

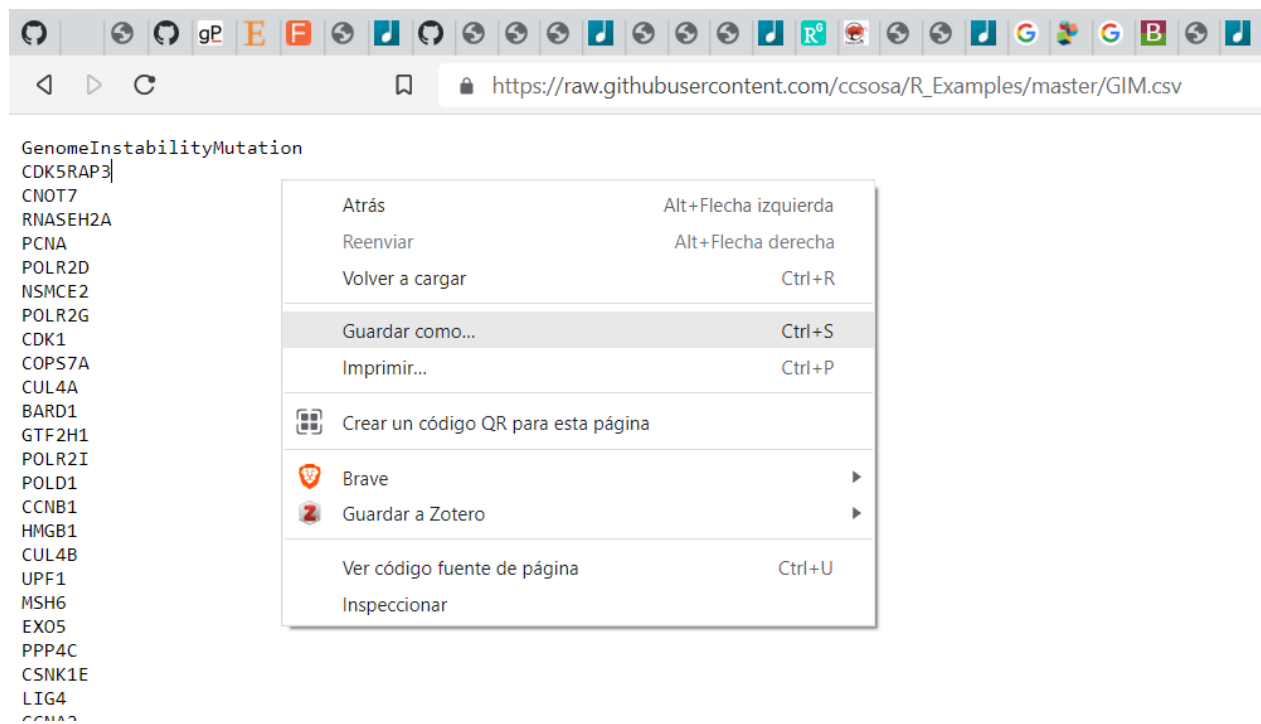
Guía para la realización de un análisis de enriquecimiento funcional usando una lista de genes

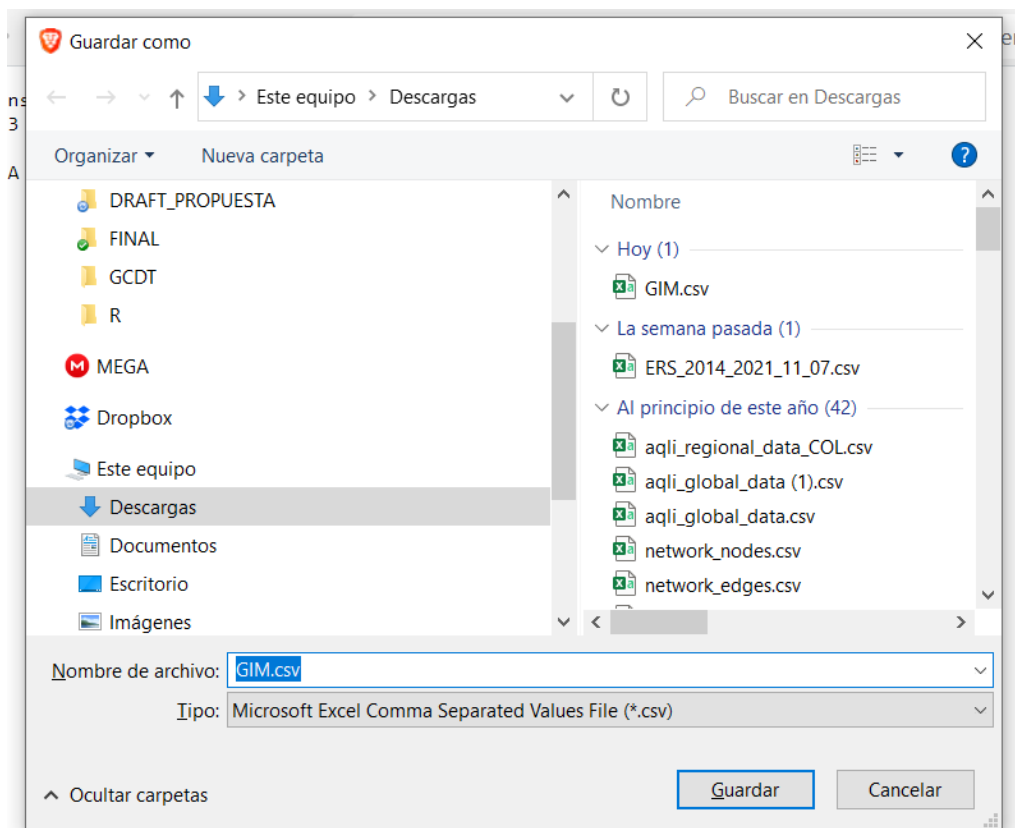
Simposio anual Ómicas - Modalidad talleres

1. Descarga de archivos necesarios:

En esta sesión se realizará análisis de enriquecimiento funcional para una única lista de genes.

- Ingrese en su navegador el siguiente enlace: https://raw.githubusercontent.com/ccsosa/R_Examples/master/GIM.csv. Posteriormente, dé clic derecho y elija la opción Guardar como... para empezar la descarga del archivo GIM.csv en su computador local como se muestra a continuación:





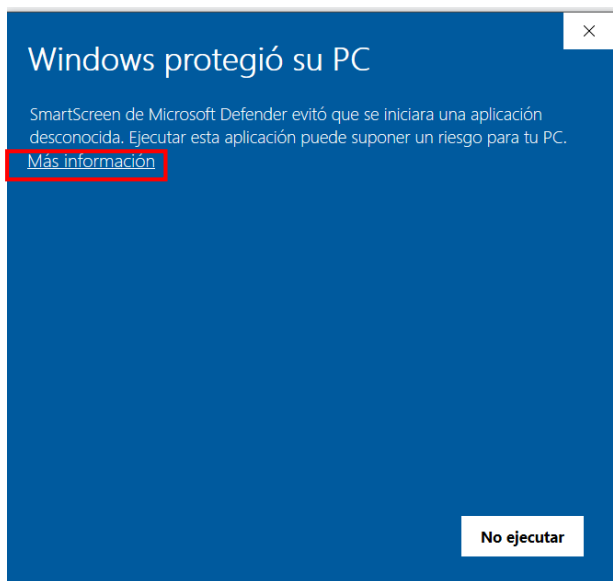
2. INSTALACIÓN DE CYTOSCAPE Y BiNGO

- Ingrese a la página <https://cytoscape.org/download.html> y de clic en el recuadro marcado. Su navegador descargará el archivo desde la dirección web: https://github.com/cytoscape/cytoscape/releases/download/3.9.0/Cytoscape_3_9_0_windows_64bit.exe

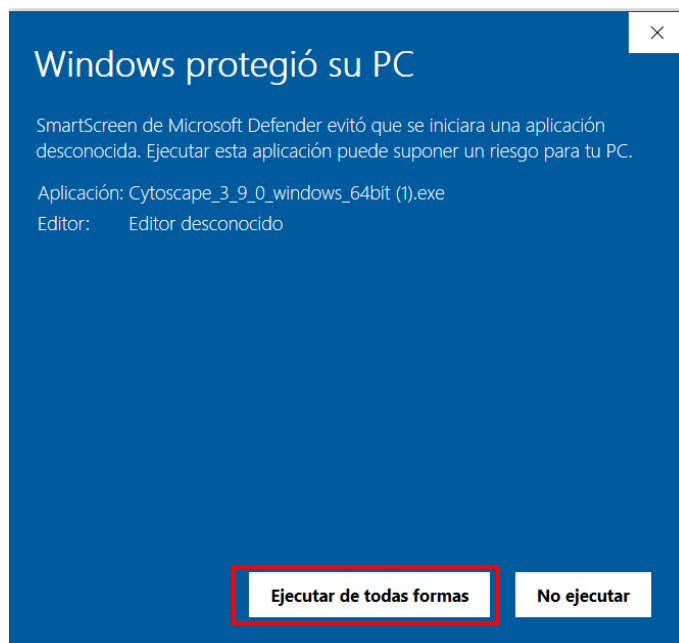
Download Cytoscape 3.9.0



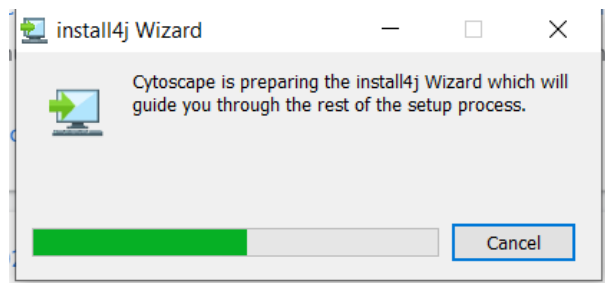
- Al descargarse la aplicación, aparecerá la siguiente ventana, en la cual debe dar clic en donde dice Más información



- Una nueva ventana emergerá y la opción ejecutar de todas formas estará disponible. Por favor, dé clic en el icono.



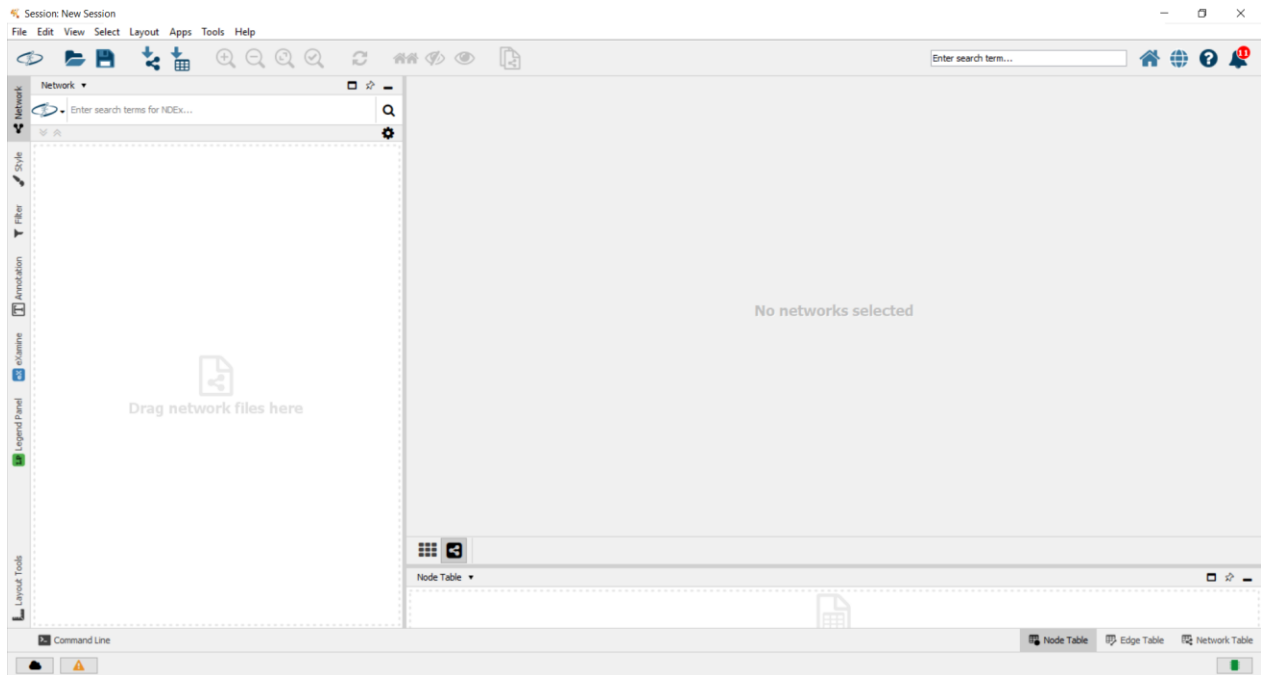
- El instalador se abrirá, luego debe dar clic sobre aceptar para iniciar la instalación



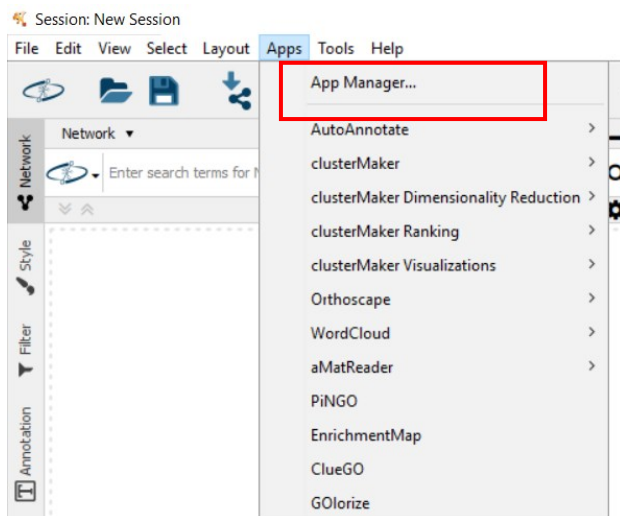
- Al abrirse Cytoscape se verá así:



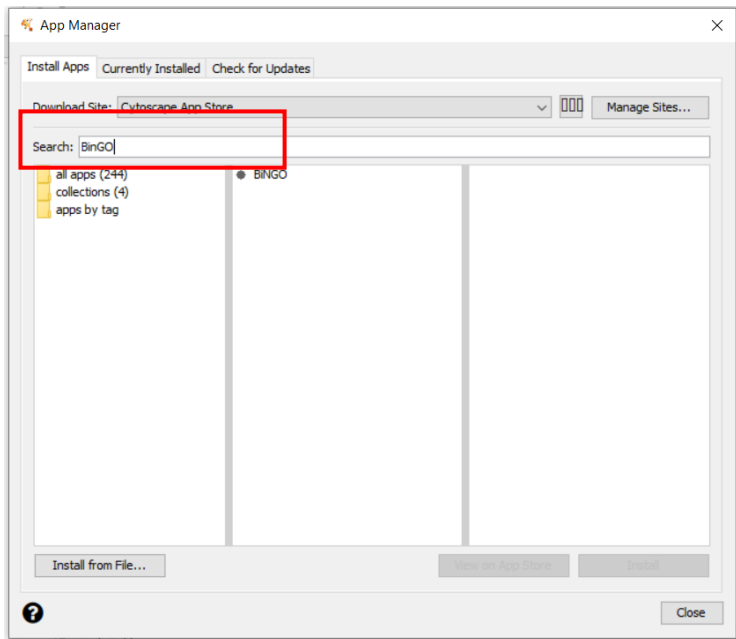
- Luego de la imagen de carga aparecerá el programa, el cual visualmente se ve así:



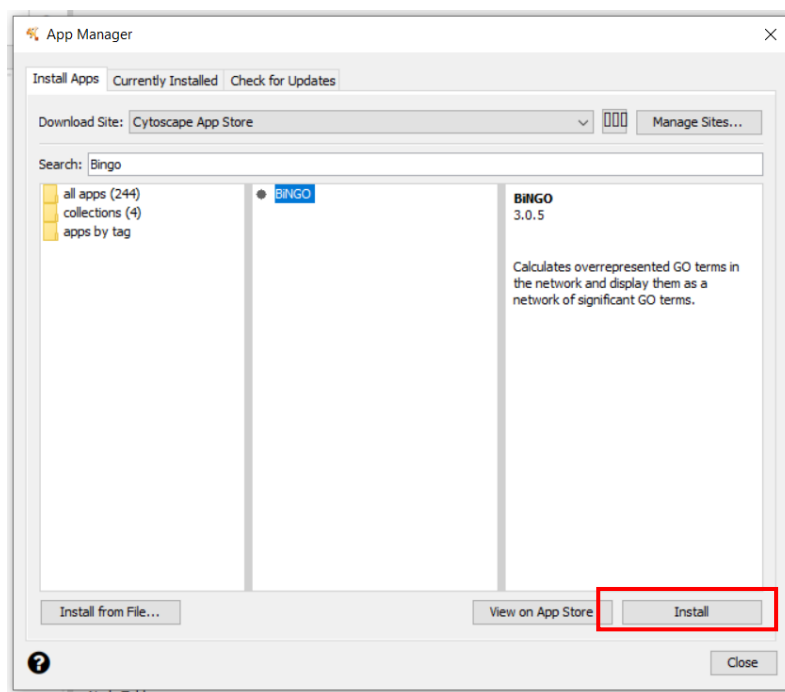
- Para instalar el plugin BiNGO, deberá buscar el menú Apps y dar clic en la opción *App Manager*



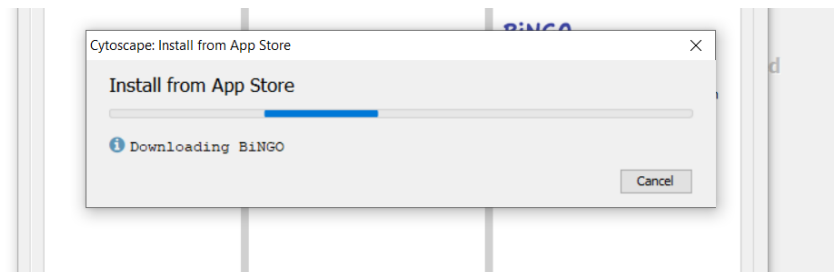
- Al dar clic, la siguiente ventana aparecerá, en la barra de búsqueda que dice Search, escriba BiNGO



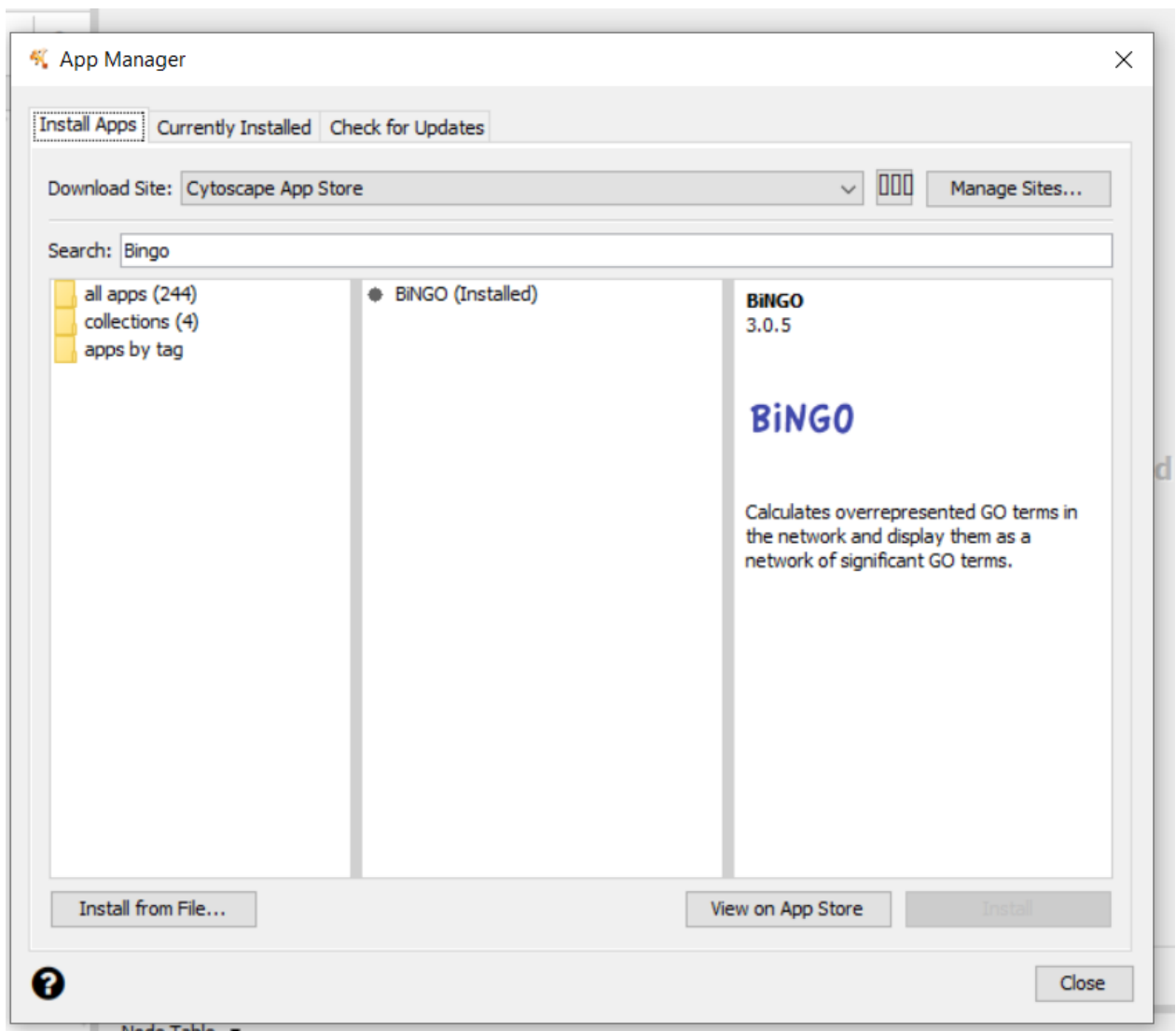
- Dé clic en BiNGO, al hacerlo, aparecerá en la parte inferior derecha la opción de instalar, por favor dé clic en instalar



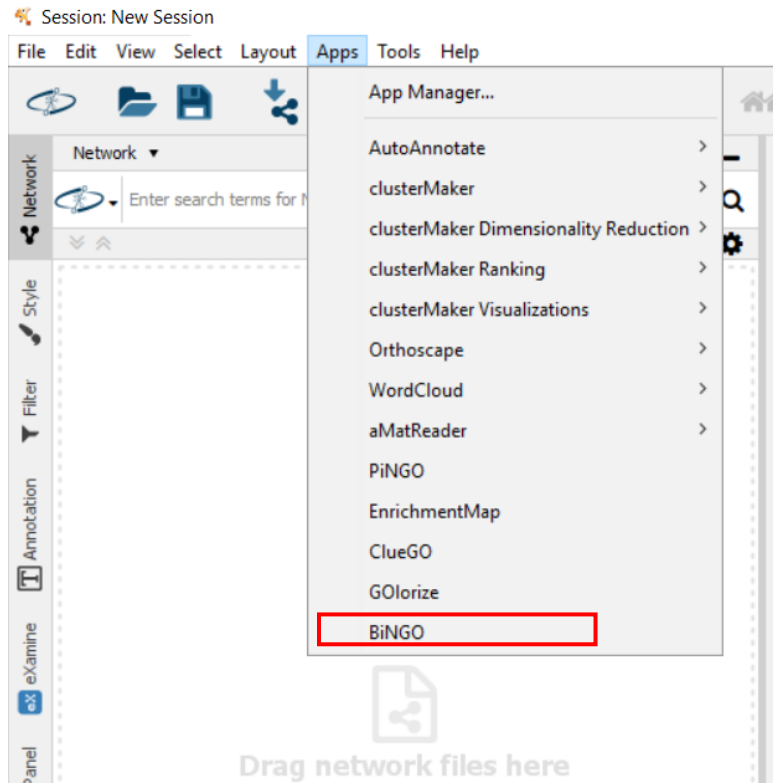
- Al dar clic en instalar, aparecerá un cuadro de progreso mostrando la descarga de BiNGO



- Al terminar la descarga, aparecerá el botón Close, además que se mostrará que BiNGO fue instalado lo cual indicará que ya se instaló el *plugin*

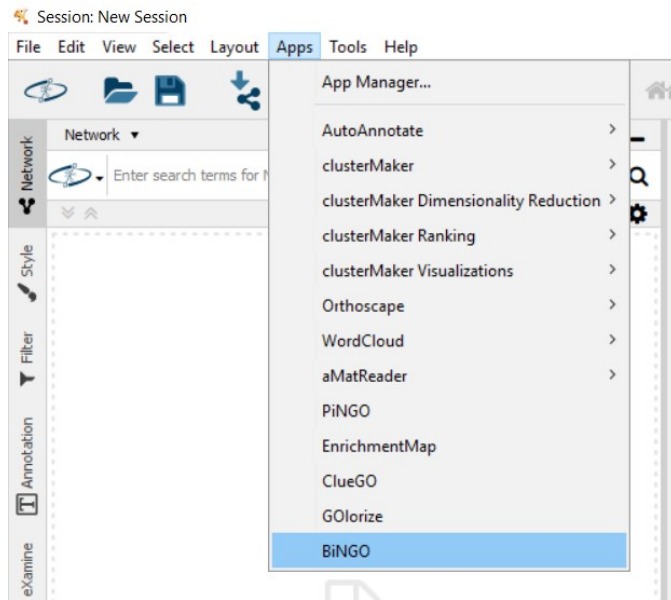


- Para usar BinGO deberá acceder de nuevo al menú Apps, en donde observará que BiNGO está ahora disponible para su uso.



3. Uso de BiNGO

- Abra BiNGO desde el menú Apps



- BiNGO se abrirá y tendrá el siguiente aspecto en donde está la siguiente información en orden descendente (En rojo, se señalan los parámetros más importantes):
 - Nombre del análisis
 - Espacio para incluir los id de los genes
 - Selección de evaluación a ser realizada
 - Prueba estadística
 - Selección del método de corrección de hipótesis múltiples
 - Elección del alfa
 - Selección de los GO a ser visualizados en un grafo
 - Selección de la referencia a usar
 - Información para obtener
 - Selección del organismo a ser analizado
 - Espacio para guardar el archivo localmente
 - Botón para correr el análisis

BiNGO Settings

BiNGO settings

Save settings as default Help

Cluster name:

☐ Get cluster from network
 ☒ Paste genes from text

Q0DD08

Assess: ☒ Overrepresentation ☐ Underrepresentation ☒ Visualization

Select a statistical test:

Hypergeometric test

Select a multiple testing correction:

Benjamini & Hochberg False Discovery Rate (FDR) correction

Choose a significance level:

0.05

Select the categories to be visualized:

Overrepresented categories after correction

Select reference set:

Use whole annotation as reference set

Select ontology file:

GO_Biological_Process

Select namespace:

biological_process

Select organism/annotation:

Oryza sativa

Discard the following evidence codes:

Save BiNGO data file in: D:\DD Browse...

Start BiNGO

- Copie los genes del archivo GIM.csv descargado previamente y péguelos en el espacio provisto, posteriormente elija organismo la opción *Homo sapiens* y desactive la opción *Save BiNGO data file in*, si es el caso, para evitar guardar archivos localmente. Finalmente, dé clic en el botón *Start BiNGO* para iniciar el análisis.

BiNGO Settings

Save settings as default Help

Cluster name:
GIM

☐ Get cluster from network ☒ Paste genes from text

POLR2E
USP7
AURKA
CCND1
RAD17
COPS4
ERCC2
CHD1L
CCNH
ERCC8
ZNF830
RBL2

Assess: ☒ Overrepresentation ☐ Underrepresentation ☒ Visualization

Select a statistical test:
Hypergeometric test

Select a multiple testing correction:
Benjamini & Hochberg False Discovery Rate (FDR) correction

Choose a significance level:
0.05

Select the categories to be visualized:
Overrepresented categories after correction

Select reference set:
Use whole annotation as reference set

Select ontology file:
GO_Biological_Process

Select namespace:
biological_process

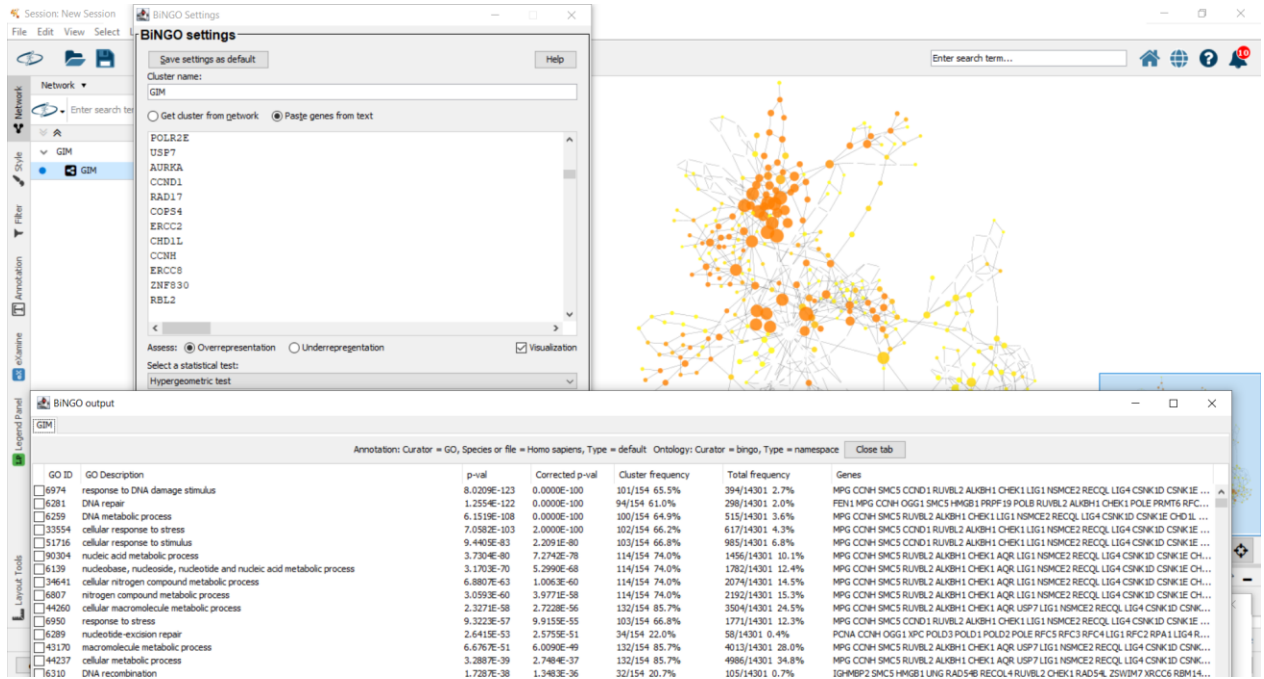
Select organism/annotation:
Homo sapiens

Discard the following evidence codes:
|

☐ Save BiNGO data file in: D:\DD

Start BiNGO

- Al cabo de unos segundos aparecerán los resultados en la interfaz de Cytoscape, obtendrá un grafo dirigido y una tabla con los resultados:



- La tabla con los resultados del análisis aparecerá en la parte inferior de Cytoscape. Esta tabla tiene siete columnas: GO ID, GO Description, p-val, Corrected p-val, Cluster frequency, Total frequency y Genes, las cuales representan: el id del término GO, el nombre del GO, el p valor, el p valor después de usar la corrección de Bonferroni o *false discovery rate*, la frecuencia del término GO en la muestra de genes, la frecuencia total del GO en la base de datos de anotación de la especie y los genes que pertenecen al término GO.

BiNGO output

GIM

Annotation: Curator = GO, Species or file = Homo sapiens, Type = default Ontology: Curator = bingo, Type = namespace

	GO ID	GO Descri...	p-val	Corrected p-val	Cluster frequency	Total frequency	Genes
<input type="checkbox"/>	6974	response t...	8.0209E-123	0.0000E-100	101/154 65.5%	394/14301 2.7%	MPG CCNH ...
<input type="checkbox"/>	6281	DNA repair	1.2554E-122	0.0000E-100	94/154 61.0%	298/14301 2.0%	FEN1 MPG ...
<input type="checkbox"/>	6259	DNA metab...	6.1519E-108	0.0000E-100	100/154 64.9%	515/14301 3.6%	MPG CCNH ...
<input type="checkbox"/>	33554	cellular res...	7.0582E-103	2.0000E-100	102/154 66.2%	617/14301 4.3%	MPG CCNH ...
<input type="checkbox"/>	51716	cellular res...	9.4405E-83	2.2091E-80	103/154 66.8%	985/14301 6.8%	MPG CCNH ...
<input type="checkbox"/>	90304	nucleic acid...	3.7304E-80	7.2742E-78	114/154 74.0%	1456/14301 10.1%	MPG CCNH ...
<input type="checkbox"/>	6139	nucleobase...	3.1703E-70	5.2990E-68	114/154 74.0%	1782/14301 12.4%	MPG CCNH ...
<input type="checkbox"/>	34641	cellular nitr...	6.8807E-63	1.0063E-60	114/154 74.0%	2074/14301 14.5%	MPG CCNH ...
<input type="checkbox"/>	6807	nitrogen co...	3.0593E-60	3.9771E-58	114/154 74.0%	2192/14301 15.3%	MPG CCNH ...
<input type="checkbox"/>	44260	cellular mac...	2.3271E-58	2.7228E-56	132/154 85.7%	3504/14301 24.5%	MPG CCNH ...
<input type="checkbox"/>	6950	response t...	9.3223E-57	9.9155E-55	103/154 66.8%	1771/14301 12.3%	MPG CCNH ...
<input type="checkbox"/>	6289	nucleotide-...	2.6415E-53	2.5755E-51	34/154 22.0%	58/14301 0.4%	PCNA CCN...
<input type="checkbox"/>	43170	macromole...	6.6767E-51	6.0090E-49	132/154 85.7%	4013/14301 28.0%	MPG CCNH ...
<input type="checkbox"/>	44237	cellular met...	3.2887E-39	2.7484E-37	132/154 85.7%	4986/14301 34.8%	MPG CCNH ...
<input type="checkbox"/>	6310	DNA recom...	1.7287E-38	1.3483E-36	32/154 20.7%	105/14301 0.7%	IGHMBP2 S...
<input type="checkbox"/>	44238	primary me...	3.5899E-37	2.6251E-35	133/154 86.3%	5284/14301 36.9%	MPG CCNH ...
<input type="checkbox"/>	6260	DNA replica...	1.1688E-32	8.0440E-31	34/154 22.0%	185/14301 1.2%	FEN1 RNAS...
<input type="checkbox"/>	8152	metabolic p...	7.0191E-31	4.5624E-29	133/154 86.3%	5955/14301 41.6%	MPG CCNH ...
<input type="checkbox"/>	50896	response t...	6.8469E-30	4.2162E-28	106/154 68.8%	3631/14301 25.3%	MPG CCNH ...
<input type="checkbox"/>	34645	cellular mac...	2.7339E-29	1.5993E-27	60/154 38.9%	1030/14301 7.2%	FEN1 CCN...
<input type="checkbox"/>	9059	macromole...	8.2153E-29	4.5771E-27	60/154 38.9%	1051/14301 7.3%	FEN1 CCN...
<input type="checkbox"/>	44265	cellular mac...	3.4092E-27	1.8131E-25	41/154 26.6%	442/14301 3.0%	VCP FEN1 ...
<input type="checkbox"/>	6302	double-str...	1.4695E-26	7.4755E-25	21/154 13.6%	62/14301 0.4%	VCP XRCC6...
<input type="checkbox"/>	6297	nucleotide-...	1.0209E-25	4.9769E-24	14/154 9.0%	17/14301 0.1%	RFC5 RFC3...
<input type="checkbox"/>	718	nucleotide-...	2.4189E-25	1.1320E-23	15/154 9.7%	22/14301 0.1%	CCNH RPA...
<input type="checkbox"/>	9057	macromole...	5.7565E-25	2.5904E-23	41/154 26.6%	504/14301 3.5%	VCP FEN1 ...
<input type="checkbox"/>	6308	DNA catab...	8.6479E-24	3.7474E-22	19/154 12.3%	58/14301 0.4%	MPG CCNH ...
<input type="checkbox"/>	7049	cell cycle	1.1127E-23	4.6496E-22	48/154 31.1%	794/14301 5.5%	PSMD14 C...
<input type="checkbox"/>	75	cell cycle	1.8484E-22	7.8646E-21	22/154 14.2%	107/14301 0.7%	PSMD14 C...

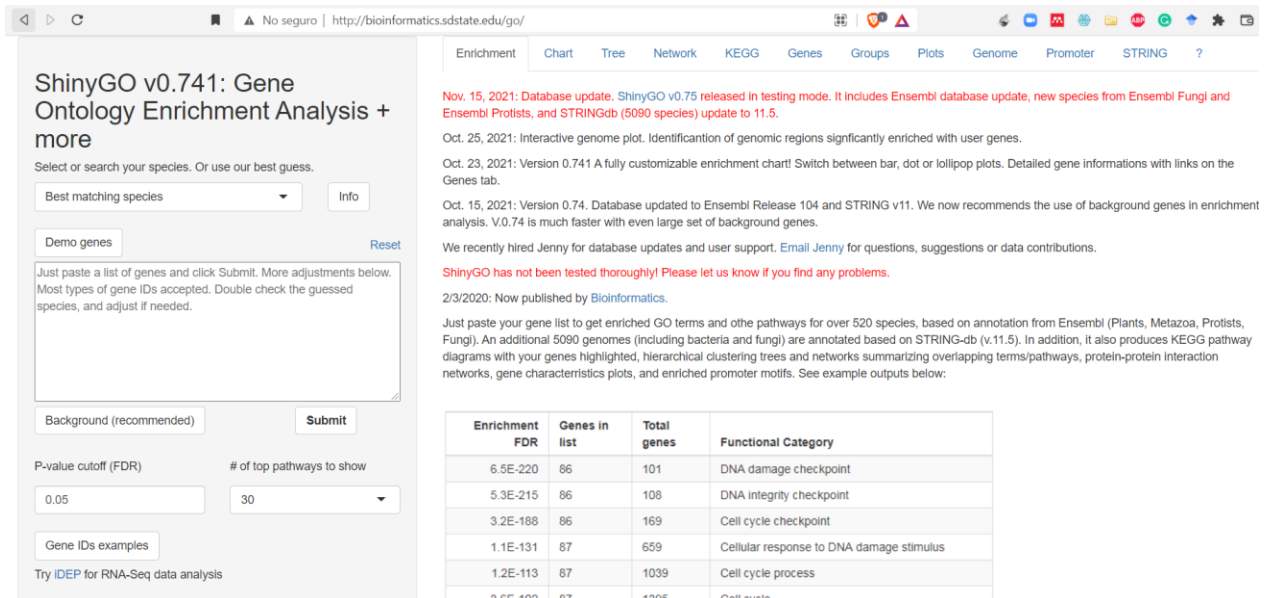
Select All Unselect All Select nodes

4. Uso de ShinnyGO

ShinnyGO es una nueva herramienta para análisis de enriquecimiento funcional que a diferencia de otras herramientas combina el uso de la base de datos de *Gene ontology*, con datos de *KEGG*, *Reactome* y *STRING*. La herramienta tiene diez opciones:

Opción	Descripción
Enrichment	Lista de términos GO enriquecidos
Chart	Gráficos descriptivos para los términos GO enriquecidos
Tree	Análisis de conglomerados jerárquico
Network	Permite obtener un grafo con pesos para los GO enriquecidos
KEGG	Si se usa la opción KEGG permite obtener que rutas metabólicas están sobrerrepresentadas
Genes	Permite obtener información adicional acerca de los genes usados
Group	Agrupar los términos GO enriquecidos y los genes a los cuales están asociados
Plots	Permite obtener gráficos para comparar la lista de genes y el genoma del organismo
Genome	Permite realizar un gráfico para mapear en que cromosomas están los genes usados
Promoter	Permite buscar factores de transcripción para los genes analizados

- Abra el enlace: <http://bioinformatics.sdstate.edu/go/>, en este encontrará la aplicación que consta de un panel izquierdo en donde encontrará el organismo el cual será evaluado, el espacio para ingresar los genes, el botón *Submit*, que activará el programa, la opción *P-value cutoff (FDR)*, para elegir, el punto critico para definir que GO están enriquecidos y la opción *# of top pathways to show*, que servirá para definir cuantos GO enriquecidos desea que aparezcan.



ShinyGO v0.741: Gene Ontology Enrichment Analysis + more

Select or search your species. Or use our best guess.

Best matching species Info

Demo genes

Just paste a list of genes and click Submit. More adjustments below. Most types of gene IDs accepted. Double check the guessed species, and adjust if needed.

Background (recommended)

P-value cutoff (FDR) # of top pathways to show

Gene IDs examples

Enrichment Chart Tree Network KEGG Genes Groups Plots Genome Promoter STRING ?

Nov. 15, 2021: Database update. ShinyGO v0.75 released in testing mode. It includes Ensembl database update, new species from Ensembl Fungi and Ensembl Protists, and STRINGdb (5090 species) update to 11.5.

Oct. 25, 2021: Interactive genome plot. Identification of genomic regions significantly enriched with user genes.

Oct. 23, 2021: Version 0.741 A fully customizable enrichment chart! Switch between bar, dot or lollipop plots. Detailed gene informations with links on the Genes tab.

Oct. 15, 2021: Version 0.74. Database updated to Ensembl Release 104 and STRING v11. We now recommends the use of background genes in enrichment analysis. V0.74 is much faster with even large set of background genes.

We recently hired Jenny for database updates and user support. Email Jenny for questions, suggestions or data contributions.

ShinyGO has not been tested thoroughly! Please let us know if you find any problems.

2/3/2020: Now published by Bioinformatics.

Just paste your gene list to get enriched GO terms and other pathways for over 520 species, based on annotation from Ensembl (Plants, Metazoa, Protists, Fungi). An additional 5090 genomes (including bacteria and fungi) are annotated based on STRING-db (v.11.5). In addition, it also produces KEGG pathway diagrams with your genes highlighted, hierarchical clustering trees and networks summarizing overlapping terms/pathways, protein-protein interaction networks, gene characteristics plots, and enriched promoter motifs. See example outputs below:

Enrichment FDR	Genes in list	Total genes	Functional Category
6.5E-220	86	101	DNA damage checkpoint
5.3E-215	86	108	DNA integrity checkpoint
3.2E-188	86	169	Cell cycle checkpoint
1.1E-131	87	659	Cellular response to DNA damage stimulus
1.2E-113	87	1039	Cell cycle process
3.6E-102	87	1395	Cell cycle

- Copie los genes del archivo GIM.csv descargado previamente y péguelos en el espacio provisto, posteriormente elija como organismo la opción *Human*, elija 0.05 como FDR y seleccione que solo quiere el top 30 de términos GO enriquecidos. Finalmente, dé clic en el botón *Submit*.

ShinyGO v0.741: Gene Ontology Enrichment Analysis + more

Select or search your species. Or use our best guess.

Human ▼ Info

Demo genes Reset

RAD17
COPS4
ERCC2
CHD1L
CCNH
ERCC8
ZNF830
RBL2

😊 10

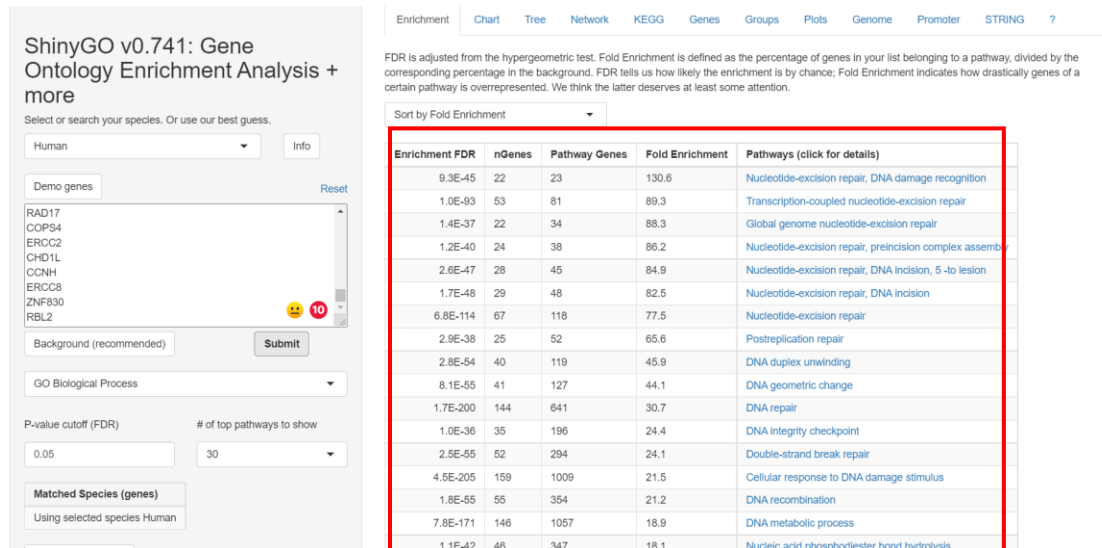
Background (recommended) Submit

P-value cutoff (FDR) 0.05 # of top pathways to show 30 ▼

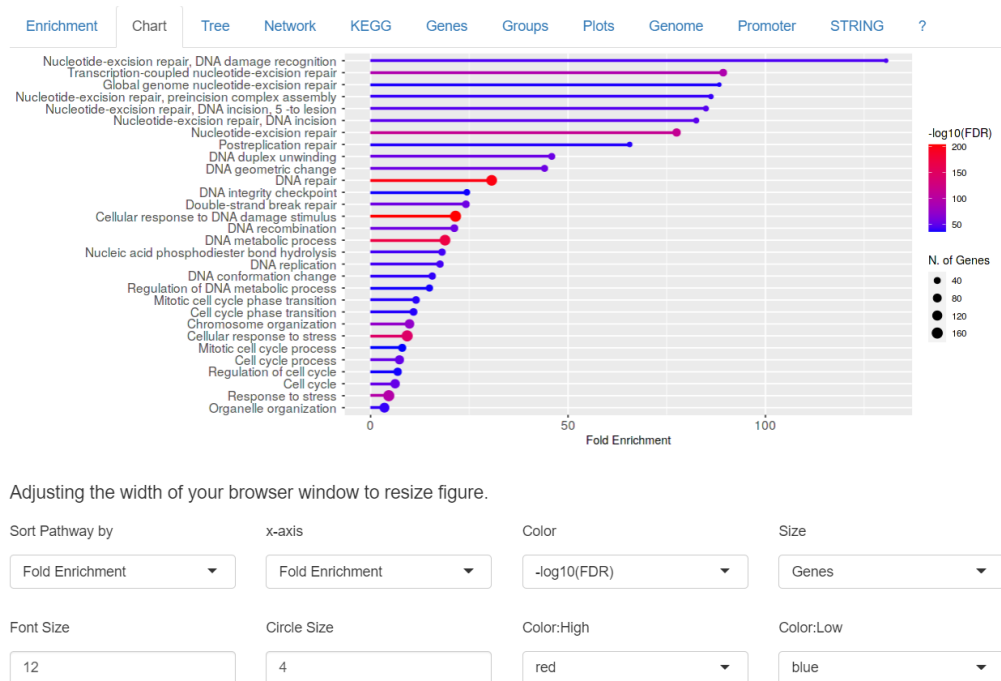
Gene IDs examples

Try [iDEP](#) for RNA-Seq data analysis

- Al cabo de unos segundos los resultados del análisis serán mostrados, estos están organizados de acuerdo con el $Fold\ enrichment = \frac{\% \text{ genes de lista en una ruta}}{\% \text{ genes en background}}$, que representa cuán drásticamente están sobrerrepresentados los genes de una determinada ruta.



- Al correr el análisis, diez opciones están disponibles para su uso. Al dar clic en *chart*, podrá obtener un gráfico de los términos GO enriquecidos, donde, *x* es el *fold enrichment*, y son los términos GO obtenidos y los colores representan el FDR obtenido.



- Si se usa la opción Network aparecerá el siguiente grafo, ¿qué representa?

