia y la entrada al ciclo celular de los megacariocitos. *Rasa1* esta asociada a una ruta de señalización (*Ras*/*MAPK*) que estimula la maduración la formación, proliferación, adhesión y agregación de plaquetas (Genever *et al.* 1999; Hitchcock *et al.* 2003; Skerry y Genever 2001). *Itk* es una proteína asociada a señalización de quimiocinas, su rol exacto en la división de megacariocitos aun no se ha establecido. *Fgr* es una tirosina cinasa asociada a la activación de plaquetas (Robinson *et al.* 2000). De las proteínas candidato señaladas *Fgr*, se ha asociado a las respuesta al virus sarcoma felino; *Rasa1* y *Notch1* se han asociado entre otros virus al de la hepatitis C.

Dentro de las proteínas candidato asociadas a vías de señalización; la tirosina cinasa asociada a megacariocitos (*Matk*), es una proteína cuyas interacciones podrían potencialmente ayudar a explicar la anomalía en la producción de plaquetas durante la fase grave del dengue. Estudios funcionales de *Matk* en ratones han demostrado que el bloqueo del gen que codifica para esta cinasa inhibe significativamente la producción de células progenitoras de los megacariocitos (Avraham *et al.* 1995). Dentro de la red generada *Matk*, (Figura 8) interactúa con las proteínas *Tnfr10*, *Itk*, *Jak2*, *Akt1*, *Rasa1*,

Fgfr2, y Notch3, implicadas en la ruta de señalización de la megacariopoyésis, también interactúa con EphA3, Stat3, Stat5a, Pik31; que son proteínas que se han reportado como asociadas a dengue y también asociados a división de megacariocitos. Otra interacción relevante que se pudo identificar dentro de la red de es con Src; este es un proto-oncogen presente en rutas de señalización de megacariopoyesis y se tiene la noción de que este podría actuar como regulador del crecimiento del linaje megacariocítico (Avraham *et al.* 1995).

La sobre-representación términos de GO, de proteínas asociadas a señalización concuerda con los resultados presentados obtenidos en el trabajo experimental desarrollado por Khadka *et al.*, (2011), quienes encontraron sobrerrepresentadas proteínas asociadas a coagulación; proceso en el cual las plaquetas juegan un rol protagónico. Así mismo, las rutas de señalización asociadas a coagulación se han asociado a la formas severas de dengue (Martina *et al.* 2009).

Otra proteína analizada de forma particular es la Dihidrolipamida deshidrogenasa (Dld), es una enzima codificada por un gen mitocondrial; es un componente de el complejo alfa cetoacídico y del sistema de ruptura de glicina, alteraciones en esta enzima se han ligado con enfermedades asociadas al metabolismo (Roche y Reed 1974). Dentro del análisis topológico de la red de proteínas de humano en respuesta a dengue, Dld presenta el valor más alto de entrelazamiento 0.32 e interacciona con 122 proteínas las cuales habían sido implicadas con la respuesta dengue. El valor de entrelazamiento de una proteína refleja la capacidad de esta de conectar diversos grupos dentro de una red, como se observa en la figura 7, Dld interconecta módulos de proteínas asociados a funciones como procesamiento de antígenos, señalización, respuesta estrés entre otros.

Navratil *et al.*, (2011) señala que proteínas de humano que concentran gran número de interacciones y con valores altos de entrelazamiento son blancos predilectos de las proteínas virales.

Como se ha señalado DId, interactúa dentro de la red con proteínas asociadas a señalización, de forma particular se analizaron las interacciones con proteínas de la ruta de señalización de la megacariopoyesis. La disminución en el conteo de plaquetas (trombocitopenia) no es una característica exclusiva de los casos graves de dengue, también se han identificado en infecciones de otro virus como VIH y VHC. La disminución en la producción de plaquetas y la destrucción de estas son dos causas de la trombocitopenia; la destrucción de plaquetas se cree que está asociado a eventos de autoinmunidad. Los eventos autoinmunes relacionados con dengue, en especial los asociados con la disminución de plaquetas se estudiaron experimentalmente en ratones, se encontró que la presencia de un anticuerpo contra la proteína viral NS1, reduce el número de plaquetas; además se señala que puede existir una reacción cruzada de anticuerpos (Sun *et al.* 2007).

La relación entre DId y la autoinmunidad en DENV, aun no es clara. Sin embargo, se ha reportado que esta proteína está asociada con la respuesta a la infección por VHC y se han encontrado anticuerpos anti-DId en el suero sanguíneo de pacientes con casos graves de hepatitis C (Wu *et al.* 2002). Esto cobra mayor importancia si consideramos que más del 40 % de proteínas implicadas en la infección por DENV, están implicadas en la infección con VHC (Khadka *et al.* 2011)

## 10. CONCLUSIONES

En el presente trabajo se encontraron 238 nuevos genes o proteínas de humano con alta probabilidad de estar implicados a la respuesta a la infección por el virus del dengue.

Las principales conclusiones que se desprenden del presente proyecto incluyen:

- Se construyó una red de alta confiabilidad de proteínas de humano la cual ésta conformada por 32,931 proteínas y 198,819 interacciones.
- La red integrada es la que en términos probabilísticos permite identificar de forma precisa proteínas asociadas a la respuesta a dengue.
- El análisis del contexto funcional (GUD), resultó ser un buen predictor para identificar nuevas asociaciones funcionales de proteínas que no habían sido relacionadas con anterioridad a dengue. Dicho análisis permitió obtener una red de proteínas asociadas a la respuesta a dengue que incluye 1858 proteínas y 32595 interacciones.
- Se encontraron 16 proteínas previamente implicadas en la producción de plaquetas de las cuales 6 ya se habían asociado a dengue y 9 son proteínas candidato resultado de este trabajo. Este hallazgo es importante debido a que la disminución en el conteo de plaquetas es una característica propia de la fase grave del dengue.
- La probabilidad que proteína Matk, de estar implicada en la respuesta a la infección por DENV, es alta, especialmente si se considera que esta implicación

podría ayudar a explicar el proceso de trombocitopenia en las etapas graves de la enfermedad.

- El análisis topológico de la red sugiere que la proteína Dld juega un papel importante en la respuesta a dengue con posible implicación en fenómenos de autoinmunidad. Esta proteína no había sido asociada previamente a dengue. Sin embargo existe evidencia que asocia a Dld con la respuesta a la infección por VHC y un posible evento de autoinmunidad.

## 11. REFERENCIAS BIBLIOGRÁFICAS

- Aragués R, Sander C, and Oliva B. 2008. Predicting cancer involvement of genes from heterogeneous data. *BMC Bioinformatics* 9:172.
- Avraham S, Jiang S, Ota S, Fu Y, Deng B, Dowler LL, White RA, and Avraham H. 1995. Structural and functional studies of the intracellular tyrosine kinase *MATK* gene and its translated product. *J Biol Chem* 270(4):1833-1842.
- Barabasi AL, and Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5(2):101-113.
- Bartenschlager R, and Miller S. 2008. Molecular aspects of Dengue virus replication. *Future Microbiol* 3(2):155-165.
- Bartley TD, Bogenberger J, Hunt P, Li YS, Lu HS, Martin F, Chang MS, Samal B, Nichol JL, Swift S et al. . 1994. Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor *Mpl*. *Cell* 77(7):1117-1124.
- Bielefeldt-Ohmann H. 2000. Measuring virulence without a target. *Trends Microbiol* 8(6):265-266.
- Burke DS, and Kliks S. 2006. Antibody-dependent enhancement in dengue virus infections. *J Infect Dis* 193(4):601-603; author reply 603-604.
- Clyde K, Kyle JL, and Harris E. 2006. Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J Virol* 80(23):11418-11431.
- Cusick ME, Klitgord N, Vidal M, and Hill DE. 2005. Interactome: gateway into systems biology. *Hum Mol Genet* 14 Spec No. 2:R171-181.

- Chatr-aryamontri A, Kerrien S, Khadake J, Orchard S, Ceol A, Licata L, Castagnoli L, Costa S, Derow C, Huntley R et al. . 2008. MINT and IntAct contribute to the Second BioCreative challenge: serving the text-mining community with high quality molecular interaction data. *Genome Biol* 9 Suppl 2:S5.
- Chua HN, Sung WK, and Wong L. 2007. An efficient strategy for extensive integration of diverse biological data for protein function prediction. *Bioinformatics* 23(24):3364-3373.
- de Chasse B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaogue S, Meiffren G, Pradezynski F, Faria BF, Chantier T et al. . 2008. Hepatitis C virus infection protein network. *Mol Syst Biol* 4:230.
- Dirección General de Epidemiología, Secretaría de Salud. Panorama Epidemiológico (México). México, DF: DGE/SS, 2008.
- Doolittle JM, and Gomez SM. 2011. Mapping protein interactions between Dengue virus and its human and insect hosts. *PLoS Negl Trop Dis* 5(2):e954.
- Dutta AK, Biswas A, Baruah K, and Dhariwal AC. 2011. National guidelines for diagnosis and management of dengue fever/dengue haemorrhagic fever and dengue shock syndrome. *J Indian Med Assoc* 109(1):30-35.
- Ge H, Walhout AJ, and Vidal M. 2003. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet* 19(10):551-560.
- Gautier VW, Gu L, O'Donoghue N, Pennington S, Sheehy N, and Hall WW. 2009. In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology* 6:47
- Genever PG, Wilkinson DJ, Patton AJ, Peet NM, Hong Y, Mathur A, Erusalimsky JD, and Skerry TM. 1999. Expression of a functional N-methyl-D-aspartate-type

- glutamate receptor by bone marrow megakaryocytes. *Blood* 93(9):2876-2883.
- Gmeiner WH, Hellmann GM, and Shen P. 2008. Tissue-dependent and -independent gene expression changes in metastatic colon cancer. *Oncol Rep* 19(1):245-251.
- Gonzalez Morales JE, Leal Villarreal L, Guzman Lopez S, Guzman Trevino GM, and Gonzalez Martinez NA. 2004. [Helicobacter pylori and disease]. *Rev Alerg Mex* 51(6):218-225.
- Grgurevich S, Linnekin D, Musso T, Zhang X, Modi W, Varesio L, Ruscetti FW, Ortaldo JR, and McVicar DW. 1997. The Csk-like proteins Lsk, Hyl, and Matk represent the same Csk homologous kinase (Chk) and are regulated by stem cell factor in the megakaryoblastic cell line MO7e. *Growth Factors* 14(2-3):103-115.
- Gubler D. 2001. Dengue and West Nile virus--an interview with Duane Gubler, Sc.D., reported by Vicki Glaser. *Vector Borne Zoonotic Dis* 1(1):81-88.
- Gubler DJ. 1987. Dengue and dengue hemorrhagic fever in the Americas. *P R Health Sci J* 6(2):107-111.
- Guha-Sapir D, and Schimmer B. 2005. Dengue fever: new paradigms for a changing epidemiology. *Emerg Themes Epidemiol* 2(1):1.
- Guzman MG. 2005. Global voices of science. Deciphering dengue: the Cuban experience. *Science* 309(5740):1495-1497.
- Halstead SB. 1988. Pathogenesis of dengue: challenges to molecular biology. *Science* 239(4839):476-481.



- Harris E, Videa E, Perez L, Sandoval E, Tellez Y, Perez ML, Cuadra R, Rocha J, Idiaquez W, Alonso RE et al. . 2000. Clinical, epidemiologic, and virologic features of dengue in the 1998 epidemic in Nicaragua. *Am J Trop Med Hyg* 63(1-2):5-11.
- Hitchcock IS, Skerry TM, Howard MR, and Genever PG. 2003. NMDA receptor-mediated regulation of human megakaryocytopoiesis. *Blood* 102(4):1254-1259.
- Ho LJ, Wang JJ, Shaio MF, Kao CL, Chang DM, Han SW, and Lai JH. 2001. Infection of human dendritic cells by dengue virus causes cell maturation and cytokine production. *J Immunol* 166(3):1499-1506.
- Holden KL, Stein DA, Pierson TC, Ahmed AA, Clyde K, Iversen PL, and Harris E. 2006. Inhibition of dengue virus translation and RNA synthesis by a morpholino oligomer targeted to the top of the terminal 3' stem-loop structure. *Virology* 344(2):439-452.
- Huck K, Feyen O, Niehues T, Ruschendorf F, Hubner N, Laws HJ, Telieps T, Knapp S, Wacker HH, Meindl A et al. . 2009. Girls homozygous for an IL-2-inducible T cell kinase mutation that leads to protein deficiency develop fatal EBV-associated lymphoproliferation. *J Clin Invest* 119(5):1350-1358.
- Jeong H, Mason SP, Barabasi AL, and Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* 411(6833):41-42.
- Kaushansky K, Lok S, Holly RD, Broudy VC, Lin N, Bailey MC, Forstrom JW, Buddle MM, Oort PJ, Hagen FS et al. . 1994. Promotion of megakaryocyte

progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin. *Nature* 369(6481):568-571.

Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, Sengupta R, Sahasrabudhe S, Randall G, Gribskov M et al. . 2011. A physical interaction network of dengue virus and human proteins. *Mol Cell Proteomics* 10(12):M111 012187.

Khan NA, Azhar EI, El-Fiky S, Madani HH, Abuljadial MA, Ashshi AM, Turkistani AM, and Hamouh EA. 2008. Clinical profile and outcome of hospitalized patients during first outbreak of dengue in Makkah, Saudi Arabia. *Acta Trop* 105(1):39-44.

Kielian M. 1995. Membrane fusion and the alphavirus life cycle. *Adv Virus Res* 45:113-151.

Kouri GP, Guzman MG, Bravo JR, and Triana C. 1989. Dengue haemorrhagic fever/dengue shock syndrome: lessons from the Cuban epidemic, 1981. *Bull World Health Organ* 67(4):375-380.

Kroeger A, and Nathan MB. 2006. Dengue: setting the global research agenda. *Lancet* 368(9554):2193-2195.

Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG et al. . 2002. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* 108(5):717-725.

- Lim J, Hao T, Shaw C, Patel AJ, Szabo G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE et al. . 2006. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 125(4):801-814.
- Lu YE, Cassese T, and Kielian M. 1999. The cholesterol requirement for sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. *J Virol* 73(5):4272-4278.
- Mackenzie JS, Gubler DJ, and Petersen LR. 2004. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nat Med* 10(12 Suppl):S98-109.
- Marianneau P, Steffan AM, Royer C, Drouet MT, Jaeck D, Kirn A, and Deubel V. 1999. Infection of primary cultures of human Kupffer cells by Dengue virus: no viral progeny synthesis, but cytokine production is evident. *J Virol* 73(6):5201-5206.
- Martina BE, Koraka P, and Osterhaus AD. 2009. Dengue virus pathogenesis: an integrated view. *Clin Microbiol Rev* 22(4):564-581.
- McBride WJ, and Bielefeldt-Ohmann H. 2000. Dengue viral infections; pathogenesis and epidemiology. *Microbes Infect* 2(9):1041-1050.
- Noisakran S, Chokeyhaibulkit K, Songprakhon P, Onlamoon N, Hsiao HM, Villinger F, Ansari A, and Perng GC. 2009. A re-evaluation of the mechanisms leading to dengue hemorrhagic fever. *Ann N Y Acad Sci* 1171 Suppl 1:E24-35.
- O'Brien KP, Remm M, and Sonnhammer EL. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* 33(Database issue):D476-480.

- Oishi K, Saito M, Mapua CA, and Natividad FF. 2007. Dengue illness: clinical features and pathogenesis. *J Infect Chemother* 13(3):125-133.
- Oyama T, Kitano K, Satou K, and Ito T. 2002. Extraction of knowledge on protein-protein interaction by association rule discovery. *Bioinformatics* 18(5):705-714.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, and Yeates TO. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96(8):4285-4288.
- Puig B, Tortosa A, and Ferrer I. 2001. Cleaved caspase-3, caspase-7 and poly (ADP-ribose) polymerase are complementarily but differentially expressed in human medulloblastomas. *Neurosci Lett* 306(1-2):85-88.
- Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B et al. . 2007. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* 39(11):1338-1349.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J et al. . 2012. The Pfam protein families database. *Nucleic Acids Res* 40(Database issue):D290-301.
- Rico-Hesse R, Harrison LM, Salas RA, Tovar D, Nisalak A, Ramos C, Boshell J, de Mesa MT, Nogueira RM, and da Rosa AT. 1997. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* 230(2):244-251.

- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, and Seraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17(10):1030-1032.
- Robinson DR, Wu YM, and Lin SF. 2000. The protein tyrosine kinase family of the human genome. *Oncogene* 19(49):5548-5557.
- Roche TE, and Reed LJ. 1974. Monovalent cation requirement for ADP inhibition of pyruvate dehydrogenase kinase. *Biochem Biophys Res Commun* 59(4):1341-1348.
- Segura C. A. 2009. Predicción y análisis computacional de nuevas interacciones entre proteínas que participan en la patogénesis de *Helicobacter pylori*. Tesis de maestría. Centro de Biotecnología Genómica, Instituto Politécnico Nacional. Reynosa-Tamulipas. México. 14p.
- Shoemaker BA, and Panchenko AR. 2007a. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput Biol* 3(3):e42.
- Shoemaker BA, and Panchenko AR. 2007b. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol* 3(4):e43.
- Skerry TM, and Genever PG. 2001. Glutamate signalling in non-neuronal tissues. *Trends Pharmacol Sci* 22(4):174-181.
- Smit JM, Bittman R, and Wilschut J. 1999. Low-pH-dependent fusion of Sindbis virus with receptor-free cholesterol- and sphingolipid-containing liposomes. *J Virol* 73(10):8476-8484.}

- Sun DS, King CC, Huang HS, Shih YL, Lee CC, Tsai WJ, Yu CC, and Chang HH. 2007. Antiplatelet autoantibodies elicited by dengue virus non-structural protein 1 cause thrombocytopenia and mortality in mice. *J Thromb Haemost* 5(11):2291-2299.
- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguetz P, Doerks T, Stark M, Muller J, Bork P et al. . 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 39(Database issue):D561-568
- Thein S, Aung MM, Shwe TN, Aye M, Zaw A, Aye K, Aye KM, and Aaskov J. 1997. Risk factors in dengue shock syndrome. *Am J Trop Med Hyg* 56(5):566-572.
- Vidal M, Cusick ME, and Barabasi AL. 2011. Interactome networks and human disease. *Cell* 144(6):986-998.
- Vlasblom J, and Wodak SJ. 2009. Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC Bioinformatics* 10:99.
- von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, and Bork P. 2005. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33(Database issue):D433-437.
- von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, and Bork P. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417(6887):399-403.

- Wachi S, Yoneda K, and Wu R. 2005. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 21(23):4205-4208.
- Wu SF, Liao CL, Lin YL, Yeh CT, Chen LK, Huang YF, Chou HY, Huang JL, Shaio MF, and Sytwu HK. 2003. Evaluation of protective efficacy and immune mechanisms of using a non-structural protein NS1 in DNA vaccine against dengue 2 virus in mice. *Vaccine* 21(25-26):3919-3929.
- Wu YY, Hsu TC, Chen TY, Liu TC, Liu GY, Lee YJ, and Tsay GJ. 2002. Proteinase 3 and dihydrolipoamide dehydrogenase (E3) are major autoantigens in hepatitis C virus (HCV) infection. *Clin Exp Immunol* 128(2):347-352.
- Wu SJ, Paxton H, Hanson B, Kung CG, Chen TB, Rossi C, Vaughn DW, Murphy GS, and Hayes CG. 2000. Comparison of two rapid diagnostic assays for detection of immunoglobulin M antibodies to dengue virus. *Clin Diagn Lab Immunol* 7(1):106-110.
- Xu T, Sampath A, Chao A, Wen D, Nanao M, Chene P, Vasudevan SG, and Lescar J. 2005. Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 Å. *J Virol* 79(16):10278-10288.
- Zheng C, Yang R, Han Z, Zhou B, Liang L, and Lu M. 2008. TPO-independent megakaryocytopoiesis. *Crit Rev Oncol Hematol* 65(3):212-222.
- Zuo Z, Liew OW, Chen G, Chong PC, Lee SH, Chen K, Jiang H, Pua CM, and Zhu W. 2009. Mechanism of NS2B-mediated activation of NS3pro in dengue virus: molecular dynamics simulations and bioassays. *J Virol* 83(2):1060-1070.

## 12. APÉNDICE

Lista de proteínas candidato con alta probabilidad de estar implicadas en la respuesta a la infección por el virus del dengue.

<b>Id. Uniprot</b>	<b>Nombre</b>	<b>GUD</b>
A2AAT3	Complejo mayor de histocompatibilidad clase II, B	125
B0S7Z5	Complejo mayor de histocompatibilidad clase I, C	124
B0S7V1	Complejo mayor de histocompatibilidad clase II, B	123
B0S7Z4	Complejo mayor de histocompatibilidad clase I, C	123
B0S7Z6	Complejo mayor de histocompatibilidad clase I, C	123
A2ABE5	Complejo mayor de histocompatibilidad clase I, C	122
A2ABE6	Complejo mayor de histocompatibilidad clase I, C	122
A2ABE7	Proteína no caracterizada	122
A2AEA2	Proteína no caracterizada	122
A2AEA4	Proteína no caracterizada	122
P09622	Dihidrolipoamida dehidrogenasa	122
A2AAR9	Proteína no caracterizada	121
A2ACI6	Proteína no caracterizada	121
A2ARK8	Complejo de histocompatibilidad mayor clase I, A	121
A2BDY9	Complejo mayor de histocompatibilidad clase I, A	121
A2BER6	Secuencia relacionada con el complejo mayor de histocompatibilidad	121
A2BF24	Complejo mayor de histocompatibilidad clase I, C	121
A2BF25	Complejo mayor de histocompatibilidad clase I, C	121
A2BF26	Complejo mayor de histocompatibilidad clase I, C	121
A8MVU7	Proteína no caracterizada	121
A9R9N7	Complejo de mayor de histocompatibilidad clase I, B	121
A9R9N8	Complejo mayor de histocompatibilidad clase I, A	121
B0UXQ0	Antígeno MHC clase I	121
B0UXQ1	Complejo mayor de histocompatibilidad clase I, A	121
A2AAS2	Complejo mayor de histocompatibilidad clase I, B	113
A2ACF0	Complejo mayor de histocompatibilidad clase I, F	113
A2ACI5	Complejo mayor de histocompatibilidad clase I, F	113
A2AEA1	Complejo mayor de histocompatibilidad clase I, B	113
A2AQM5	Proteína no caracterizada	113
A2BF29	Proteína no caracterizada	113
B0S7Z9	Proteína no caracterizada	113
B0S895	Proteína no caracterizada	113
B0S7Y6	Complejo mayor de histocompatibilidad clase II	77
B0S7Y9	Complejo mayor de histocompatibilidad clase II, D	73
A2AAY8	Proteína no caracterizada	72
A2AAZ0	Proteína no caracterizada	72
A2ADX3	Complejo mayor de histocompatibilidad clase II	72
A2ALJ5	Complejo mayor de histocompatibilidad clase II, D	72
A2ARK5	Complejo mayor de histocompatibilidad clase II, D	72
A2BEU3	Complejo mayor de histocompatibilidad clase II, D	72



A2BFX2	Complejo mayor de histocompatibilidad clase II, D	72
A2BFX3	Complejo mayor de histocompatibilidad clase II, D	72
B0S7Z0	Complejo mayor de histocompatibilidad clase II, D	72
B0UXD8	Antígeno MHC clase 2	72
A2ALJ3	Complejo mayor de histocompatibilidad clase II, DP	71
A2BET9	Complejo mayor de histocompatibilidad clase II, DP	71
A2AAT4	Proteína no caracterizada	64
A2AAT5	Proteína no caracterizada	64
A2AAT8	Unidad alfa del complejo mayor de histocompatibilidad clase II	64
A2AB90	Tapasina	64
A2ABB9	Tapasina	64
A8MUD0	Proteína no caracterizada	64
B0S7N3	Complejo mayor de histocompatibilidad clase II, DM	64
B0S7N4	Complejo mayor de histocompatibilidad clase II, DM	64
A6YF59	RAET1E2	57
E7ES57	Complejo mayor de histocompatibilidad clase I, B	56
P13611	Proteína núcleo Versican	34
Q96GW7	Proteína núcleo Brevican	29
P16112	Proteína núcleo Aggrecan	28
O14594	Proteína núcleo Neurocan	26
P00519	Proteína-tirosina cinasa ABL1	24
Q8WXD0	Receptor de la relaxina 2	24
Q9UBG0	Receptor tipo manosa 2	24
B2R6Q4	Isoforma Src	23
B7UEF3	Proteína-tirosina cinasa Abelson	23
B7UEF4	Proteína-tirosina cinasa Abelson	23
B7UEF5	Proteína-tirosina cinasa Abelson	23
P41240	Proteína-tirosina cinas CSK	23
P42680	Proteína-tirosina cinasa Tec	23
P42684	Proteína-tirosina cinasa Abelson 2	23
Q06187	Proteína-tirosina cinasa BTK	23
Q86WK6	Proteína inductora de anfoterina 1	23
B3KNZ9	Tirosina cinas asociada a megacariocitos	22
P07948	Proteína-tirosina cinasa Lyn	22
P09769	Proteína-tirosina cinasa Fgr	22
P12931	Proto-oncogen proteína-tirosina cinasa Src	22
P35968	Receptor del factor de crecimiento endotelial	22
P42679	Tirosina cinas asociada a megacariocitos	22
P42685	Proteína tirosina cinasa FRK	22
Q13018	Receptor 2 de la fosfolipasa secretoria	22
Q5TGA3	Peptidil-prolil isomerasa E	22
Q9HBX9	Receptor de relaxina 1	22
Q9NST8	Tirosina cinas asociada a megacariocitos	22
Q9UNP9	Peptidil-prolil isomerasa E	22
B4DHD1	Receptor tipo relaxina	21

O94813	Proteína homóloga slit2	21
P08631	Proteína tirosina cinasa HCK	21
P16591	Proteína-tirosina cinasa Fer	21
P30530	Receptor de la proteína-tirosina cinasa UFO	21
P42681	Proteína tirosina cinasa TXK	21
P51813	Proteína-tirosina cinasa citoplasmática BMx	21
P98175	Proteína de unión al RNA 10	21
Q08881	Proteína-tirosina cinasa ITK/TSK	21
Q3KU23	LGR8.1	21
Q4LE67	Variante proteína Neurican	21
Q5T0V0	Slit	21
Q92626	Homóloga a peroxidasa	21
Q9HBX8	Repeticiones ricas en leucina	21
O75093	Proteína homóloga Slit 1	20
O75473	Receptor de acoplamiento 49 de la proteína G	20
P06239	Proteína-tirosina cinasa Lck	20
P15498	Proto-oncogen vav	20
P51451	Proteína-tirosina cinasa Blk	20
Q573B4	Tirosina cinasa específica de linfocitos	20
Q5JY90	Tirosina cinasa agamaglobulina	20
Q5T0V1	Slit	20
Q6PJG9	Repeticiones ricas en leucina y dominio tipo fibronectina 3	20
Q96IN1	Tirosina cinasa linfoide	20
A2A284	Fosfolipasa C	19
A4D0T1	Repetición rica en leucina, neuronal	19
P31749	Proteína-treonina cinasa alfa serina-RAC	19
P43405	Proteína-tirosina cinasa SYK	19
P52756	Proteína de unión al RNA 5	19
Q13882	Proteína-tirosina cinasa 6	19
Q15746	Proteína relacionada con cinasas	19
Q5VST9	Obscurina	19
Q6UY18	Proteína neuronal con repeticiones ricas en leucina	19
Q9BTN0	Similar a la molécula de adhesión sináptica 4	19
Q9BXB1	Receptor de acoplamiento 48 de la proteína G	19
Q9H3W5	Proteína neuronal con repeticiones ricas en leucina 3	19
Q9H3Y6	Proteína-tirosina cinasa Srms	19
Q9NPY3	Receptor de un factor de complemento	19
Q9P2V4	Fotoreceptor asociado a la superfamilia de proteínas LRR	19
Q9ULH4	Similar a la molecular de adhesión sináptica 1	19
O75094	Proteína homóloga Slit 3	18
P16581	Selectina-E	18
P19174	Fosfolipasa C-gama-1	18
P43403	Proteína-tirosina cinasa ZAP-70	18
P52735	Factor de intercambio de guanina 2	18
Q5TGY6	Proteína asociada al sarcoma viral felino	18

Q5VSN0	Dominio SH3	18
Q8WZ42	Titina	18
Q9UKW4	Factor de intercambio de guanina 3	18
Q9UMR7	Dominio tipo lectina C familia 4 miembro A	18
Q9UPX0	Miembro de la sueprfamilia de las inmunoglobulina	18
A6NHA5	Dominio tipo lectina C familia 4 miembro D	17
A6NNE8	Peptidil-prolil isomerasa E	17
B0LM41	RBM14	17
B0LPF3	Proteína de unión al factor de complemento	17
B4DTL2	Proteína activadora de la GTPasa	17
O14936	Proteína periférica de la membrana plasmatica CASK	17
O60229	Kalirina	17
O60449	Antígeno linfocítico 75	17
O60500	Nefrina	17
O75962	Proteína con dominios triple funcionales	17
O94898	Repeticiones ricas en leucina y dominios tipo inmunoglobulina	17
P0C6S8	Proteína neuronal con repeticiones ricas en leucina	17
P20936	Proteína 1 activadora de la GTPasa Ras	17
P23945	Receptor de la hormona foliculo estimulante	17
P46531	Proteína homologa a locus neurogenico Notch 1	17
P62993	Proteína de unión al receptor del factor de crecimiento	17
Q05AH0	Receptor de la hormona foliculo estimulante	17
Q38SD2	Proteína-treonina cinasa 1	17
Q4VXU2	Proteína de unión a poliadenilato 1	17
Q5KU26	Colectina 12	17
Q5S007	Proteína-treonina cinasa 2	17
Q5TAX3	Uridiltransferasa terminal 4	17
Q5TI72	Selectina E	17
Q5TI73	Selectina E	17
Q5TI74	Selectina E	17
Q5TI75	Selectina E	17
Q8IUN9	Dominio tipo lectina C familia 10 miembro A	17
Q8TBF4	Dedos de Zinc	17
Q8WTT0	Dominio tipo lectina C familia 4 miembro C	17
Q8WXI8	Dominio tipo lectina C familia 4 miembro D	17
Q9HBW1	Repetición rica en leucina	17
Q9HCU0	Endosialina	17
Q9P232	Contactina 3	17
Q9UHP7	Dominio tipo lectina C familia 2 miembro D	17
Q9ULY5	Dominio tipo lectina C familia 4 miembro F	17
A4QPA5	Dominio tipo lectina familia 4 miembro F	16
A6NMB1	Lectina para unión a ácido sialico tipo inmunoglobulina G	16
A6NNE7	Peptidil-prolil isomerasa E	16
O43699	Proteína de unión al ácido sialico 6	16
O75325	Proteína neuronal con repeticiones ricas en leucina 2	16

P06734	Receptor de membrana FC de la inmunoglobulina E de baja afinidad	16
P07332	Proteína-tirosina cinasa Fes/Fps	16
P14151	Selectina-L	16
P16885	Fosfolipasa C-gama-2	16
P98161	Policistina 1	16
Q01974	Receptor transmembranal de proteínas tirosina cinasas ROR2	16
Q04721	Proteína homologa a locus <b>neurogénico</b> Notch 2	16
Q08ET2	Proteína de unión a ácido sialico 14	16
Q13588	Proteína adaptador relacionada con GBRB2	16
Q14160	Proteína homologa Scribble	16
Q16288	Receptor del factor de crecimiento NT-3	16
Q2HXU8	Dominio tipo lectina C familia 12 miembro B	16
Q3SXY7	Repeticiones ricas en leucina, dominio tipo inmunoglobulina	16
Q6EIG7	Dominio tipo lectina C familia 6 miembro A	16
Q6UXM1	Repeticiones ricas en leucina y dominios tipo inmunoglobulina	16
Q6UXN8	Dominio tipo lectina C familia 9 miembro A	16
Q7L3H5	Dominio cinasa LIM	16
Q8IXC7	Dominio tipo lectina C familia 10 miembro A	16
Q8N1N0	Dominio tipo lectina C familia 4 miembro F	16
Q92478	Dominio tipo lectina C familia 2 miembro B	16
Q96JA1	Repeticiones ricas en leucina y dominios tipo inmunoglobulina	16
Q96LC7	Proteína de unión al ácido sálico 10	16
Q96RL6	proteína de unión al ácido sálico 11	16
Q9HCJ2	Ligando netrina G1	16
Q9NYZ4	proteína de unión al ácido sialico8	16
Q9UJ43	Selectina-L	16
Q9UM47	Peroxidasina tipo proteina	16
A1KZ92	Peroxidasina tipo proteina	15
A8K4G3	Cinasa de células hematopoyéticas	15
A9NIU9	Peptidil-prolil isomerasa E	15
B4DIP2	proteína similar a LAP2	15
B4DUE3	Cinasa de miosina de cadena liviana	15
B4E0T7	proteína no caracterizada	15
B5MC51	Dominio cinasa LIM	15
O43639	proteína citoplasmática NCK2	15
O60674	proteína-tirosina cinasa JAK2	15
P05129	proteína-cinasa C tipo g	15
P05452	Tetranectina	15
P06731	Antígeno carcinoembriónico relacionado con adhesión celular molécula 5	15
P0C7M8	Dominio tipo lectina C familia 2 miembro L	15
P13688	Antígeno carcinoembriónico relacionado con adhesión celular molécula 1	15
P16333	proteína citoplasmática NCK1	15
P31997	Antígeno carcinoembriónico relacionado con adhesión celular molécula 8	15
P40199	Antígeno carcinoembriónico relacionado con adhesión celular molécula 6	15
Q01973	Receptor transmembranal de proteínas tirosina cinasas ROR1	15

Q08G24	Antígeno carcinoembrionario relacionado a adhesión celular	15
Q0Z7S6	Antígeno carcinoembrionario relacionado a adhesión celular	15
Q14449	Receptor del factor de crecimiento 14	15
Q4G0T1	Receptor con dominios ricos en cisteína	15
Q53ZX0	Antígeno CD69	15
Q68CU6	proteína no caracterizada	15
Q6UVW9	Dominio tipo lectina C familia 2 miembro A	15
Q6UXB4	Dominio tipo lectina C familia 4 miembro C	15
Q7L985	Repetición rica en leucina y dominio similar a inmunoglobulina	15
Q96E93	Receptor tipo lectina subfamilia G miembro 1	15
Q96FE5	Repetición rica en leucina	15
Q96II8	Repetición rica en leucina y dominio homólogo a calponina	15
Q96NW7	Densina 180	15
Q9BZZ2	Sialoadesina	15
Q9NR99	proteína asociada a la remodelación de la matriz celular	15
Q9UN19	Adaptador dual fosfotirosina	15
Q9Y624	Molécula de adhesión A	15
Q9Y6Z7	Colectina 10	15

---