

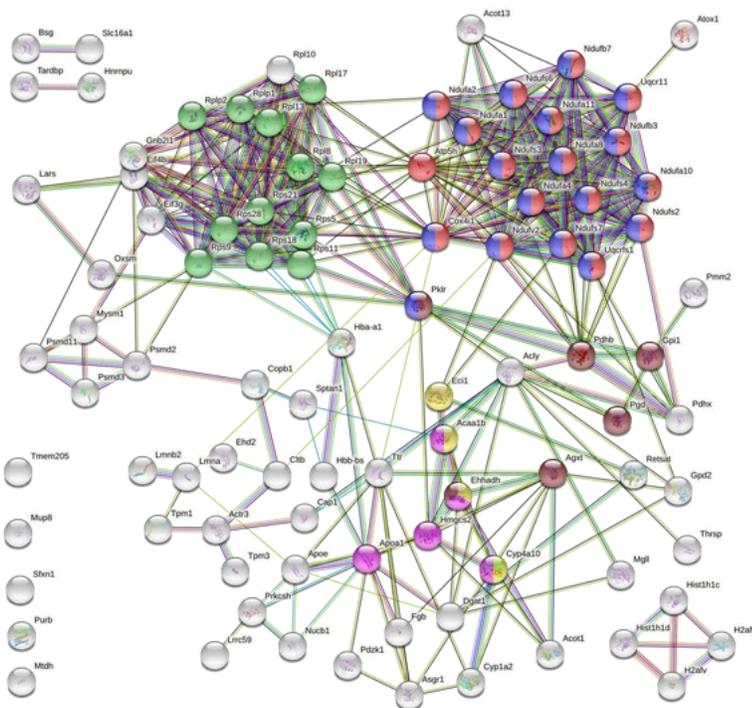
UNIVERSIDAD DE
GUANAJUATO



XXVI
de Verano
de la Ciencia
Universidad de Guanajuato



Análisis bioinformático.



Ana Sinaí Pérez Hernández; Eder Jahir Nieva Ruiz; Karla San Elías Landin; Olga Regina De La Torre Morrill; Oscar Gómez Rodríguez; Oscar Gerardo Silva Gaona; Victoriano Pérez Vázquez.

Universidad de Guanajuato

Introducción:

El análisis de sistemas moleculares, conocimiento de la función, interacción y comprensión de los procesos biológicos involucrados es una importante tarea que permite una mejor la comprensión de un organismo. La interpretación de todo el conjunto de datos que implica un compendio de proteínas, genes; es una tarea compleja y de gran volumen, que puede llegar a ser muy lenta y caótica si no se cuenta con la guía y herramientas adecuadas para llevarla a cabo ([Szklarczyk, 2017](#)).

Comenzaremos con que evaluar las interacciones que ocurren entre proteínas puede resultar complicado pues se trata de una gran cantidad de información y variables a tomar en cuenta, cómo su estructura molecular, función molecular, interacciones proteína-proteína ([Szklarczyk, 2015](#)); pues las proteínas entre sí pueden influir en la producción y en la vida media de cada una de las formas transcripcional y postranscripcional, intercambiar productos de reacción, participar en mecanismos de retransmisión de señales o contribuir conjuntamente a funciones específicas del organismo (Rivas, 2004). La información de estas interacciones se encuentra dispersa en múltiples recursos y plataformas, lo que dificulta la tarea de su estudio, igualmente elegir las plataformas correctas es toda una gran tarea.

Para facilitar la tarea de integrar la gran cantidad de interacciones entre proteínas utilizaremos la herramienta virtual Search Tool for the Retrieval of Inter-acting Genes/Proteins (STRING) que permite asociar las proteínas por asociación *funcional*, es decir, una relación funcional específica y productiva entre dos proteínas, probablemente contribuyendo a un propósito biológico común ([Szklarczyk, 2015](#)). Este recurso electrónico se vale de información generada por los usuarios y de otras bases de datos.

Finalmente, tras analizar las interacciones proteicas, las clasificaremos de acuerdo con sus asociados biológicos utilizando la base de datos Database for Annotation, Visualization and Integrated Discovery (DAVID) que permite el análisis de gran número de genes (Huang, 2009) usando un sistema integrado de base de datos back-end (Sherman, 2007) y un algoritmo modular avanzado en la recopilación de datos (Huang, 2007).

Objetivo:

Realizar un análisis bioinformático de una lista de genes o proteínas. Utilizar herramientas como la base de datos STRING que permite conocer y predecir interacciones funcionales proteína- proteína. Igualmente, de la herramienta de análisis funcional DAVID que nos facilita analizar el cambio de expresión proteica, y finalmente mapear los genes en una lista de asociados biológicos.

Definiciones

Bioinformática: es un campo de las ciencias computacionales que lleva a cabo el análisis de secuencias de moléculas biológicas. Normalmente se aplica a los genes, al ADN, al ARN, o a las proteínas, y resulta especialmente útil para comparar secuencias de genes y proteínas entre distintos organismos, pudiendo ver las relaciones evolutivas entre organismos, e intentando averiguar cuál es la función de dichos genes y proteínas.

STRING: Search Tool for the Retrieval of Inter-acting Genes/Proteins

Base de datos de interacciones proteína-proteína conocidas y predichas. Las interacciones incluyen asociaciones directas (físicas) e indirectas (funcionales); surgen de la predicción computacional, de la transferencia de conocimientos entre organismos y de interacciones agregadas de otras bases de datos.

DAVID: Database for Annotation, Visualization, and Integrated Discovery

Materiales:

- Computadora con conexión a internet
- Search Tool for the Retrieval of Inter-acting Genes/Proteins (STRING)
- Functional Database for Annotation, Visualization, and Integrated Discovery DAVID
- Lista de genes o proteínas

Metodología.

1. Ingresar compendio de datos a STRING

Cargar la lista de proteínas o genes en la base de datos Search Tool for the Retrieval of Inter-acting Genes/Proteins (STRING) v. 11.0 en la sección “multiple proteins” (se permite ingresar el gene, símbolo de la proteína o el nombre de la proteína), y seleccionar un organismo de estudio.

2. Lista de proteínas identificadas.

La plataforma genera una pequeña previsualización donde se pueden observar las proteínas identificadas y una pequeña revisión de estas, de igual forma muestra proteínas con gen homólogo y permite seleccionar entre ellas para trabajar con el que se necesite.

The following proteins in *Mus musculus* appear to match your input. Please review the list, then click 'Continue' to proceed.

77 query items showing page 1 of 4 • first • previous • next • last

1) **Tardbp**:

Tardbp - TAR DNA-binding protein 43; DNA and RNA-binding protein which regulates transcription and splicing. Involved in the regulation of CFTR splicing. It promotes CFTR exon 9 skipping by binding to the UG repeated motifs in the polymorphic region near the 3'-splice site of this exon. The resulting aberrant splicing is associated with pathological features typical of cystic fibrosis. May also be involved in microRNA biogenesis, apoptosis and cell division. Can repress HIV-1 transcription by binding to the HIV-1 long terminal repeat. Stabilizes the low molecular weight neurofilament (NFL) mRNA [...]

2) **Ndufa4**:

Ndufa4 - Cytochrome c oxidase subunit **NDUFA4**; Cytochrome c oxidase (COX, complex IV) is the terminal component of the mitochondrial respiratory chain that catalyzes the reduction of oxygen to water. Required for complex IV maintenance (By similarity)

Ndufa4l2 - NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2; Belongs to the complex I **NDUFA4** subunit family [a.k.a. *XP_006513898.1*, *ENSMUSG00000040280*, *BAE20749.1*]

3) **Lmnb2**:

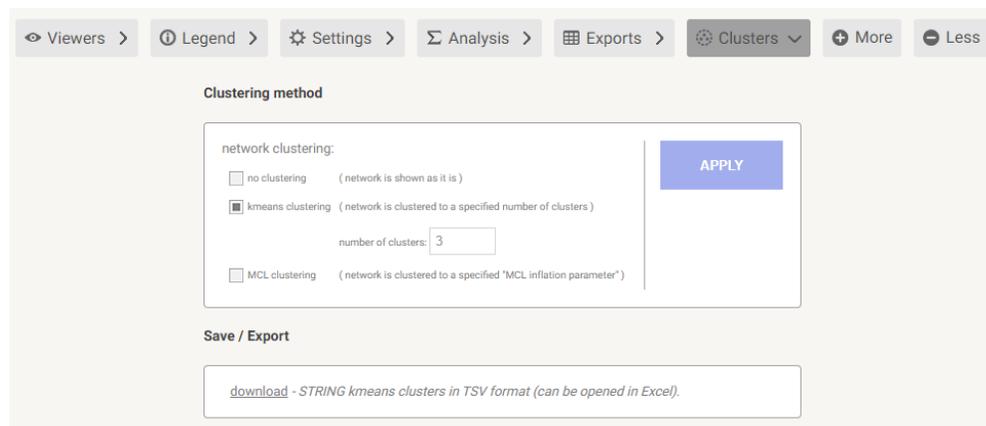
Lmnb2 - Lamin-B2; Lamins are components of the nuclear lamina, a fibrous layer on the nucleoplasmic side of the inner nuclear membrane, which is thought to provide a framework for the nuclear envelope and may also interact with chromatin

4) **Pdzk1**:

Pdzk1 - Na(+)/H(+) exchange regulatory cofactor NHE-RF3; A scaffold protein that connects plasma membrane proteins and regulatory components, regulating their surface expression in epithelial cells apical domains. May be involved in the coordination of a diverse range of regulatory processes for ion transport and second messenger cascades. In complex with SLC9A3R1, may cluster proteins that are functionally dependent in a mutual fashion and modulate the trafficking and the activity of the associated membrane proteins. May play a role in the cellular mechanisms associated with multidrug resistanc [...]

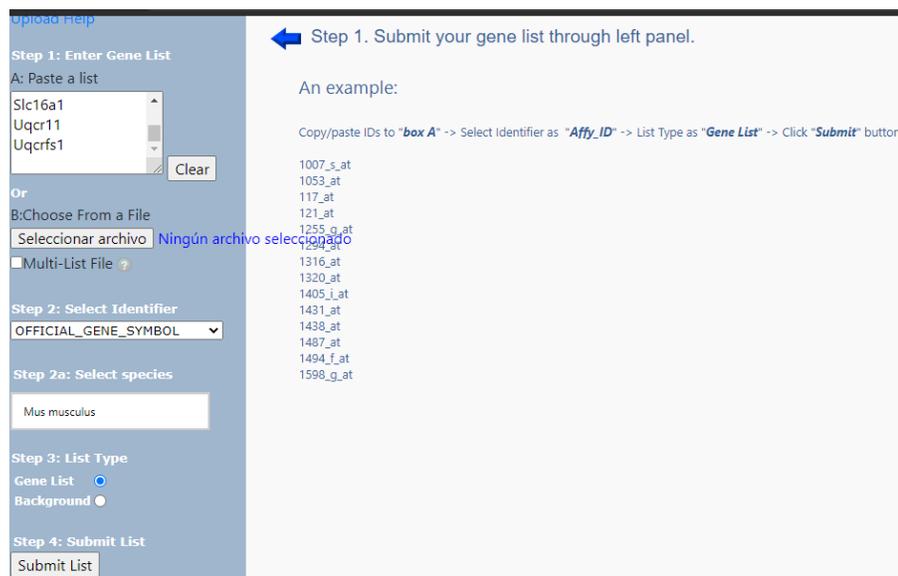
Pdzk1ip1 - **PDZK1** interacting protein 1 [a.k.a. *Map17*, *ENSMUST00000106548*, *Pdzk1ip1-005*]

Terminado nuestra primera parte del análisis, seleccionar la pestaña "clúster". En esta pestaña obtener la lista de proteínas identificadas y enriquecidas generadas del procesamiento de STRING.



6. Ingreso de datos en DAVID

Cargar el compendio de proteínas obtenido de STRING (se pueden abrir con Excel) en la herramienta de anotación funcional Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.8. en la pestaña "star analysis". El programa acepta una gran variedad de símbolos, en este caso utilizar "Official Gene Symbol" porque es el generado por STRING, finalmente seleccionar nuestro organismo de estudio.



7. Análisis de los datos.

Acceder en la herramienta de anotación funcional de DAVID, desplegar la sección de ontología génica

The screenshot shows the DAVID Analysis Wizard interface. On the left is the Gene List Manager, and on the right is the Analysis Wizard. The Gene List Manager shows a list of species with 'Mus musculus(95)' selected. The Analysis Wizard shows Step 1: 'Successfully submitted gene list' and Step 2: 'Analyze above gene list with one of DAVID tools'. Below Step 2, there are several tool options, including 'Functional Annotation Tool', 'Gene Functional Classification Tool', 'Gene ID Conversion Tool', and 'Gene Name Batch Viewer'. The 'Functional Annotation Tool' is selected, and its sub-options are visible: 'Functional Annotation Clustering', 'Functional Annotation Chart', and 'Functional Annotation Table'. Below the wizard, there is a summary of the current gene list and background, followed by a table of GO terms and their associated statistics, with horizontal bar charts for each term.

Gene List Manager

Select to limit annotations by one or more species [Help](#)

- Use All Species -
Mus musculus(95)
Unknown(1)

Select Species

List Manager [Help](#)

List_1

Select List to:
Use Rename
Remove Combine
Show Gene List

Analysis Wizard

Tell us how you like the tool
Contact us for questions

Step 1. Successfully submitted gene list
Current Gene List: List_1
Current Background: Mus musculus

Step 2. Analyze above gene list with one of DAVID tools
Which DAVID tools to use?

[Functional Annotation Tool](#)
• [Functional Annotation Clustering](#)
• [Functional Annotation Chart](#)
• [Functional Annotation Table](#)

[Gene Functional Classification Tool](#)
[Gene ID Conversion Tool](#)
[Gene Name Batch Viewer](#)

Current Gene List: List_1
Current Background: Mus musculus
95 DAVID IDs
Check Defaults Clear All

Functional Categories (3 selected)
 Gene Ontology (3 selected)

Category	Percentage	Count	Chart
<input type="checkbox"/> GOTERM_BP_1	95.8%	91	Chart
<input type="checkbox"/> GOTERM_BP_2	95.8%	91	Chart
<input type="checkbox"/> GOTERM_BP_3	95.8%	91	Chart
<input type="checkbox"/> GOTERM_BP_4	86.3%	82	Chart
<input type="checkbox"/> GOTERM_BP_5	84.2%	80	Chart
<input type="checkbox"/> GOTERM_BP_ALL	95.8%	91	Chart
<input checked="" type="checkbox"/> GOTERM_BP_DIRECT	95.8%	91	Chart
<input type="checkbox"/> GOTERM_BP_FAT	95.8%	91	Chart
<input type="checkbox"/> GOTERM_CC_1	100.0%	95	Chart
<input type="checkbox"/> GOTERM_CC_2	98.9%	94	Chart
<input type="checkbox"/> GOTERM_CC_3	98.9%	94	Chart
<input type="checkbox"/> GOTERM_CC_4	98.9%	94	Chart
<input type="checkbox"/> GOTERM_CC_5	97.9%	93	Chart
<input type="checkbox"/> GOTERM_CC_ALL	100.0%	95	Chart
<input checked="" type="checkbox"/> GOTERM_CC_DIRECT	100.0%	95	Chart
<input type="checkbox"/> GOTERM_CC_FAT	98.9%	94	Chart
<input type="checkbox"/> GOTERM_MF_1	94.7%	90	Chart
<input type="checkbox"/> GOTERM_MF_2	92.6%	88	Chart
<input type="checkbox"/> GOTERM_MF_3	90.5%	86	Chart
<input type="checkbox"/> GOTERM_MF_4	88.4%	84	Chart

8. Obtención de resultados obtenidos de DAVID

Analizar las proteínas de acuerdo con los proceso biológico, función molecular y componente celular, igualmente con las vías metabólicas de señalización e interacciones mediante KEGG (Por sus siglas en inglés Kyoto Encyclopedia of Genes and Genomes). Acceder a los genes involucrados en cada elemento en el botón “chart” de cada categoría. El programa permite filtrar aún más los datos por número de genes involucrados y su valor estadístico.

Tablas de ontología génica y vías

Functional Annotation Chart

[Help and Manual](#)

Current Gene List: List_1

Current Background: Mus musculus

95 DAVID IDs

Options

Rerun Using Options

Create Sublist

16 chart records

[Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_DIRECT	transport	RT		24	25.3	2.1E-5	2.6E-3
<input type="checkbox"/>	GOTERM_BP_DIRECT	oxidation-reduction process	RT		23	24.2	1.4E-12	7.0E-10
<input type="checkbox"/>	GOTERM_BP_DIRECT	translation	RT		14	14.7	9.1E-8	2.3E-5
<input type="checkbox"/>	GOTERM_BP_DIRECT	lipid metabolic process	RT		12	12.6	1.7E-5	2.6E-3
<input type="checkbox"/>	GOTERM_BP_DIRECT	metabolic process	RT		9	9.5	2.2E-3	1.2E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	fatty acid metabolic process	RT		7	7.4	1.3E-4	1.3E-2
<input type="checkbox"/>	GOTERM_BP_DIRECT	cytoplasmic translation	RT		4	4.2	6.4E-4	5.4E-2
<input type="checkbox"/>	GOTERM_BP_DIRECT	response to cAMP	RT		4	4.2	2.1E-3	1.2E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	carbohydrate metabolic process	RT		4	4.2	8.3E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	mitochondrial electron transport, NADH to ubiquinone	RT		3	3.2	1.6E-3	1.1E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	triglyceride catabolic process	RT		3	3.2	4.4E-3	2.2E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	acyl-CoA metabolic process	RT		3	3.2	1.0E-2	4.7E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	cholesterol biosynthetic process	RT		3	3.2	1.1E-2	4.7E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	fatty acid beta-oxidation	RT		3	3.2	2.0E-2	7.9E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	receptor-mediated endocytosis	RT		3	3.2	3.6E-2	8.3E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	chromatin silencing	RT		3	3.2	3.9E-2	8.3E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	response to nutrient	RT		3	3.2	5.1E-2	9.5E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	fatty acid biosynthetic process	RT		3	3.2	5.3E-2	9.5E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	steroid metabolic process	RT		3	3.2	6.6E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	cholesterol metabolic process	RT		3	3.2	7.3E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	nucleosome assembly	RT		3	3.2	9.5E-2	1.0E0
<input type="checkbox"/>	GOTERM_BP_DIRECT	cellular oxidant detoxification	RT		2	2.1	2.5E-2	8.3E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	regulation of Cdc42 protein signal transduction	RT		2	2.1	2.5E-2	8.3E-1
<input type="checkbox"/>	GOTERM_BP_DIRECT	lipoprotein biosynthetic process	RT		2	2.1	2.9E-2	8.3E-1

Gráficos

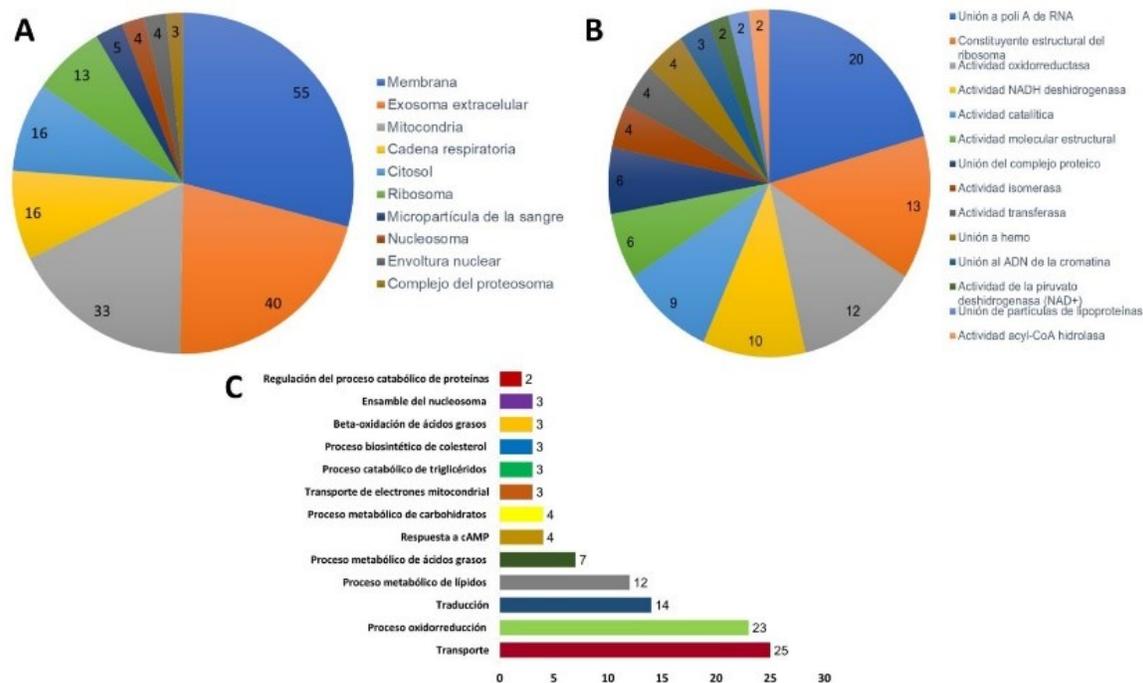


Diagrama de flujo del procedimiento:

