



Análisis de la implicancia de los miARN en el hepatocarcinoma aplicando minería de texto

Natalia Riedel Cánepa

Máster Universitario en Bioinformática y Bioestadística

Computación e Inteligencia Artificial en Problemas Biológicos y Clínicos

Romina Astrid Rebrij

Antoni Pérez Navarro

Enero 2021



Esta obra está sujeta a una licencia de
Reconocimiento-NoComercial-SinObraDerivada
[3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Análisis de la implicancia de los miARN en el hepatocarcinoma aplicando minería de texto
Nombre del autor:	<i>Natalia Riedel Cánepa</i>
Nombre del consultor/a:	<i>Romina Astrid Rebrij</i>
Nombre del PRA:	<i>Antoni Pérez Navarro</i>
Fecha de entrega (mm/aaaa):	01/2021
Titulación:	<i>Máster Universitario en Bioinformática y Bioestadística</i>
Área del Trabajo Final:	<i>Computación e Inteligencia Artificial en Problemas biológicos y Clínicos</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>hepatocarcinoma, miARN, minería de texto</i>
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i></p>	
<p>A través de la minería de texto, se ha logrado recopilar información proveniente de diferentes publicaciones, relacionadas con el hepatocarcinoma (HCC) y los miARN. A partir de acá se han obtenido genes que estarían implicados con esta enfermedad, miARN que juegan un papel importante y mutaciones que podrían conducir a la formación del HCC .</p> <p>El análisis realizado se visualiza a través de una aplicación web, HmiR², que fue llevada a cabo con el framework Shiny. Esta permite la integración con varias bases de datos que aportan información relevante sobre el contexto en el que se desarrolla esta enfermedad, como cuál es la función de los genes, cuál es el target al que se dirigen los miARN, qué tipo de mutaciones y dónde ocurren las mismas.</p> <p>Este trabajo podría funcionar como punto de partida hacia nuevas líneas de investigación que incluyan a los miARN y su participación en el cáncer.</p>	

Abstract (in English, 250 words or less):

Through text mining, it has been possible to collect information from different publications, related to hepatocarcinoma (HCC) and miRNAs. From here we have obtained genes that would be involved in this disease, miRNAs that play an important role and mutations that could lead to the formation of HCC.

The analysis carried out is viewed through a web application, HmiR², which was carried out with the Shiny framework. This allows integration with several databases that provide relevant information about the context in which this disease develops, such as what is the function of genes, what is the target to which miRNAs are directed, what type of mutations and where they happen.

This work could serve as a starting point for new lines of research that include miRNAs and their role in cancer.

Tabla de contenido

1.Introducción	9
1.1 Contexto y justificación del trabajo	9
1.2 Objetivos del trabajo	10
1.2.1 Objetivo General:	10
1.2.2 Objetivos Específicos:	10
1.3 Enfoque y método seguido	10
1.4 Plan de trabajo	12
1.5 Breve sumario de productos obtenidos	14
1.6 Breve descripción de los otros capítulos de la memoria	14
2.Contexto de trabajo	14
2.1 Hepatocarcinoma	14
2.2 MicroARN	16
3.Metodología	16
3.1 Minería de texto	17
3.1.1 Análisis del Corpus Primario	18
3.1.2 Obtención de Corpus Secundario	19
3.2 Aplicación Shiny	19
Shiny vincula valores de entrada con los de salida(output) a partir de programación Reactiva.	20
4.Desarrollo del Trabajo	21
4.1 Genes	23
4.2 miARN	25
4.3 Mutaciones	28
4.4 Entorno HCC	30
5.Resultados	31
5.1 Genes-HCC	31
5.2 miARN-HCC	35
5.3Mutaciones-miARN	38
6.Conclusiones	40
7.Glosario:	42
8.Bibliografía:	42
9.Anexo	43

Tabla de Ilustraciones

Figura 1: Workflow basado en Minería de texto (Rany Jyoti et al,2015).....	11
Figura 2: Calendario realizado a través de "Gantt Project"	13
Figura 3: Anatomía del hígado.....	15
Figura 4: Respuesta obtenida desde Pubmed.....	18
Figura 5: Widgets de Shiny estándar.....	20
Figura 6: Salidas reactivas estándar.....	21
Figura 7: Pantalla inicial de HmiR ²	22
Figura 8: Búsqueda de Pubmed de genes.....	23
Figura 9: Resultados de búsqueda GO.....	25
Figura 10: Input miARN.....	26
Figura 11: Búsqueda de Pubmed de miARN.....	27
Figura 12: miARN involucrados.....	28
Figura 13: Genes diana.....	29
Figura 14: Búsqueda de Pubmed de mutaciones.....	29
Figura 15: Información sobre mutaciones.....	30
Figura 16: Información sobre el entorno HCC.....	32
Figura 17: Resultados frecuencia de genes.....	32
Figura 18: Búsqueda publicaciones PTEN.....	33
Figura 19: Búsqueda publicaciones AFP.....	34
Figura 20: Búsqueda publicaciones STAT3.....	34
Figura 21: Resultados obtenidos de la base de datos GO para PTEN.....	35
Figura 22: Resultados obtenidos de la base de datos GO para AFP.....	36
Figura 23: Resultados obtenidos de la base de datos GO para STAT3.....	36
Figura 24: Resultados de miARN más frecuentes.....	36
Figura 25: Búsqueda publicaciones miR-122.....	37
Figura 26: Búsqueda publicaciones miR-21.....	37
Figura 27: Búsqueda publicaciones miR-221.....	38

Figura 28: Resultados obtenidos de miARN involucrados con el gen DICER1.....	38
Figura 29: Resultados obtenidos de miARN involucrados con el gen AFP.....	39
Figura 30: Resultados obtenidos de miARN involucrados con el genGNMT.....	39
Figura 31: Resultados de genes relacionado con mutaciones más frecuentes.....	39
Figura 32: Publicaciones sobre mutaciones.....	39
Figura 33: Resultados de mutaciones en el gen AFP.....	40
Figura 34: Resultados de mutaciones en el gen DICER1.....	40
Figura 35: Resultados de mutaciones en el gen GNMT.....	40

Este trabajo está dedicado:

-Al Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, que me permite colaborar con su investigación.

-A mi familia, que a pesar de que un océano nos separa, siempre están conmigo de forma incondicional.

-Y a Martín, mi novio y compañero de vida que cree en mí y me apoya en cada paso que doy.

1.Introducción

1.1 Contexto y justificación del trabajo

El hepatocarcinoma (HCC), es el tipo más común de cáncer primario de hígado y la sexta neoplasia más frecuente a nivel mundial. En la mayoría de las ocasiones, asienta sobre una enfermedad hepática previa, y se conocen muchos de los factores de riesgo para desarrollarlo (hepatitis virales, cirrosis alcohólica, esteatohepatitis no alcohólica, entre otros). Las anomalías moleculares más comunes en esta neoplasia son mutaciones en los genes del promotor TERT, TP53, CTNNB1, AXIN1, ARID1A, CDKN2A y CCND1 [1, 2].

Por otro lado, los ácidos microrribonucleicos (miARN) son una clase de ARN pequeño no codificante, conservado evolutivamente con entre 21 y 24 nucleótidos de longitud. Desempeñan un papel crucial en diferentes procesos biológicos como la proliferación, diferenciación, apoptosis y metabolismo. Se ha informado que más del 50% de los genes de miARN se localizan a menudo en sitios frágiles de los cromosomas y en las regiones genómicas que son propensas a la delección, amplificación y mutaciones. Además se reconoce que la expresión aberrante de varios miARN está estrechamente relacionada con la etiología y el resultado clínico de varios cánceres humanos, incluido el hepatocarcinoma. A estos miARN se los conoce como "oncomiR" si promueven el desarrollo de tumores y "anti-oncomiR" si inhibe el crecimiento tumoral [3].

A lo largo de todos estos años se han realizado miles de trabajos de investigación en distintas partes del mundo y abarcando varias aristas relacionadas con este tipo de enfermedad. La cantidad de publicaciones científicas inconexas que hoy se encuentran en la web hace inviable que se puedan leer todas manualmente y sacar un patrón observable que sirva como puntapié para, quizás, comenzar una nueva línea de investigación sobre este tema.

Para este trabajo se utilizó justamente la minería de texto que tiene como finalidad, descubrir nueva información a partir de la información ya existente, deduciendo relaciones que se encuentren explícitas en los documentos que forman parte del corpus del trabajo[4]. Y se desarrolló una aplicación amigable que sirve al usuario como herramienta para poder encontrar información relevante sobre la enfermedad.

1.2 Objetivos del trabajo

1.2.1 Objetivo General:

Desarrollar una aplicación web interactiva, que permita dilucidar cuáles son los posibles roles que cumplen los miARN en el hepatocarcinoma.

1.2.2 Objetivos Específicos:

- ✓ Buscar genes potencialmente implicados en el desarrollo de esta enfermedad
- ✓ Determinar la presencia de prooncomiR y AntioncomiR
- ✓ Determinar la relación entre las mutaciones y los miARN

1.3 Enfoque y método seguido

Se propuso trabajar con el entorno R. Esto se debe, por un lado a la amplia utilización que se le ha dado a esta herramienta a través de las diferentes materias cursadas. Por otro lado, porque R presenta librerías que son idóneas para el tipo de trabajo que se planteó, que es el de minería de texto, con lo cual se determinó utilizar este software para realizar este Trabajo Final de Máster (TFM).

Para cualquier trabajo de investigación es fundamental la búsqueda y el respaldo bibliográfico y uno de los motores de búsqueda más importantes y reconocidos a nivel biomédico es PubMed que comprende más de 30 millones de citas de MEDLINE, revistas de ciencias biológicas y libros en línea [5]. Se comenzó con la búsqueda de información relacionada con el tema planteado, en la página web antes mencionada. Esto permitió obtener un corpus principal desde el cual se hicieron los primeros análisis, filtros, clasificación y se avanzó con el TFM a partir de este.

Uno de los métodos que se siguió, fue el del artículo Rany Jyoti *et al.* [6], que a través de diferentes funciones del paquete *pubmed.mineR*, busca crear un corpus primario, filtrarlo y obtener uno secundario, obtener términos que serán sometidos a Análisis Semántico Latente (LSA) y posterior similitud por coseno, según presenta el siguiente flujo de trabajo:

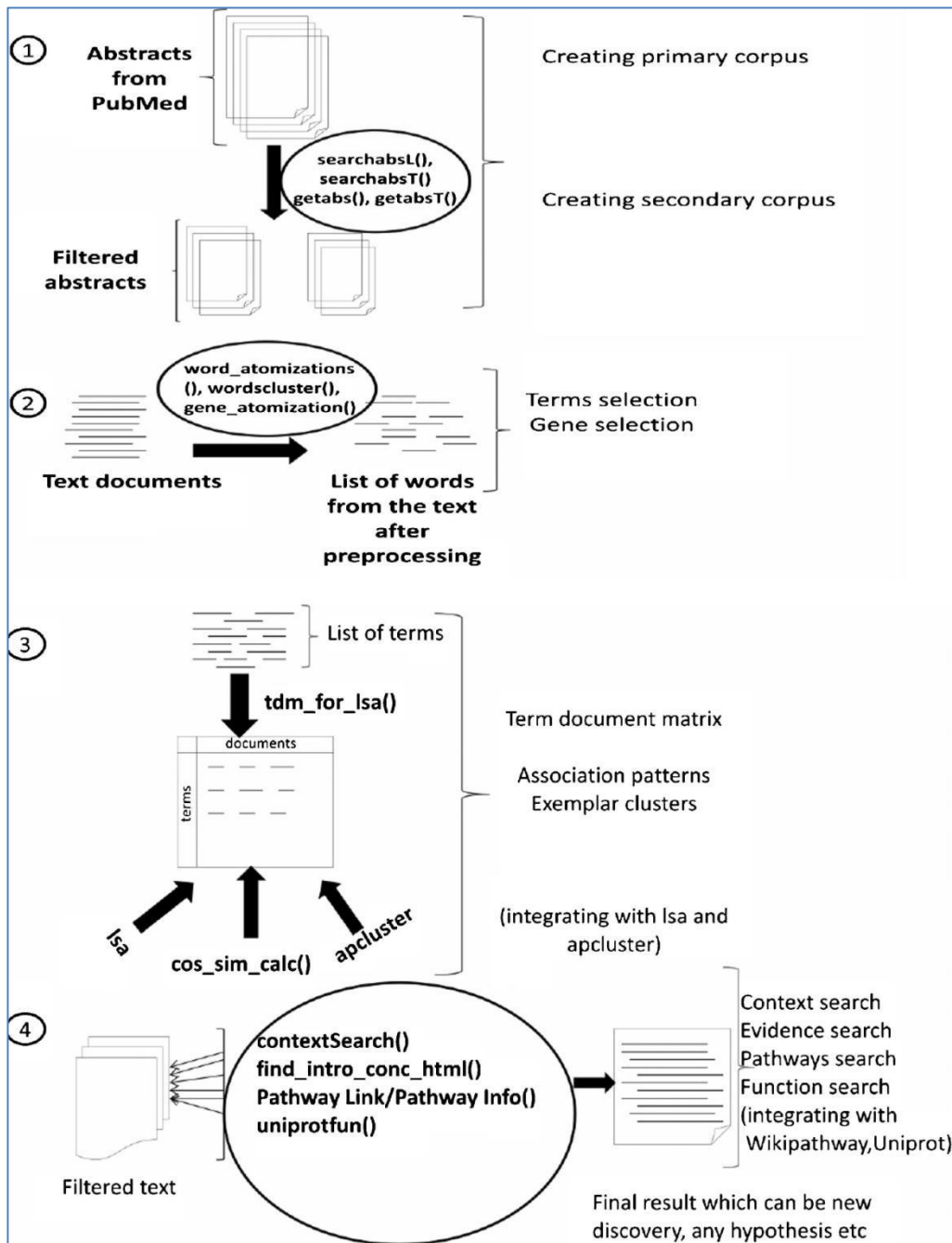


Figura 1: Workflow basado en Minería de texto (Rany Jyoti *et al*,2015)

De esta primera parte se pudo extraer los abstracts relacionados con la enfermedad en cuestión que es el hepatocarcinoma, sobre prooncomiR y AntioncomiR (miARN), genes y mutaciones que estén relacionados con el tema principal. A su vez para complementar y poder llegar al final de este TFM con una deducción más concluyente se hizo una integración con otras bases de datos relacionadas, a través de palabras claves.

Se cuenta actualmente con muchas bases de datos que permiten obtener una visión más completa sobre los distintos temas de investigación , en este caso se decidió la asociación con "Gene Ontology" (GO), ya que obtendremos una descripción del gen y sus atributos, con distintas bases de datos relacionadas con mis miARN y con información sobre mutaciones de genes.

Finalmente se implementó el script de minería de texto sobre una aplicación web a través de la librería *Shiny* de R a fin de que se pueda visualizar de manera organizada y amigable los datos obtenidos.

1.4 Plan de trabajo

Para poder llevar adelante este trabajo se organizó en bloques que envuelvan los objetivos planteados. Los bloques quedaron estructurados con sus respectivos hitos de la siguiente manera:

- ✓ Módulo Inicial: Preparar la base para cumplimentar los siguientes módulos
- ✓ Módulo Genes-HCC: El objetivo fue poder mostrar las publicaciones correspondientes, en un periodo de 20 años, que se presentan en Pubmed y conectar, a través de una tabla, con la base de datos "Gene Ontology (GO)"
- ✓ Módulo miARN-HCC: El objetivo fue mostrar las publicaciones correspondientes al tema, en un periodo de 20 años, que se presentan en Pubmed y conectar, a través de una tabla, con la bases de datos que impartan información sobre los miARN
- ✓ Módulo Mutaciones-miARN: El objetivo de este módulo fue mostrar las publicaciones correspondientes al tema, en un también de periodo de 20 años, que se representan en Pubmed y conectar, a través de una tabla, con bases de datos que aporten información relevante sobre genes que habrían sido mutados.

A su vez se realizó una aplicación a través de la librería "Shiny" de R (desde ahora "HmiR²"), que es desde donde se muestran los resultados obtenidos.

Luego de diseñar estos módulos generales y viendo los tiempos de los que se disponía, se decidió separar el desarrollo del trabajo según el cuadro que más abajo se detalla.

GANTT project			
Nombre	Fecha de i...	Fecha de ...	
Definición de los contenidos de trabajo	16/09/20	28/09/20	
Elaboración de la propuesta de TFM	16/09/20	22/09/20	
Corrección	23/09/20	28/09/20	
Plan de estudio	29/09/20	13/10/20	
Definir contenidos, tareas y tiempos de elaboración ...	29/09/20	8/10/20	
Corrección	9/10/20	13/10/20	
Desarrollo del trabajo: Fase 1	14/10/20	16/11/20	
MODULO Inicial	14/10/20	2/11/20	
Investigación y descargas de librerías idoneas p...	14/10/20	16/10/20	
Descarga de abstracts y elaboración y posterior...	19/10/20	20/10/20	
Determinación de frecuencia y relación entre t...	21/10/20	22/10/20	
Investigación y diseño de aplicación web intera...	23/10/20	2/11/20	
MODULO Genes-HCC	3/11/20	6/11/20	
Recolección de abstracts pertinentes	3/11/20	4/11/20	
Tabla de resultados	5/11/20	6/11/20	
MODULOS miARN-HCC	9/11/20	12/11/20	
Recolección de abstracts pertinentes	9/11/20	10/11/20	
Tabla de resultados	11/11/20	12/11/20	
MODULO mutaciones -HCC	13/11/20		
Recolección de abstracts pertinentes	13/11/20	13/11/20	
Tabla de resultados	16/11/20	16/11/20	
Desarrollo Web	23/10/20	16/11/20	
Desarrollo del trabajo: Fase2	17/11/20	14/12/20	
MODULO Genes-HCC	17/11/20	25/11/20	
Integración con base de datos "Gene Ontology"	17/11/20	25/11/20	
MODULOS mutaciones-HCC	26/11/20	2/12/20	
Interacción con base de datos	26/11/20	2/12/20	
MODULO miARN-HCC	3/12/20	12/12/20	
Investigación y descarga de nuevas librerías idon...	3/12/20	9/12/20	
Interacción con base de datos	3/12/20	9/12/20	
Interacción con base de datos "miRBase"	7/12/20	12/12/20	
Comienzo de elaboración de memoria	14/12/20	14/12/20	
Desarrollo Web	17/11/20	14/12/20	
Elaboración de la memoria	18/12/20	8/01/21	
Ajustes y correcciones	18/12/20	8/01/21	
Elaboración de la presentación	11/01/21	12/01/21	
Realización de Power Point	11/01/21	12/01/21	
Defensa pública	13/01/21	20/01/21	

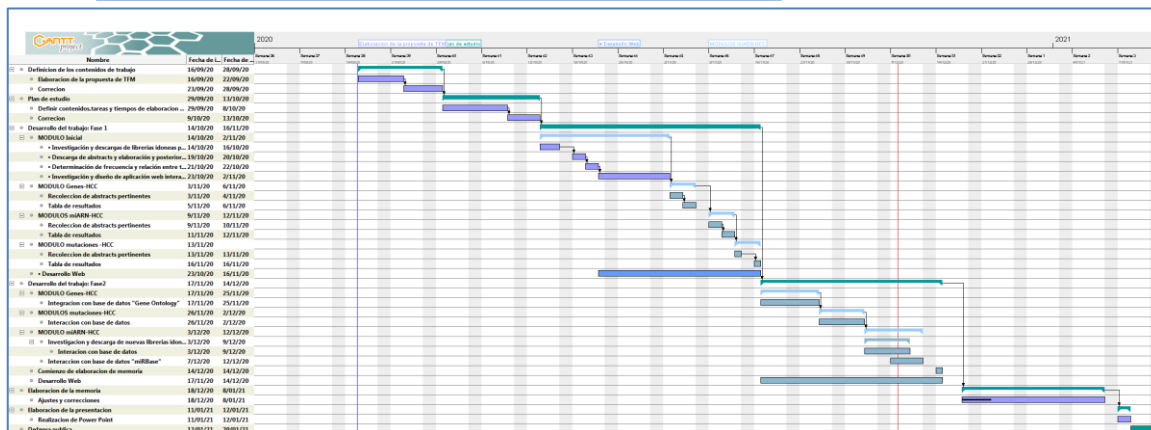


Figura 2: Calendario realizado a través de "Gantt Project"

1.5 Breve resumen de productos obtenidos

- ✓ **Plan de Trabajo:** Escrito con detalles de lo realizado en cada fase.
- ✓ **Memoria:** Documento que registra la actividad realizada en el TFM, resultados y conclusiones.
- ✓ **Aplicación Web:** La URL que permite el ingreso a la app es: <https://aconcagua.shinyapps.io/trabajofinal/>
También está disponible el código en GitHub: <https://github.com/NatyRiedel/TFM>
- ✓ **Presentación:** A través de diapositivas Power Point se realiza una exposición del trabajo realizado. Luego una autoevaluación que expone lo aprendido a lo largo del desarrollo del TFM.

1.6 Breve descripción de los otros capítulos de la memoria

- ✓ **Capítulo 2:** Breve descripción del hepatocarcinoma y miARN para dar una visión general del tema trabajado.
- ✓ **Capítulo 3:** Metodología utilizada enumerando herramientas y bases de datos.
- ✓ **Capítulo 4:** Descripción de cómo se llevó a cabo la aplicación.
- ✓ **Capítulo 5:** Resultados obtenidos.
- ✓ **Capítulo 6:** Conclusiones generales del trabajo realizado.
- ✓ **Capítulo 7:** Glosario.
- ✓ **Capítulo 8:** Bibliografía utilizada.
- ✓ **Capítulo 9:** Anexo con el código que fue usado. Se proporciona URL de la aplicación web.

2. Contexto de trabajo

2.1 Hepatocarcinoma

El hígado es el órgano más grande del cuerpo humano y está ubicado en la parte superior derecha, detrás de las costillas inferiores. Cumple muchas funciones, incluidas las siguientes:

- ✓ Almacena nutrientes previamente absorbidos por el sistema digestivo (glucosa en forma de glucógeno, hierro, vitaminas, etc.)
- ✓ Elimina productos de desecho y células no funcionales del cuerpo.
- ✓ Filtra y procesa sustancias químicas de los alimentos, el alcohol y los medicamentos

- ✓ Sintetiza elementos involucrados en la coagulación sanguínea
- ✓ Produce *bilis*, que ayuda a digerir las grasas y a eliminar los productos de desecho

La mayoría de las células que encontramos en el hígado son los hepatocitos, encargados de las funciones metabólicas del hígado. También otros tipos como las células de Kupffer, encargadas de la función inmunológica[7].

El hígado se divide en dos lóbulos(derecho e izquierdo) y se caracterizan por tener una distribución propia de ramas de arterias y venas. A cada segmento le llega una ramificación de la arteria hepática, sale una ramificación de la vena hepática y también le llega una ramificación de la vena porta. A este paquete vascular tan complejo se le suma la vía biliar, también distribuidas sus ramas desde cada segmento[8].

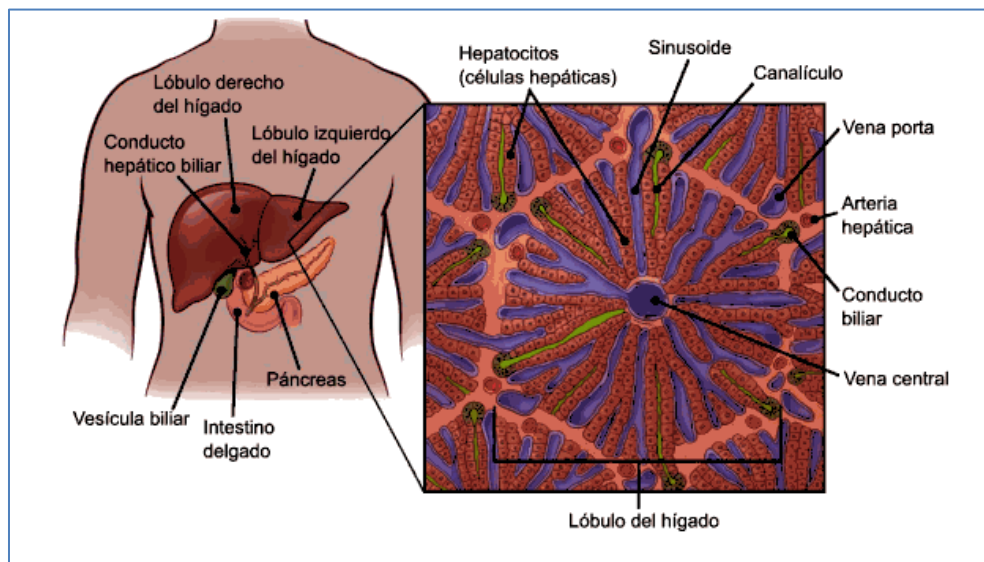


Figura 3: Anatomía del hígado

El cáncer de hígado es la quinta causa más común de cáncer en el mundo y la segunda causa principal de muerte por cáncer, lo que representa un problema de salud mundial. El hepatocarcinoma en particular representa el 90% de las neoplasias malignas primarias del hígado. El HCC es un proceso complejo y progresivo que suele detectarse en etapas tardías de la enfermedad y, por tanto, tiene un mal pronóstico y una supervivencia limitada.

Conocimiento sobre los mecanismos y la señalización que gobiernan la proliferación celular, apoptosis, migración y otros procesos relacionados con el desarrollo tumoral es un tema central para comprender las bases de la enfermedad y cubrir tratamientos innovadores[8].

El hallazgo de terapias eficaces para luchar contra el cáncer ha sido uno de los retos científicos desde hace mucho tiempo. Aunque el diagnóstico y el tratamiento del HCC han experimentado un gran avance en los últimos años, es necesaria la identificación de nuevos biomarcadores para el diagnóstico precoz del tumor e identificar nuevas dianas terapéuticas[3].

2.2 MicroARN

Los ácidos microrribonucleicos (miARN) son una clase de moléculas pequeñas de ARN no codificantes, conservadas evolutivamente y de origen natural, de aproximadamente 21 a 24 nucleótidos de longitud. Debido a que regulan alrededor del 30% de los genes, desempeñan un papel crucial. Sus funciones incluyen la regulación de procesos celulares importantes como la diferenciación, el desarrollo tisular, la apoptosis y el metabolismo. Además, evidencias recientes muestran una expresión alterada de miARN en diversas enfermedades, incluido el cáncer [3],[9].

La expresión aberrante de varios miARN está estrechamente relacionada con la etiología y el resultado clínico de varios cánceres humanos, incluido hepatocarcinoma. Los miARN involucrados en carcinogénesis pueden tener dos roles diferentes, pueden ser microRNA oncogenes, llamados oncomirs, si promueven el desarrollo de tumores a través de la regulación negativa de genes supresores de tumores. Alternativamente, los microARN supresores de tumores, llamados anti-oncomiR, pueden inhibir el crecimiento tumoral al dirigirse a los oncogenes[9].

3. Metodología

Como se ha comentado, para este trabajo se utilizó un entorno R, debido a la amplia utilización que se le ha dado a través de las diferentes materias cursadas y porque presenta librerías idóneas para el tipo de trabajo que se proponía.

A continuación se presenta las herramientas y bases de datos y algunos paquetes utilizados para llevar adelante este TFM

✓ Pubmed

Para cualquier trabajo de investigación es fundamental la búsqueda y el respaldo bibliográfico y uno de los motores de búsqueda más importantes y reconocidos a nivel

biomédico es PubMed que permite buscar en más de 30 millones de citas de MEDLINE, revistas de ciencias biológicas y libros en línea [5].

✓ Gene Ontology

El consorcio GO proporciona una representación computacional del conocimiento científico actual sobre las funciones de los genes las proteínas y moléculas de ARN no codificantes producidas por genes, de muchos organismos diferentes.

GO describe tres aspectos:

- Función Molecular: Actividades a nivel molecular realizadas por productos genéticos, como "catálisis" o "transporte".
- Componente Celular: Las ubicaciones relativas a las estructuras celulares en las que un producto génico realiza una función. Se refiere a anatomía celular.
- Proceso Biológico: Los procesos más grandes, o "programas biológicos" realizados por múltiples actividades moleculares [11].

✓ Paquete multiMiR

Este paquete se ha utilizado para recopilar información acerca de los diferentes miARN existentes. Cuenta con tres categorías a las que consulta diferentes bases de datos externas:

- Interacciones de miARN-objetivo validadas: consulta a miRecords, miRTarBase, TarBase.
- Interacciones previstas de miARN-objetivo: DIANA-microT-CDS, EIMMo, Microcosmo, miRanda, miRDB, PicTar, AGAVE, TargetScan.
- Asociaciones de miARN-enfermedades/fármacos: miR2Disease, Pharmaco-miR, Fenómero[12].

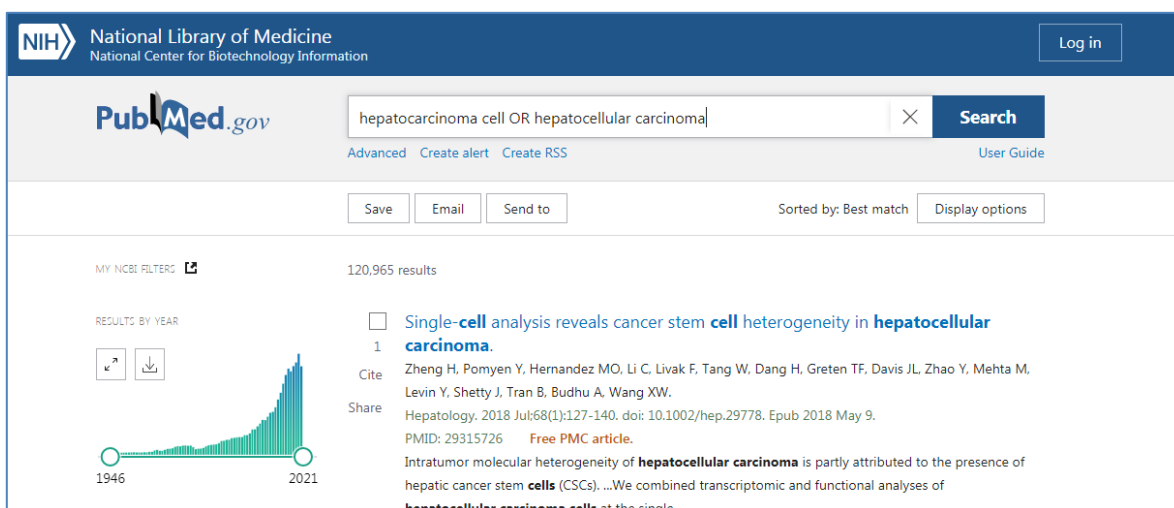
3.1 Minería de texto

Para este trabajo se utilizó minería de texto que tiene como finalidad, descubrir nueva información a partir de la información que forma parte del corpus del trabajo, facilitando el análisis y la lectura de los datos extraídos[4].

3.1.1 Análisis del Corpus Primario

Para llevar a cabo el módulo inicial que abarca la recopilación de información y los primeros análisis de la misma, se utilizó para comenzar las librerías *Pubmed.mineR* y *EasyPubmed*. Esta última permite buscar y guardar localmente los resúmenes (abstracts) de Pubmed con los términos que proponemos. Se realiza a través de la función *batch_pubmed_download()*. Se puede filtrar por argumentos, según los intereses del usuario.

Al buscar en Pubmed sobre hepatocarcinoma, han salido 120.965 resultados y la mayoría de ellos son de los últimos años, con lo cual se decidió restringir la búsqueda a los últimos 20 años y adicionar las publicaciones que a su vez estén relacionadas con miARN.



The screenshot shows the PubMed search interface. At the top, the NIH logo and 'National Library of Medicine' are visible. The search bar contains the query 'hepatocarcinoma cell OR hepatocellular carcinoma' and shows '120,965 results'. A bar chart on the left shows 'RESULTS BY YEAR' from 1946 to 2021, with a significant increase in the number of results starting around 2010. The first result is titled 'Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma' by Zheng H, Pomyen Y, Hernandez MO, et al., published in *Hepatology* in 2018. The abstract snippet indicates that intratumor molecular heterogeneity of hepatocellular carcinoma is partly attributed to the presence of hepatic cancer stem cells (CSCs).

Figura 4: Respuesta obtenida desde Pubmed

De este fichero descargado, se obtiene un objeto de tipo S4, que está formado por 3 slots

- ✓ **Journal:** vector con información sobre el autor, título, referencia, fecha de publicación.
- ✓ **Abstract:** vector con el resumen del artículo.
- ✓ **PMID:** vector con ID asignado por Pubmed.

Para obtener el corpus con el cual se harán los primeros análisis se utiliza la función *readabs()*, genera un nuevo fichero de texto tabulado llamado "newabs.txt". Para acceder a cada slot usamos @, por ejemplo corpus@journal.

3.1.2 Obtención de Corpus Secundario

Como se mencionó en capítulos anteriores se dividió el trabajo según tres módulos Genes-HCC, miARN-HCC Y mutaciones-HCC, todo lo que a continuación se presenta fue replicado para cada uno de ellos.

Se decidió avanzar en la búsqueda de los corpus secundarios, para esto, se utilizó la función *searchabsL()* que permite filtrar, a partir del corpus inicial obtenido, según la/s palabra/s que se le pase al vector "include". Para cada módulo se utilizaron las palabras "gene", "mir*" y "mutation" respectivamente.

Luego se obtuvo la frecuencia de los genes que aparecen en los corpus secundarios con la función *gene_atomization()*. Se obtiene como salida un fichero *table.txt* que se guarda en el disco local.

También se ha realizado tokenización de texto, para obtener palabras de alta frecuencia de ocurrencia. Esto se realizó con la función *word_atomizations()*. Se pasa como argumento el corpus secundario obtenido anteriormente y se obtiene un data frame con 2 columnas, la palabra y la frecuencia de la misma.

De esta forma se ha obtenido, por ejemplo, los miARN de mayor rango que luego se utilizaron en la aplicación.

3.2 Aplicación Shiny

Shiny es un paquete de R que permite construir aplicaciones web interactivas directamente desde R.

El script *app.R* tiene tres componentes:

- ✓ Un objeto de interfaz de usuario: *ui*
- ✓ Una función de servidor: *server*
- ✓ Una llamada a la función *shinyApp*.

El objeto *ui* controla el diseño y la apariencia de la aplicación, mientras que la función *server* contiene las instrucciones que la computadora necesita para construir la aplicación. Finalmente, la función *shinyApp* crea los objetos de la aplicación Shiny desde el par *ui/server*.

Shiny tiene un elemento web, llamado *Widget* con el que un usuario puede interactuar con la aplicación, proporcionándole valores de entrada (*input*). Cuando el usuario cambia

un widget, su valor también lo hace. Shiny viene con una familia de widgets incorporados, cada uno creado con una función R.

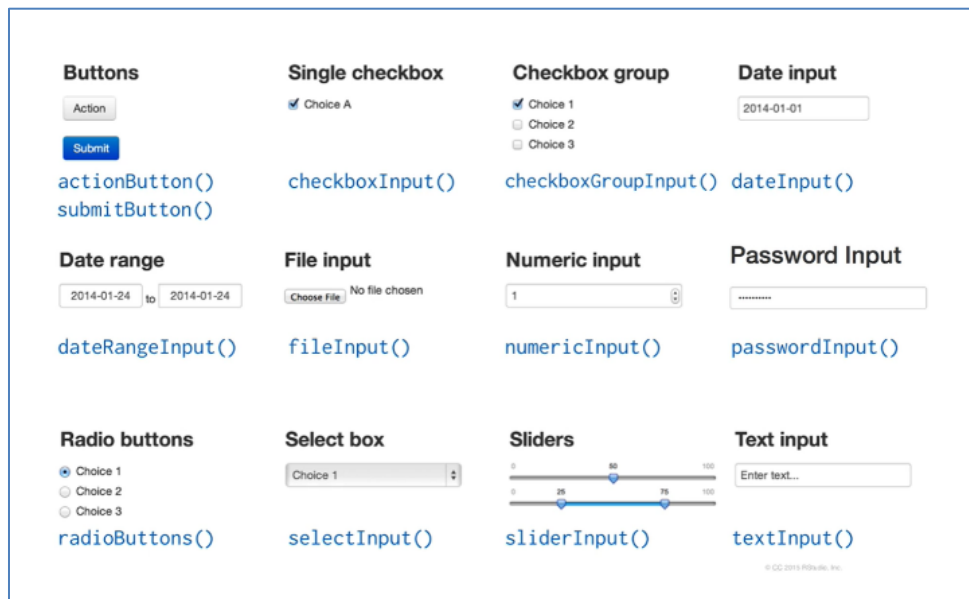


Figura 5: Widgets de Shiny estándar

Shiny vincula valores de entrada con los de salida(output) a partir de programación Reactiva.

Se pueden crear resultados reactivos en un proceso de dos pasos:

- ✓ Agregar un objeto R a la interfaz de usuario.
- ✓ Indicar a Shiny cómo construir el objeto en la función server. El objeto será reactivo si el código que lo construye usa el valor de un widget.

Shiny proporciona una familia de funciones que convierten los objetos R en salida para la interfaz de usuario. Cada función crea un tipo específico de salida[13].

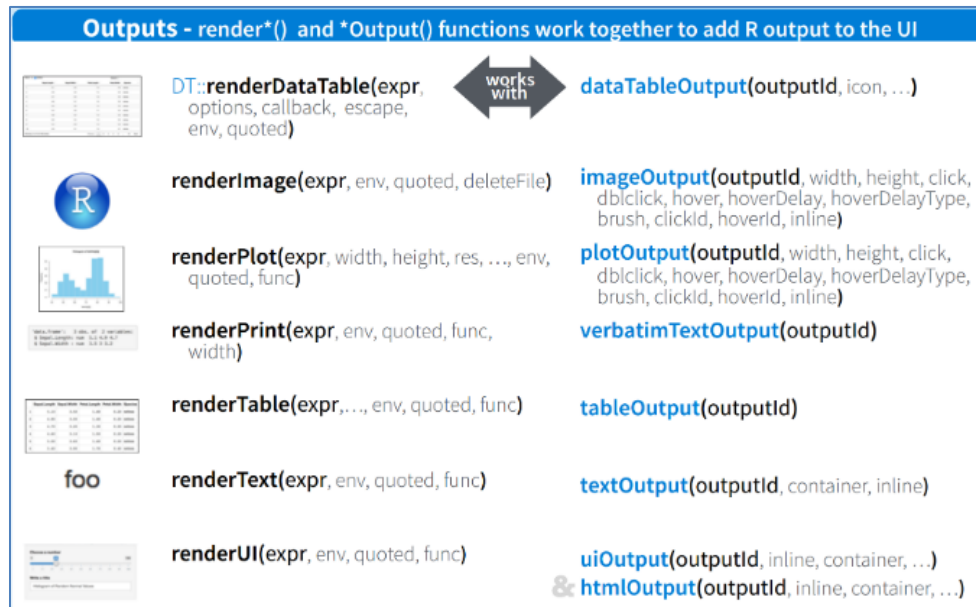


Figura 6: Salidas reactivas estándar

4.Desarrollo del Trabajo

Para el desarrollo de este trabajo se realizó una aplicación web interactiva llamada **HmiR²** en Shiny Dashboard. Se ha confeccionado con la estructura clásica que consta de la UI con una cabecera (`dashboardHeader`) que contiene el logo de la app,

```
ui <- dashboardPage(
  skin="purple",
  dashboardHeader(
    tags$li(
      class = "dropdown"
      ,tags$style(".main-header {max-height: 70px}")
      ,tags$style(".main-header .logo {height: 70px}")
    )
  ,
  title = titulo
),
```

una barra lateral (`dashboardSidebar`) con las opciones del menú, en este caso ítems con sus respectivos subítems desplegables

```
dashboardSidebar(
  sidebarMenu(
    br(),
    menuItem("Genes", tabName = "Genes", icon = icon("dna"), startExpanded = "TRUE",
      menuSubItem("Publicaciones en Pubmed", tabName = "pubmed1"),
      menuSubItem("GO", tabName = "GO1"),
      menuSubItem("Clinical Trials", tabName = "clinical")),
    menuItem("miARN", tabName = "miARN", icon = icon("vial"), startExpanded = "TRUE",
      menuSubItem("Publicaciones en Pubmed", tabName = "pubmed2"),
      menuSubItem("miRData", tabName = "miRData")),
    menuItem("Mutaciones", tabName = "Mutaciones", icon = icon("atom"), startExpanded = "TRUE",
      menuSubItem("Publicaciones en Pubmed", tabName = "pubmed3"),
      menuSubItem("My Cancer Genome", tabName = "cancer")),
    menuItem("Entorno HCC", tabName = "mapa", icon = icon("sitemap"), startExpanded = "TRUE")
  )
),
```

y el cuerpo (dashboardBody) donde se ingresan los inputs y se visualizan los resultados de las opciones elegidas, outputs.

```
dashboardBody(  
  useShinyjs(),  
  tags$style(type = "text/css",  
    ".shiny-output-error{visibility: hidden; }",  
    ".shiny-output-error:before{visibility:hidden; }"  
  ),  
  tabitems(  
    tabitem(tabName = "pubmed1",  
      h2("Busqueda de Genes en Pubmed"),  
      htmlOutput("genes"),  
      br(),  
      selectInput("pubmeda", "Selecciona un Gen", choices = choicesgen),  
      actionButton(inputId = "pub1", label = "Buscar", icon = icon("search"), block = FALSE),  
      add_busy_bar(color = "white"),  
      br(),  
      hidden(actionButton("nb", label = "Nueva Busqueda", block = FALSE)),  
      br(),br(),  
      tags$div(style= 'cursor:pointer',DT::dataTableOutput("resultados1")),  
      br(),  
      wellPanel(htmlOutput("panelPublicacionTMP"))  
    ),  
    tabitem(tabName = "pubmed2",  
      h2("Busqueda de miARN en Pubmed"),  
      htmlOutput("miarn"),  
      br().  
    ),  
    ...  
  )  
)
```

Como pantalla inicial se puede observa la siguiente presentación:

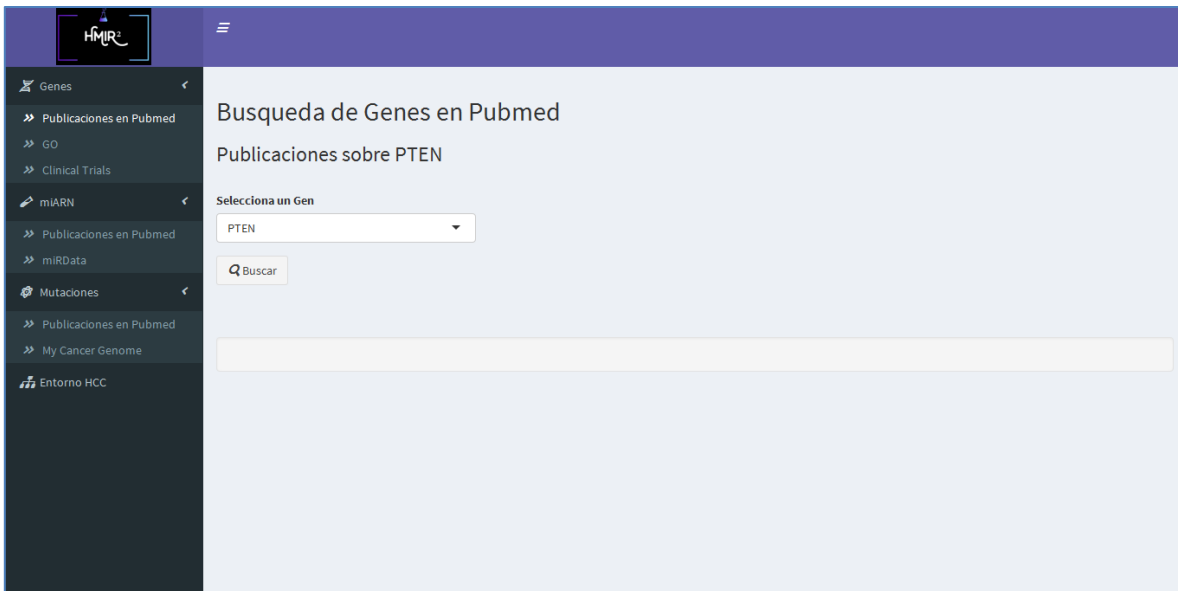


Figura 7: Pantalla inicial de HmiR²

La aplicación se ha dividido según los objetivos propuestos, con lo cual tenemos un ítem que trata de "genes", otro de "miARN" y sobre "mutaciones" . Como los procesos biológicos no son procesos aislados, sino que se relacionan e interactúan con otros genes y moléculas, se agregó el ítem "Entorno HCC", para dar una perspectiva biológica de la enfermedad.

4.1 Genes

Como primer subítem tenemos la búsqueda en Pubmed, que permite poner como input alguno de los genes obtenidos por minería de texto y filtrar sólo los artículos relacionados, a su vez se puede visualizar el resumen y se agregó un link con el que se puede acceder al artículo original publicado en Pubmed.

Publicaciones sobre PTEN

Nueva Busqueda

Mostrar 5 registros

Buscar:

Nro. Identificador	Revista
6	32163390 6. Mol Biol (Mosk). 2020 Jan-Feb;54(1):69-77. doi: 10.31857/S0026898420010152.
7	32133979 7. Cell Mol Biol (Noisy-le-grand). 2019 Dec 31;65(8):7-10.
8	32044726 8. Mol Ther Nucleic Acids. 2020 Mar 6;19:1073-1085. doi: 10.1016/j.omtn.2019.12.043.
9	32020591 9. J Cell Physiol. 2020 Sep;235(9):6154-6166. doi: 10.1002/jcp.29544. Epub 2020 Feb
10	31993313 10. Acta Pharm Sin B. 2020 Jan;10(1):159-170. doi: 10.1016/j.apsb.2019.11.001. Epub

Mostrando registros del 6 al 10 de un total de 92 registros

Anterior 1 2 3 4 5 ... 19 Siguiente

Focusing on the brighter side of Sevoflurane: Realizing true potential of an anesthetic agent as a regulator of cell signaling pathways and microRNAs in different cancers. Qureshi MZ(1), Attar R(2), Javed A(3), Sabitaliyevich UY(4), Adylova A(4), Konysbayevna KK(5), Buha A(6), Sohail MI(7), Aras A(8). Author information: (1)Deanship of Educational services, Department of Biochemistry, Qassim University, Saudi Arabia. (2)Department of Obstetrics and Gynecology, Yeditepe University, Turkey. (3)Wexham Park Hospital, Wexham Street, Slough, Berkshire, SL2 4HL, United Kingdom. (4)Department of Postgraduate Education and Research, Kazakhstan Medical University KSPH, Almaty, Kazakhstan. (5)Ministry of Health of the Republic of Kazakhstan. Zhibek zholy 5, Almaty, Kazakhstan. (6)Department of Toxicology "Akademik Danilo Soldatovic", University of Belgrade-Faculty of Pharmacy, Serbia. (7)Department of Zoology, GCU Lahore, Pakistan. (8)Department of Botany, Faculty of Science, Istanbul University, Istanbul, Turkey. Reconceptualization of different anesthetics as anticancer agents has opened new horizons for a better and sharper analysis of true potential of Sevoflurane as a promising and frontline candidate in the pipeline of anticancer agents. Sevoflurane mediated regulation of cell signaling pathways and non-coding RNAs has leveraged our understanding to another level. There have been remarkable advancements in unraveling mechanistic insights related to the ability of sevoflurane to modulate microRNAs in different cancers. Astonishingly, sevoflurane mediated regulation of miRNAs and long non-coding RNAs have been more comprehensively addressed in ischemia-reperfusion injuries. However, researchers yet have to gather missing pieces of premium research work to uncover mechanistic regulation of long non-coding RNAs by sevoflurane in various cancers.

Figura 8: Búsqueda de Pubmed de genes

Para realizar esta búsqueda y descarga en Pubmed, en el server se registra el siguiente código, con las funciones y librerías comentadas en capítulos anteriores.

```
PubMedResult<-eventReactive(input$pub1,{
  PM1 <- batch_pubmed_download(
    pubmed_query_string = paste(c(input$pubmeda, " [MH] AND" , new_query),collapse = ""),
    format = "abstract",
    batch_size = 5000,
    dest_file_prefix = "pubmed_"
  )
  pmr<-readabs("pubmed_01.txt")
})
```

Cuando el usuario elige el gen y pulsa "buscar" se activa *eventReactive()* , el fichero es leído por *readabs()*.

El output se visualiza en forma de tabla, esto se logra con el objeto `output$resultados1` y la función `renderDataTable()`.

```
output$resultados1<-renderDataTable({
  corpus1<- PubmedResult()
  pmes1<-data.frame(corpus1@PMID, corpus1@Journal)
  colnames(pmes1)<-c("Nro. Identificador", "Revista")
  datatable(pmes1, selection=list(mode="single", selected = c(1)
), options = list(language= list(url = '//cdn.datatables.net/plug-ins/1.10.11/i18n/Spanish.json'),pageLength = 5))
})
```

Para visualizar el abstract elegido, se utilizó la función `renderPrint()` y poder acceder a la página de Pubmed para ver el artículo original

```
output$panelPublicacionTMP<-renderPrint({
  s = input$resultados1_row_last_clicked
  if (length(s)) {
    corpus1<-readabs("pubmed_01.txt")
    abstract <- corpus1@Abstract[s]
    id <- corpus1@PMID[s]
    link <- a(
      "Ver Publicacion en PubMed"
      ,href= gsub(
        ""
        ,""
        ,paste(
          "https://pubmed.ncbi.nlm.nih.gov/"
          ,corpus1@PMID[s]
        )
        , fixed = TRUE)
      , target="_blank"
    )
    cat(
      abstract
    )
    cat(
      paste("<br>", link)
    )
  }
})
```

Como segundo subítem, la integración con la base de datos GO. Como la búsqueda es a través de los identificadores Entrez se han obtenido primero los correspondientes ID.

```
my.symbols <- c("PTEN", "AFP", "STAT3", "SIRT7", "GC", "XPO5", "HMGA2", "VEGFA", "FOXO1", "ZEB1", "PDCD4",
"GPC3", "EGFR", "SIRT1", "MET", "DNMT1", "IGF1R", "TET1", "CDC42", "KLF5", "GRN", "FOXA1", "MACC1", "SOCS1",
"TP53", "E2F1", "E2F3", "ROCK1", "SP1", "CTGF", "DICER1", "HGF", "ROCK2", "BMI1", "CDK1", "FUT8", "METTL3",
"PIK3R3", "AKAP12", "AURKA", "CTNNB1", "EGFL7", "HDAC6", "IGF2", "IGF2BP1", "SMAD2", "SOCS3", "UBE2I",
"VDAC1", "ZEB2", "ABC1", "AKT2", "CCNE1", "FBX011")
entrez_id<-select(org.Hs.eg.db,
  keys = my.symbols,
  columns = c("ENTREZID", "SYMBOL"),
  keytype = "SYMBOL")
```

Luego se han puesto las referencias de cada gen. Como output se obtienen todos los resultados registrados que coinciden con el identificador GO.

Identificadores Entrez

Selecciona un gen

Referencias de Identificadores

SYMBOL	ENTREZID	SYMBOL	ENTREZID	SYMBOL	ENTREZID	SYMBOL	ENTREZID	SYMBOL	ENTREZID
PTEN	5738	FOXC4	17250	GRI1	2096	ZKCR1	22405	CTNBL1	1499
JAP	174	GPC3	2719	FOXA1	3169	HGF	3082	ESPL7	5162
STEB	6774	EGFR	1956	MCC1	34639	ROCK2	9475	HDA35	10013
SIRT7	51547	SIRT1	29411	SOC1	8651	EMIL	648	HGF2	3481
GC	2638	MET	4233	TP53	7157	CDYL	983	HGF3P1	10642
XPOS	57510	DNAH1	1786	EPF1	1889	FUT8	2530	SMAD2	4087
HMG42	8091	HGF1R	3480	EPF3	1871	METL3	56339	SOC3	9021
VEGFA	7422	TEL	80312	ROCK1	6093	PKR3	8503	LIBZ1	7329
FOI1	2308	CDC42	998	SP1	6667	AKAP12	9590	VDAC1	7416
ZEB1	6935	KLF5	688	CTGF	NA	ALRKA	6790	ZEB2	9839

Resultados GO Similitud Genes

Mostrar registros Buscar:

ID GO	Termino	Definicion	IC	
GO:0001525	GO:0001525	angiogenesis	Blood vessel formation when new vessels emerge from the proliferation of pre-existing blood vessels.	5.38726510287739
GO:0001933	GO:0001933	negative regulation of protein phosphorylation	Any process that stops, prevents or reduces the rate of addition of phosphate groups to amino acids within a protein.	5.95871534845982
GO:0002902	GO:0002902	regulation of B cell apoptotic process	Any process that modulates the frequency, rate, or extent of B cell apoptotic process.	8.95444762201381
		transcription initiation from RNA	Any process involved in the assembly of the RNA polymerase II preinitiation complex (PIC) at an RNA polymerase II promoter region of a DNA template, resulting in the subsequent synthesis of RNA from that promoter. The initiation phase includes PIC assembly and the formation of the first few bonds in the RNA	

Figura 9: Resultados de búsqueda GO

Para poder mostrar los output en el servidor se ha colocado el siguiente código

```
#GO
output$resultadosgo<-renderDataTable({input$radio
x<-as.data.frame(t(getGOInfo(input$radio)))
colnames(x)<-c("ID GO", "Termino", "Definicion", "IC")
datatable(x, selection=list(mode="single", selected = c(1)
), options = list(language= list(url = '//cdn.datatables.net/plug-ins/1.10.11/i18n/Spanish.json'),pageLength = 5))
})
```

Para que la salida sea más prolija también se visualiza en forma de tabla con lo cual se utilizó *renderDataTable()*. Se puede ver a partir de un gen elegido en *input\$radio*, los términos GO relacionados y la definición. Esto se logra gracias a la función *getGOInfo()* de la librería *GOsim*.

4.2 miARN

Como primer ítem se obtuvo de igual manera los artículos correspondientes a los miARN obtenidos por minería de texto. El orden en que se presentan es según

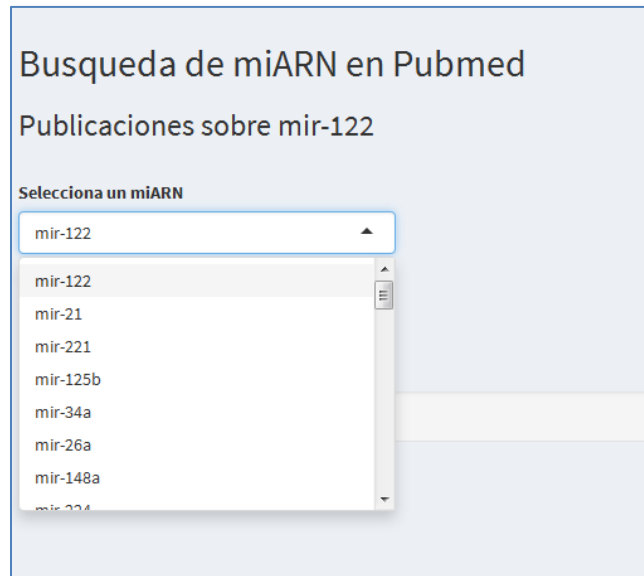


Figura 10: Input miARN

Luego de elegir un input se obtienen los artículos referentes, se puede ver el resumen del mismo e ingresar al artículo original en la página de Pubmed.

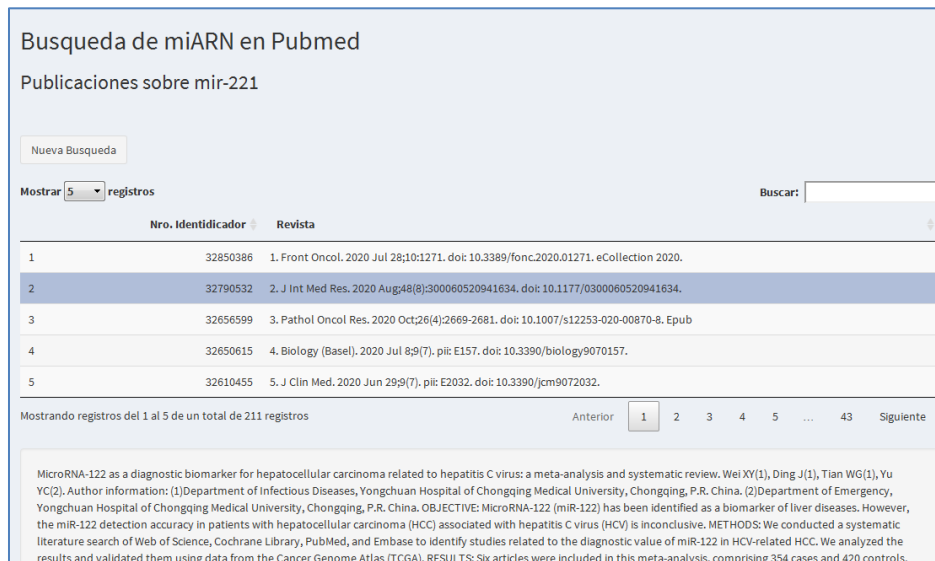


Figura 11: Búsqueda de Pubmed de miARN

Para este primer ítem se realizó en el server lo mismo explicado anteriormente para el módulo Genes.

Otro de los subítems elegidos fue poder brindar información relevante sobre cada miARN y los genes diana a los que se dirigen estos miARN.

Por un lado, a partir de un gen elegido se obtiene información de los miARN involucrados que los regulan.

Información correspondiente a miARN

A partir de un gen elegido, se muestra todos los miARN que se dirigen a el

Selecciona un gen

A partir de un miARN, se muestra todos los genes dianas a los que se dirigen

Escribe el miARN

Resultados ARN Genes Diana

show 10 entries Search:

	miARN involucrado	Gen	Score	Tipo
1	hsa-miR-1297	PTEN	0.999	predicted
2	hsa-miR-26b-5p	PTEN	0.999	predicted
3	hsa-miR-26a-5p	PTEN	0.999	predicted
4	hsa-miR-548m	PTEN	0.999	predicted
5	hsa-miR-3148	PTEN	0.999	predicted
6	hsa-miR-4775	PTEN	0.998	predicted

Figura 12: miARN involucrados

En el servidor se ha colocado la función `get_multimir()` del paquete `multiMiR` que permite obtener los miARN involucrados, según el gen elegido, por eso se coloca en el parámetro `target= input$microgen`. Como esto devuelve un data frame con columnas que no son relevantes para este trabajo se han quitado algunas.

```

#miARN
output$publicacionesarn <- renderDataTable({
  example3 <- get_multimir(org = "hsa",
    target = input$microgen,
    table = "predicted",
    summary = TRUE,
    predicted.cutoff = 35,
    predicted.cutoff.type = "p",
    predicted.site = "all"
  )
  tablamir <- example3@data[,c(-1,-2,-5:-6)]
  colnames(tablamir) <- c("miARN involucrado", "Gen", "score", "Tipo")
  datatable(tablamir, selection=list(mode="single", selected = c(1)
  ), options = list(language= list(url = '//cdn.datatables.net/plug-ins/1.10.11/i18n/Spanish.json'), pageLength = 5))
  tablamir
})

```

Por otro se obtiene a partir de un miARN que se escribe como input, los genes a los que se dirige para regularlos.

A partir de un gen elegido, se muestra todos los miARN que se dirigen a el

Selecciona un gen

A partir de un miARN, se muestra todos los genes dianas a los que se dirigen

Escribe el miARN

Resultados ARN
Genes Diana

how entries Search:

	Base de datos	miARN involucrado	Gen	Técnica	ID Pubmed	Tipo
1	mirecords	hsa-miR-338-3p	UBE2Q1	Western blot	18684991	validated
2	mirecords	hsa-miR-338-3p	UBE2Q1	Western blot	18684991	validated
3	mirecords	hsa-miR-338-3p	UBE2Q1	Western blot	18684991	validated
4	mirtarbase	hsa-miR-338-3p	ACVR1	HITS-CLIP	23313552	validated
5	mirtarbase	hsa-miR-338-3p	ADCY7	PAR-CLIP	22012620	validated
6	mirtarbase	hsa-miR-338-3p	GRK2	PAR-CLIP	26701625	validated
7	mirtarbase	hsa-miR-338-3p	ALOX5AP	PAR-CLIP	23446348	validated
8	mirtarbase	hsa-miR-338-3p	CCND1	Luciferase reporter assay	22942717	validated

Figura 13: Genes diana

Utilizando la misma función y librería anterior pero cambiando los parámetros se obtiene la tabla presentada. Se han retirado las columnas que no aportan información útil en este trabajo.

```

output$publicaciondiana<-renderDataTable({
  genesDiana<- get_multimir(mirna = input$inputarn, summary = TRUE)
  tablagen<-genesDiana@data[,c(-2,-5,-6,-8)]
  colnames(tablagen)<-c("Base de datos", "miARN involucrado", "Gen", "Técnica", "ID Pubmed", "Tipo")
  datatable(tablagen, selection=list(mode="single", selected = c(1)
  ), options = list(language= list(url = '//cdn.datatables.net/plugin-ins/1.10.11/i18n/spanish.json'),pageLength = 5))
  tablagen
})

```

4.3 Mutaciones

En el primer ítem se obtienen las publicaciones relacionadas a las mutaciones que están involucradas en el HCC. Estas se obtuvieron por minería de texto. También se puede visualizar los resúmenes seleccionados e ingresar en la página de Pubmed para ver los originales.

Busqueda de mutaciones en Pubmed

Mostrar registros

	Nro. Identificador	Revista
1	33240468	1. Comput Struct Biotechnol J. 2020 Oct 21;18:3243-3254. doi:
2	32424223	competing financial interests or personal relationships that could have appeared
3	32333311	3. Dig Dis Sci. 2020 Apr 24. doi: 10.1007/s10620-020-06260-z. [Epub ahead of print]
4	32279994	4. Biochem Biophys Res Commun. 2020 Jun 11;526(4):889-897. doi:
5	31615341	declare that they have no known competing financial interests or personal

Mostrando registros del 1 al 5 de un total de 59 registros

 Anterior 2 3 4 5 ... 12 Siguiente

to influence the work reported in this paper. 2. Sci Rep. 2020 May 18;10(1):8181. doi: 10.1038/s41598-020-65213-8. Clinical significance of combined circulating TERT promoter mutations and miR-122 expression for screening HBV-related hepatocellular carcinoma. Trung NT(1)(7). Hoan NX(3)(4). Trung PO(5)(3). Binh MT(6). Van Tonge H(7)(8). Toan NI(8).

Figura 14: Búsqueda de Pubmed de mutaciones

Al igual que pasó anteriormente para este ítem se ha utilizado el mismo código, en el server, que el utilizado en el módulo genes.

A partir de los análisis realizados con el corpus secundario se obtuvieron los genes mutados que con más frecuencia aparecieron. Según el gen elegido como input, se obtiene en una tabla información relevante sobre la mutación, por ejemplo el tipo de mutación, el cromosoma donde se encuentra el gen, etc.

Selecciona un gen

Resultados

Show entries

	Entrez ID	Gen	Tipo de mutacion	Cambio en el aa.	Nro Chr	Posicion Inicio	Posicion Final	Referencia	Variante
1	23405	DICER1	Missense_Mutation	H356L	14	95590842	95590842	T	A
2	23405	DICER1	Splice_Site	X1699_splice	14	95560494	95560494	C	A
3	23405	DICER1	Missense_Mutation	D1875N	14	95556981	95556981	C	T
4	23405	DICER1	Missense_Mutation	E1843D	14	95557445	95557445	T	A
5	23405	DICER1	Missense_Mutation	Y793C	14	95574719	95574719	T	C
6	23405	DICER1	Splice_Region	A752=	14	95577654	95577654	T	A
7	23405	DICER1	Frame_Shift_Del	F1666Lfs*39	14	95562262	95562262	C	-

Showing 1 to 7 of 7 entries

 Previous Next

Figura 15: Información sobre mutaciones

Para obtener esta información se recurrió a la librería *CGDS*. Como obtiene la información de varias bases de datos desde su portal con la función del mismo nombre se menciona el link al que recurre y luego con *getMutationData()*, obtenemos la tabla con la información. Como parámetro se menciona qué tipo de datos necesito en este caso mutaciones y sobre la enfermedad en cuestión y en genes utilizamos los elegidos por el usuario en *input\$mutacion*. El código utilizado fue el siguiente en el servidor

```
#Mutaciones
output$publicacionmut <- renderDataTable({
  mycgds = CGDS("http://www.cbioportal.org/")

  y = getMutationData(mycgds, caseList = "lihc_tcga_all", geneticProfile = "lihc_tcga_mutations", genes = input$mutacion)

  tabla_mut <- y[, c(-3:-5, -7, -9:-12, -14-15, -18:-22)]
  colnames(tabla_mut) <- c("Entrez ID", "Gen", "Tipo de mutacion", "Cambio en el aa.", "Nro Chr", "Posicion Inicio", "Posicion Final")
  datatable(y, selection=list(mode="single", selected = c(1))
  ), options = list(language= list(url = '//cdn.datatables.net/plug-ins/1.10.11/i18n/Spanish.json'), pageLength = 5))
  tabla_mut
})
```

4.4 Entorno HCC

A modo adicional y como referencia se incorporó una imagen obtenida de la Enciclopedia de genes y genomas de Kioto. Si se desea más información se ofrece un link para ingresar a la página web de KEGG.

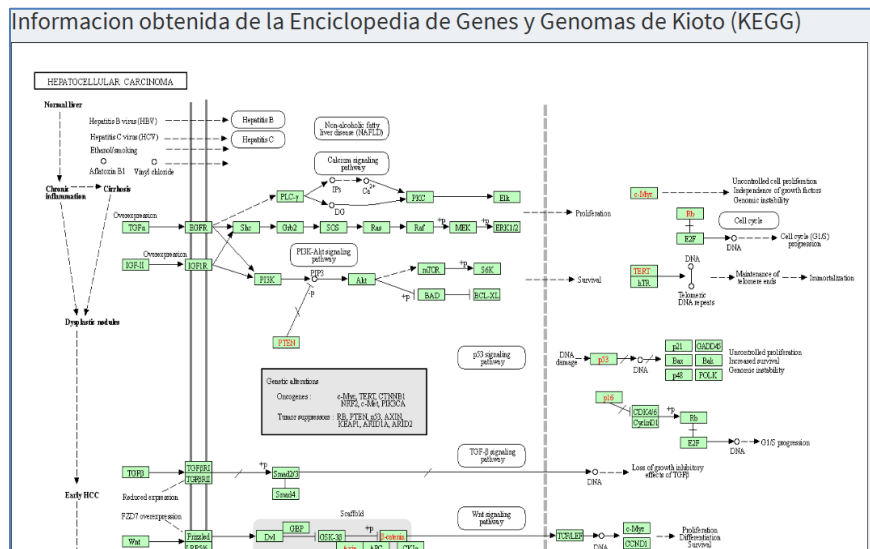


Figura 16: Información sobre el entorno HCC

Al tratarse sólo de una imagen, se puede visualizar al colocar en la UI, *img*, pasando como argumento *src* con el nombre de la imagen en formato .png y previamente colocada en la carpeta "www", que es donde busca este tipo de recurso digital.

```

tabItem(tabName = "mapa",
  h2("Información obtenida de la Enciclopedia de Genes y Genomas de Kioto (KEGG)",
  tags$div(img(src = "hsa05225.png", height = 890, width = 1060)),
  wellPanel(uioutput("mapakegg")))
),

```

Para poder realizar el vínculo con la página KEGG se ha colocado el siguiente código en el servidor

```

hiperlink <- a("Ver detalles del mapa en la pagina web de KEGG",href=
  "https://www.genome.jp/dbget-bin/www_bget?pathway+hsa05225")

output$mapakegg<-renderUI({
  tagList("URL link:", hiperlink)
})

```

5.Resultados

Por una cuestión de extensión sólo se mostrarán los primeros 3 resultados a modo representativo. Por otro lado, como forma de unificar ciertas funciones, se ha trabajado con genes que pertenecen al *Homo sapiens*.

5.1 Genes-HCC

A partir del análisis realizado con minería de texto se ha obtenido que estos son los genes que con más frecuencia aparecen en las publicaciones relacionadas con HCC y miARN.

	Gene_symbol	Genes	Freq
1	PTEN	phosphatase and tensin homolog	159
2	AFP	alpha-fetoprotein	92
3	EZH2	enhancer of zeste homolog 2 (Drosophila)	62
4	T	T, brachyury homolog (mouse)	56
5	STAT3	signal transducer and activator of transcription 3 (acute-...	51
6	SIRT7	sirtuin 7	39
7	GC	group-specific component (vitamin D binding protein)	38
8	XPO5	exportin 5	35
9	HMGGA2	high mobility group AT-hook 2	34
10	VEGFA	vascular endothelial growth factor A	33
11	FOXO1	forkhead box O1	32
12	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	31
13	ZEB1	zinc finger E-box binding homeobox 1	30
14	PDCC4	programmed cell death 4 (neoplastic transformation inhi...	29

Figura 17: Resultados frecuencia de genes

En cuanto a publicaciones para el gen **PTEN** tenemos un total de 92 publicaciones

Busqueda de Genes en Pubmed

Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma descendiente

Publicaciones sobre PTEN

Nueva Busqueda

Mostrar registros Buscar:

Nro.	Identificador	Revista
1	32974191	1. Front Oncol. 2020 Aug 19;10:1563. doi: 10.3389/fonc.2020.01563. eCollection 2020.
2	32850833	2. Front Cell Dev Biol. 2020 Jul 31;8:714. doi: 10.3389/fcell.2020.00714.
3	32774734	3. Am J Transl Res. 2020 Jul 15;12(7):3780-3791. eCollection 2020.
4	32319540	
5	32167152	5. Biosci Rep. 2020 Mar 27;40(3). pii: BSR-20182466_RET. doi:

Mostrando registros del 1 al 5 de un total de 92 registros Anterior 2 3 4 5 ... 19 Siguiente

2. Front Cell Dev Biol. 2020 Jul 31;8:714. doi: 10.3389/fcell.2020.00714.

Figura 18: Búsqueda publicaciones PTEN

Para el gen **AFP** también 92 publicaciones

Busqueda de Genes en Pubmed

Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma descendiente

Publicaciones sobre AFP

Nueva Busqueda

Mostrar **5** registros Buscar:

	Nro. Identificador	Revista
1	32974191	1. Front Oncol. 2020 Aug 19;10:1563. doi: 10.3389/fonc.2020.01563. eCollection 2020.
2	32850833	2. Front Cell Dev Biol. 2020 Jul 31;8:714. doi: 10.3389/fcell.2020.00714.
3	32774734	3. Am J Transl Res. 2020 Jul 15;12(7):3780-3791. eCollection 2020.
4	32319540	
5	32167152	5. Biosci Rep. 2020 Mar 27;40(3). pii: BSR-20182466_RET. doi:

Mostrando registros del 1 al 5 de un total de 92 registros Anterior **1** 2 3 4 5 ... 19 Siguiente

[2. Front Cell Dev Biol. 2020 Jul 31;8:714. doi: 10.3389/fcell.2020.00714.](#)

Figura 19: Búsqueda publicaciones AFP

Para el gen **STAT3** contamos con 22 registros

Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma descendiente

Publicaciones sobre STAT3

Nueva Busqueda

Mostrar **5** registros Buscar:

	Nro. Identificador	Revista
1	31648164	1. Biomed Pharmacother. 2019 Dec;120:109551. doi: 10.1016/j.biopha.2019.109551. Epub
2	30981758	2. Cancer Lett. 2019 Jul 10;454:158-170. doi: 10.1016/j.canlet.2019.04.008. Epub
3	30370582	3. J Cell Physiol. 2019 Apr;234(4):3500-3514. doi: 10.1002/jcp.26863. Epub 2018 Oct
4	30274685	4. Pathol Res Pract. 2018 Dec;214(12):1980-1986. doi: 10.1016/j.prp.2018.08.005.
5	30132576	5. Oncol Rep. 2018 Nov;40(5):2806-2813. doi: 10.3892/or.2018.6651. Epub 2018 Aug 17.

Mostrando registros del 1 al 5 de un total de 22 registros Anterior **1** 2 3 4 5 Siguiente

[2. Front Cell Dev Biol. 2020 Jul 31;8:714. doi: 10.3389/fcell.2020.00714.](#)

Figura 20: Búsqueda publicaciones STAT3

Al analizar los resultados GO obtenemos para el gen **PTEN**

GO:0002902	GO:0002902	regulation of B cell apoptotic process	Any process that modulates the frequency, rate, or extent of B cell apoptotic process.	8.95444762201381
GO:0006915	GO:0006915	apoptotic process	A programmed cell death process which begins when a cell receives an internal (e.g. DNA damage) or external signal (e.g. an extracellular death ligand), and proceeds through a series of biochemical events (signaling pathway phase) which trigger an execution phase. The execution phase is the last step of an apoptotic process, and is typically characterized by rounding-up of the cell, retraction of pseudopodes, reduction of cellular volume (pyknosis), chromatin condensation, nuclear fragmentation (karyorrhexis), plasma membrane blebbing and fragmentation of the cell into apoptotic bodies. When the execution phase is completed, the cell has died.	3.50340916844811
GO:0010666	GO:0010666	positive regulation of cardiac muscle cell apoptotic process	Any process that increases the rate or extent of cardiac cell apoptotic process, a form of programmed cell death induced by external or internal signals that trigger the activity of proteolytic caspases whose actions dismantle a cardiac muscle cell and result in its death.	11.4393542718018
GO:0033032	GO:0033032	regulation of myeloid cell apoptotic process	Any process that modulates the frequency, rate, or extent of myeloid cell apoptotic process.	9.0414589900344
GO:0043066	GO:0043066	negative regulation of apoptotic process	Any process that stops, prevents, or reduces the frequency, rate or extent of cell death by apoptotic process.	4.85318261694713

Figura 21: Resultados obtenidos de la base de datos GO para PTEN

Para el gen **AFP**

GO:0001542	GO:0001542	ovulation from ovarian follicle	The process leading to the rupture of the follicle, releasing the centrally located oocyte into the oviduct. An example of this is found in <i>Mus musculus</i> .	8.95444762201381
GO:0006810	GO:0006810	transport	The directed movement of substances (such as macromolecules, small molecules, ions) or cellular components (such as complexes and organelles) into, out of or within a cell, or between cells, or within a multicellular organism by means of some agent such as a transporter, pore or motor protein.	2.63083565301899
GO:0042448	GO:0042448	progesterone metabolic process	The chemical reactions and pathways involving progesterone, a steroid hormone produced in the ovary which prepares and maintains the uterus for pregnancy. Also found in plants.	8.87440491434027
GO:0043687	GO:0043687	post-translational protein modification	The process of covalently altering one or more amino acids in a protein after the protein has been completely translated and released from the ribosome.	6.10663547853644
GO:0044267	GO:0044267	cellular protein metabolic process	The chemical reactions and pathways involving a specific protein, rather than of proteins in general, occurring at the level of an individual cell. Includes cellular protein modification.	2.58041704983526
GO:0060395	GO:0060395	SMAD protein signal transduction	The cascade of processes by which a signal interacts with a receptor, causing a change in the activity of a SMAD protein, and ultimately effecting a change in the functioning of the cell.	8.60614092774559

Figura 22: Resultados obtenidos de la base de datos GO para AFP

Para el gen **STAT3**

GO:0000122	GO:0000122	negative regulation of transcription by RNA polymerase II	Any process that stops, prevents, or reduces the frequency, rate or extent of transcription mediated by RNA polymerase II.	5.29088597588416
GO:0006355	GO:0006355	regulation of transcription, DNA-templated	Any process that modulates the frequency, rate or extent of cellular DNA-templated transcription.	3.03667446717433
GO:0006357	GO:0006357	regulation of transcription by RNA polymerase II	Any process that modulates the frequency, rate or extent of transcription mediated by RNA polymerase II.	4.14565655120037
GO:0007165	GO:0007165	signal transduction	The cellular process in which a signal is conveyed to trigger a change in the activity or state of a cell. Signal transduction begins with reception of a signal (e.g. a ligand binding to a receptor or receptor activation by a stimulus such as light), or for signal transduction in the absence of ligand, signal-withdrawal or the activity of a constitutively active receptor. Signal transduction ends with regulation of a downstream cellular process, e.g. regulation of transcription or regulation of a metabolic process. Signal transduction covers signaling from receptors located on the surface of the cell and signaling via molecules located within the cell. For signaling between cells, signal transduction is restricted to events at and within the receiving cell.	2.22522110158693
GO:0007259	GO:0007259	receptor signaling pathway via JAK-STAT	Any process in which STAT proteins (Signal Transducers and Activators of Transcription) and JAK (Janus Activated Kinase) proteins convey a signal to trigger a change in the activity or state of a cell. The receptor signaling pathway via JAK-STAT begins with activation of a receptor and proceeds through STAT protein activation by members of the JAK family of tyrosine kinases. STAT proteins dimerize and subsequently translocate to the nucleus. The pathway ends with regulation of target gene expression by STAT proteins.	6.25196846596105

Figura 23: Resultados obtenidos de la base de datos GO para STAT3

5.2 miARN-HCC

words	Freq
mir-	All
mir-122	914
mir-21	403
mir-221	212
mir-125b	197
mir-34a	197
mir-26a	159
mir-148a	146
mir-224	145
mir-101	140
mir-145	137
mir-155	136
mir-146a	131
mir-375	127
mir-7	127
mir-195	126
mir-214	126

Luego del análisis realizado con minería de texto se ha obtenido las palabras que con más frecuencia aparecen en las publicaciones relacionadas con HCC y miARN, a partir de esta lista se ha filtrado para obtener cuales son los miARN que aparecen, dando lugar a los que se muestran.

Figura 24: Resultados de miARN más frecuentes

Las publicaciones que pudimos obtener de los miR más frecuentes fue para **miR-122**, hay 211 registros

Publicaciones sobre mir-122

Nueva Búsqueda

Mostrar 5 registros

Buscar:

	Nro. Identificador	Revista
1	32850386	1. Front Oncol. 2020 Jul 28;10:1271. doi: 10.3389/fonc.2020.01271. eCollection 2020.
2	32790532	2. J Int Med Res. 2020 Aug;48(8):300060520941634. doi: 10.1177/0300060520941634.
3	32656599	3. Pathol Oncol Res. 2020 Oct;26(4):2669-2681. doi: 10.1007/s12253-020-00870-8. Epub
4	32650615	4. Biology (Basel). 2020 Jul 8;9(7). pii: E157. doi: 10.3390/biology9070157.
5	32610455	5. J Clin Med. 2020 Jun 29;9(7). pii: E2032. doi: 10.3390/jcm9072032.

Mostrando registros del 1 al 5 de un total de 211 registros

Anterior 1 2 3 4 5 ... 43 Siguiente

2. J Int Med Res. 2020 Aug;48(8):300060520941634. doi: 10.1177/0300060520941634.

Figura 25: Búsqueda publicaciones miR-122

Para miR-21 un total de 211 registros

Publicaciones sobre mir-21

Selecciona un miARN

mir-21

Buscar

Mostrar 5 registros

Buscar:

	Nro. Identificador	Revista
1	32850386	1. Front Oncol. 2020 Jul 28;10:1271. doi: 10.3389/fonc.2020.01271. eCollection 2020.
2	32790532	2. J Int Med Res. 2020 Aug;48(8):300060520941634. doi: 10.1177/0300060520941634.
3	32656599	3. Pathol Oncol Res. 2020 Oct;26(4):2669-2681. doi: 10.1007/s12253-020-00870-8. Epub
4	32650615	4. Biology (Basel). 2020 Jul 8;9(7). pii: E157. doi: 10.3390/biology9070157.
5	32610455	5. J Clin Med. 2020 Jun 29;9(7). pii: E2032. doi: 10.3390/jcm9072032.

Mostrando registros del 1 al 5 de un total de 211 registros

Anterior 1 2 3 4 5 ... 43 Siguiente

2. J Int Med Res. 2020 Aug;48(8):300060520941634. doi: 10.1177/0300060520941634.

Figura 26: Búsqueda publicaciones miR-21

Para miR-211, 211 publicaciones

Publicaciones sobre mir-221

Nueva Búsqueda

Mostrar registros

Buscar:

	Nro. Identificador	Revista
1	32850386	1. Front Oncol. 2020 Jul 28;10:1271. doi: 10.3389/fonc.2020.01271. eCollection 2020.
2	32790532	2. J Int Med Res. 2020 Aug;48(8):300060520941634. doi: 10.1177/0300060520941634.
3	32656599	3. Pathol Oncol Res. 2020 Oct;26(4):2669-2681. doi: 10.1007/s12253-020-00870-8. Epub
4	32650615	4. Biology (Basel). 2020 Jul 8;9(7). pii: E157. doi: 10.3390/biology9070157.
5	32610455	5. J Clin Med. 2020 Jun 29;9(7). pii: E2032. doi: 10.3390/jcm9072032.

Mostrando registros del 1 al 5 de un total de 211 registros

Anterior 2 3 4 5 ... 43 Siguiente

[2. J Int Med Res. 2020 Aug;48\(8\):300060520941634. doi: 10.1177/0300060520941634.](#)

MicroRNA-122 as a diagnostic biomarker for hepatocellular carcinoma related to hepatitis C virus: a meta-analysis and systematic review. Wei XY(1), Ding J(1), Tian WG(1), Yu

Figura 27: Búsqueda publicaciones miR-221

A partir de los genes obtenidos, que podrían sufrir mutaciones se pudo ver cuáles son los miARN que lo regulan, para el gen **DICER1** se obtuvo

	miARN involucrado	Gen	Score	Tipo
1	hsa-miR-1270	AFP	0.975	predicted
2	hsa-miR-3925-5p	AFP	0.973	predicted
3	hsa-miR-1323	AFP	0.944	predicted
4	hsa-miR-548a-3p	AFP	0.91	predicted
5	hsa-miR-548f-3p	AFP	0.901	predicted
6	hsa-miR-548o-3p	AFP	0.892	predicted
7	hsa-miR-620	AFP	0.887	predicted
8	hsa-miR-216a-5p	AFP	0.877	predicted
9	hsa-miR-4311	AFP	0.862	predicted
10	hsa-miR-146a-5p	AFP	0.853	predicted

Figura 28: Resultados obtenidos de miARN involucrados con el gen DICER1

Para el gen **AFP**

	miARN involucrado	Gen	Score	Tipo
1	hsa-miR-1270	AFP	0.975	predicted
2	hsa-miR-3925-5p	AFP	0.973	predicted
3	hsa-miR-1323	AFP	0.944	predicted
4	hsa-miR-548a-3p	AFP	0.91	predicted
5	hsa-miR-548f-3p	AFP	0.901	predicted
6	hsa-miR-548o-3p	AFP	0.892	predicted
7	hsa-miR-620	AFP	0.887	predicted
8	hsa-miR-216a-5p	AFP	0.877	predicted
9	hsa-miR-4311	AFP	0.862	predicted
10	hsa-miR-146a-5p	AFP	0.853	predicted

Figura 29: Resultados obtenidos de miARN involucrados con el gen AFP

Para el gen **GnMT**

show 10 entries		Search:		
	miARN involucrado	Gen	Score	Tipo
1	hsa-miR-873-5p	GNMT	0.793	predicted
2	hsa-miR-1324	GNMT	0.735	predicted
3	hsa-miR-548u	GNMT	0.669	predicted
4	hsa-miR-522-5p	GNMT	0.365	predicted
5	hsa-miR-518f-5p	GNMT	0.365	predicted
6	hsa-miR-519c-5p	GNMT	0.365	predicted
7	hsa-miR-518d-5p	GNMT	0.365	predicted
8	hsa-miR-520c-5p	GNMT	0.365	predicted
9	hsa-miR-523-5p	GNMT	0.365	predicted

Figura 30: Resultados obtenidos de miARN involucrados con el gen GNMT

5.3 Mutaciones-miARN

A partir del análisis realizado con minería de texto se ha obtenido que estos son los genes que estarían mutados, que con más frecuencia aparecen en las publicaciones relacionadas con HCC y miARN.

Gene_symbol	Genes	Freq
1	YTHDF2	14
2	AFP	13
3	DICER1	11
4	GNMT	11
5	CREB1	10
6	ADI1	9
7	CYP2C9	9
8	SIRT7	9
9	UGT2B4	8
10	METTL3	7
11	BMI1	6
12	CTNNB1	6
13	PPP1R15B	6
14	PTEN	6
15	TP53	6

Figura 31: Resultados de genes relacionado con mutaciones más frecuentes

Las publicaciones que se encuentran registradas referidas a mutaciones son 59

Busqueda de mutaciones en Pubmed

Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma descendente

Q Buscar

Mostrar 5 registros

Buscar:

Nro.	Identificador	Revista
1	33240468	1. Comput Struct Biotechnol J. 2020 Oct 21;18:3243-3254. doi:
2	32424223	competing financial interests or personal relationships that could have appeared
3	32333311	3. Dig Dis Sci. 2020 Apr 24. doi: 10.1007/s10620-020-06260-z. [Epub ahead of print]
4	32279994	4. Biochem Biophys Res Commun. 2020 Jun 11;526(4):889-897. doi:
5	31615341	declare that they have no known competing financial interests or personal

Mostrando registros del 1 al 5 de un total de 59 registros

Anterior 1 2 3 4 5 ... 12 Siguiente

Figura 32: Publicaciones sobre mutaciones

Para el gen **AFP** se ha obtenido como resultado

Entrez ID	Gen	Tipo de mutacion	Cambio en el aa.	Nro Chr	Posicion Inicio	Posicion Final	Referencia	Variante	
1	174	AFP	Missense_Mutation	E57Q	4	74303922	74303922	G	C
2	174	AFP	Splice_Site	X595_splice	4	74319616	74319616	T	C

Figura 33: Resultados de mutaciones en el gen AFP

Para el gen DICER1

Entrez ID	Gen	Tipo de mutacion	Cambio en el aa.	Nro Chr	Posicion Inicio	Posicion Final	Referencia	Variante	
1	23405	DICER1	Missense_Mutation	H356L	14	95590842	95590842	T	A
2	23405	DICER1	Splice_Site	X1699_splice	14	95560494	95560494	C	A
3	23405	DICER1	Missense_Mutation	D1875N	14	95556981	95556981	C	T
4	23405	DICER1	Missense_Mutation	E1843D	14	95557445	95557445	T	A
5	23405	DICER1	Missense_Mutation	Y793C	14	95574719	95574719	T	C
6	23405	DICER1	Splice_Region	A752=	14	95577654	95577654	T	A
7	23405	DICER1	Frame_Shift_Del	F1666Lfs*39	14	95562262	95562262	C	-

Figura 34: Resultados de mutaciones en el gen DICER1

Para GNMT

Entrez ID	Gen	Tipo de mutacion	Cambio en el aa.	Nro Chr	Posicion Inicio	Posicion Final	Referencia	Variante	
1	27232	GNMT	Missense_Mutation	A230D	6	42931160	42931160	C	A
2	27232	GNMT	Missense_Mutation	R157Q	6	42930828	42930828	G	A

Figura 35: Resultados de mutaciones en el gen GNMT

6. Conclusiones

Gracias al análisis realizado en minería de texto se ha podido obtener, de diversas publicaciones, cuales son los genes que están involucrados en el hepatocarcinoma. Entre ellos podemos nombrar al gen PTEN, AFP, TP53, FOXO1, STAT3, ZEB1 entre otros. Los genes codifican para proteínas que no actúan solas, sino que forman parte de un "todo" que permite el correcto funcionamiento del hígado y del organismo en general. Por eso es importante poder conocer en qué procesos está involucrado cada uno de estos genes y esto se ha logrado con la información facilitada por GO, a su vez se proporcionó un respaldo bibliográfico, pudiendo acceder a las publicaciones en las que aparecen.

Los miARN regulan la expresión de miles de genes y esto implica, de alguna manera, la regulación de procesos biológicos. Esto es un evento fundamental, ya que no todos los genes se expresan en todas las células, pero su incorrecto funcionamiento puede conducir a la formación de tumores. De este trabajo se han recopilado algunos de los miARN que participan en el hígado y de alguna forma en el HCC. Estos son miR-122, miR-21, miR-7, miR-145, miR-148-a, miR-155. Como regulan la expresión de genes es importante saber

hacia qué genes se dirigen cada uno de los miARN. Esto puede darnos una visión de en qué punto podría producirse el error, que conduce a la formación del HCC.

Por último se han obtenido algunos genes que habrían sido mutados, CDC25A, SIRT7, ADI1, y como es de esperar algunos de ellos coinciden con los anteriores mencionados, ya que uno de los motivos que permite la evolución de un tumor es justamente la mutación de genes que participan en el buen desempeño del órgano. Encontramos, entre otros, MET, PTEN, AFP, TP53, etc. Se ha proporcionado la posibilidad de buscar, a partir de diferentes bases de datos, cuáles son las mutaciones que se producen y dónde ocurren. También se puede recurrir a la publicación proporcionada por Pubmed.

Para optimizar la aplicación HmiR², se podrían implementar gráficos que faciliten visualmente la integración y el estudio del HCC. Por otro lado se podría ampliar los objetos de búsquedas para que dé lugar al análisis de nuevas enfermedades.

La realización de este TFM ha permitido dilucidar que hay menos información del HCC, si se compara con otro tipo de cáncer como de mama, pulmón, colorrectal por ejemplo. A su vez, el estudio de miARN y su participación en la formación de tumores es algo relativamente "nuevo". Esto ha hecho que se dificulte mucho encontrar paquetes idóneos que permitan relacionar el HCC con miARN y por ende hacer un mejor análisis de cómo participan los miARN en este tipo de cáncer, que era lo que se pretendía con este TFM.

Por este motivo creo que este trabajo, que permitió, a partir de la minería de texto, unificar información, que de forma inconexa se encontraba en diferentes publicaciones, puede funcionar de puntapié para continuar con esta investigación pero teniendo presente la visión general del panorama en la que se encuentra este estudio actualmente.

7.Glosario:

GO: Gene Ontology

HCC: Hepatocarcinoma (en inglés, Hepatocellular carcinoma)

KEGG: Enciclopedia de genes y genomas de Kioto

miARN: Ácido microrribonucleicos

TFM: Trabajo Final de Máster

UI: Interfaz de Usuario(en inglés, User Interface)

URL: Localizador de recursos uniforme(en inglés, Uniform Resource Locator)

8.Bibliografía:

- 1) Iñarrairaegui, M.(2016). El hepatocarcinoma. Anales del Sistema Sanitario de Navarra, 39(2), 179-180
- 2) Khemlina, G., Ikeda, S., & Kurzrock, R. (2017). The biology of Hepatocellular carcinoma: implications for genomic and immune therapies. *Molecular cancer*, 16(1), 149.
- 3) Bronte, F., Bronte, G., Fanale, D., Caruso, S., Bronte, E., Bavetta, M. G., Fiorentino, E., Rolfo, C., Bazan, V., Di Marco, V., & Russo, A. (2016). HepatomiRNoma: The proposal of a new network of targets for diagnosis, prognosis and therapy in hepatocellular carcinoma. *Critical reviews in oncology/hematology*, 97, 312–321.
- 4) Brun, R.E., & Senso, J. (2004). Minería Textual. *El profesional de la información*, 13(1), 11-27
- 5) National Library of Medicine. [Internet]. National Center for Biotechnology information(US). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/>
- 6) Rani, J., Shah, A. B., & Ramachandran, S. (2015). pubmed.mineR: an R package with text-mining algorithms to analyse PubMed abstracts. *Journal of biosciences*, 40(4), 671–682.

- 7) Asociación Española contra el Cáncer [internet](ES). Disponible en: <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-higado/anatomia>
- 8) Riedel, R., Pérez-Pérez, A., Carmona-Fernández, A., Jaime, M., Casale, R., Dueñas, J. L., Guadix, P., Sánchez-Margalet, V., Varone, C. L., & Maymó, J. L. (2019). Human amniotic membrane conditioned medium inhibits proliferation and modulates related microRNAs expression in hepatocarcinoma cells. *Scientific reports*, 9(1), 14193. <https://doi.org/10.1038/s41598-019-50648-5>
- 9) Sangro, B. Micro-ARN. Posibles implicaciones terapéuticas en las enfermedades hepáticas. (2011). *GH CONTINUADA*, 10(4), 170-174
- 10) Mou, T., Zhu, D., Wei, X., Li, T., Zheng, D., Pu, J., Guo, Z., & Wu, Z. (2017). Identification and interaction analysis of key genes and microRNAs in hepatocellular carcinoma by bioinformatics analysis. *World journal of surgical oncology*, 15(1), 63. <https://doi.org/10.1186/s12957-017-1127-2>
- 11) Gene Ontology, Unifying Biology [Internet]. Human Genome Research Institute (NHGRI)(US). Disponible en: <http://geneontology.org/>
- 12) The multiMiR R package and database [Internet]. Colorado School of Public Health. University of Colorado (US). Disponible en: <http://multimir.org>
- 13) Rpubs by RStudio. [Internet]. Disponible en: https://rpubs.com/cmlopera/Rday_Shiny

9. Anexo

- La app se encuentra disponible en el servidor de shiny :

<https://aconcagua.shinyapps.io/trabajofinal/>

-Código Shiny:

```
library(multiMiR)
```

```
library(shiny)
```

```
library(shinydashboard)
```

```
library(shinyjs)
```

```
library(shinybusy)
```

```
library(DT)
```

```

library(ngram)

library(easyPubMed)

library(pubmed.mineR)

library(lsa)

library(gdata)

library(stringr)

library(org.Hs.eg.db)

library(rclinicaltrials)

library(GOSim)

library(miRBaseConverter)

library(cgdsr)

choicesgen<- c("PTEN", "AFP", "STAT3", "SIRT7", "GC", "XPO5", "HMGA2", "VEGFA", "FOXO1",
"ZEB1", "PDCD4", "GPC3", "EGFR", "SIRT1", "MET", "DNMT1", "IGF1R", "TET1", "CDC42", "KLF5",
"GRN", "FOXA1", "MACC1", "SOCS1", "TP53", "E2F1", "E2F3", "ROCK1", "SP1", "CTGF", "DICER1",
"HGF", "ROCK2", "BMI1", "CDK1", "FUT8", "METTL3", "PIK3R3", "AKAP12", "AURKA", "CTNNB1",
"EGFL7", "HDAC6", "IGF2", "IGF2BP1", "SMAD2", "SOCS3", "UBE2I", "VDAC1", "ZEB2", "ABC1",
"AKT2", "CCNE1", "FBXO11")

choicesmutation<-c("AFP", "DICER1", "GNMT", "CREB1", "ADI1", "CYP2C9", "SIRT7", "UGT2B4",
"METTL3", "BMI1", "CTNNB1", "PPP1R15B", "PTEN", "TP53", "UGT2B7", "TERT", "XPO5", "AGA",
"CAV1", "CEP55", "PBX3", "ADAMTS4", "CCNE1", "CDC25A", "FGF19", "ZC3H13", "CSMD3",
"CTHRC1", "E2F7", "IGF2BP1", "MET", "NRAS", "PCNA", "RHEB", "RICTOR", "TG", "TSC1", "TSC2",
"VEGFA")

choicesmiarn<-c("mir-122", "mir-21", "mir-221", "mir-125b", "mir-34a", "mir-26a", "mir-148a",
"mir-224", "mir-101", "mir-145", "mir-155", "mir-146a", "mir-375", "mir-7", "mir-195", "mir-214",
"mir-200a", "mir-124", "mir-16", "mir-223", "mir-9", "mir-203", "mir-18a", "mir-29", "mir-139",
"mir-181a", "mir-29a", "mir-126", "mir-141", "mir-106b", "mir-23a", "mir-199a-3p", "mir-338-3p",
"mir-1", "mir-22", "mir-143", "mir-499", "mir-100", "mir-130b", "mir-638", "mir-222", "mir-503",
"mir-125a", "mir-449a")

new_query<-c("(hepatocellular carcinoma[TIAB] OR HCC [all fields] OR hepatocarcinoma
cells[TIAB]) AND (microRNA[TIAB] OR miARN[TIAB] OR microRNAs[TIAB]) AND
(2000:2020/09[dp])")

titulo<-tags$img(src="logo.png", height=110, width= 110)

entrez_id<-select(org.Hs.eg.db,

```

```

        keys = choicesgen,
        columns = c("ENTREZID", "SYMBOL"),
        keytype = "SYMBOL")
ui <- dashboardPage(
  skin="purple",title = "HmiR2",
  dashboardHeader(
    tags$li(
      class = "dropdown"
      ,tags$style(".main-header {max-height: 70px}")
      ,tags$style(".main-header .logo {height: 70px}")
    ),
    title = titulo
  ),
  dashboardSidebar(
    sidebarMenu(
      br(),
      menuItem("Genes", tabName = "Genes", icon = icon("dna"), startExpanded = "TRUE",
        menuSubItem("Publicaciones en Pubmed", tabName = "pubmed1"),
        menuSubItem("GO", tabName = "GO1")),
      menuItem("miARN", tabName = "miARN", icon = icon("vial"), startExpanded = "TRUE",
        menuSubItem("Publicaciones en Pubmed", tabName = "pubmed2"),
        menuSubItem("Datos miR", tabName = "miRData")),
      menuItem("Mutaciones", tabName = "Mutaciones", icon = icon("atom"), startExpanded =
"TRUE",
        menuSubItem("Publicaciones en Pubmed", tabName = "pubmed3"),
        menuSubItem("Info mutacion", tabName = "cancer")),

```

```

    menuItem("Entorno HCC", tabName = "mapa", icon = icon("sitemap"), startExpanded =
"TRUE")
  )
),
dashboardBody(
  useShinyjs(),
  tags$style(type = "text/css",
    ".shiny-output-error{visibility: hidden; }",
    ".shiny-output-error:before{visibility:hidden; }"
  ),
  tabItems(
    tabItem(tabName = "pubmed1",
      h1("Busqueda de Genes en Pubmed"),
      br(),
      h3("Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma
descendiente"),
      br(),
      htmlOutput("genes"),
      br(),
      selectInput("pubmeda", "Selecciona un Gen", choices = choicesgen),
      actionButton(inputId = "pub1", label = "Buscar", icon = icon("search"), block = FALSE),
      add_busy_bar(color = "white"),
      br(),
      hidden(actionButton("nb", label = "Nueva Busqueda", block = FALSE)),
      br(),br(),
      tags$div(style = 'cursor:pointer',DT::dataTableOutput("resultados1")),

```

```

br(),
wellPanel(htmlOutput("panelPublicacionTMP"))
),
tabItem(tabName = "pubmed2",
  h1("Busqueda de miARN en Pubmed"),
  h3("Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma
descendiente"),
  br(),
  htmlOutput("miarn"),
  br(),
  selectInput("pubmedb", "Selecciona un miARN", choices = choicesmiarn),
  actionButton(inputId = "pub2", label = "Buscar", icon = icon("search"), block= FALSE),
  add_busy_bar(color = "white"),
  br(),
  hidden(actionButton("nb1", label = "Nueva Busqueda", block = FALSE )),
  br(),br(),
  tags$div(style= 'cursor:pointer',DT::dataTableOutput("resultados2")),
  br(),
  wellPanel(htmlOutput("panelPublicacionTMPS"))
),
tabItem(tabName = "pubmed3",
  h1("Busqueda de mutaciones en Pubmed"),
  h3("Las publicaciones comprenden un periodo de 20 años(2000-2020), se mostraran de forma
descendiente"),
  br(),
  br(),

```

```

actionButton(inputId = "pub3", label = "Buscar", icon = icon("search"), block= FALSE),
add_busy_bar(color = "white"),
br(),
hidden(actionButton("nb2", label = "Nueva Búsqueda", block = FALSE)),
br(),br(),
tags$div(style= 'cursor:pointer',DT::dataTableOutput("resultados3")),
br(),
wellPanel(htmlOutput("panelPublicacion2"))
),
tabItem(tabName = "GO1",
h2("Ontología de Genes"),
h3("Referencias de Identificadores",style = "text-align: right;"),
fluidPage(
fluidRow(
tags$head(tags$style(HTML(".small-box {height: 250px}")))),
valueBox(
h2("Identificadores Entrez"), width = 5, color = "purple",
div( selectInput(inputId = "radio", label = "Selecciona un gen", choices =
entrez_id$ENTREZID))
),
valueBox(tags$div(img(src = "referencias.png", height = "230", width = "550"), style = "text-align: center;"),color = "aqua", width = 7 ,subtitle = "Referencia")
)
),
br(),
add_busy_bar(color = "white"),

```



```

fluidRow(
  tabsetPanel(
    tabPanel(title="Resultados GO", width = 12,
      tags$div(style= 'cursor:pointer',DT::dataTableOutput("resultadosgo"))
    ),
    tabPanel(title= "Similitud Genes", width = 12,
      h2("Similitud de genes basados en terminos GO", align= "center"),
      h3("Los genes deben ir con el identificador Entrez"),
      div(style = "margin-left: 50px;", div(textInput(inputId = "input1", label = "Escribe el primer
gen"))),
      div(textInput(inputId = "input2", label = "Escribe el segundo gen")
      ),tableOutput("simgen")
    )
  )
),
  tabItem(tabName = "cancer",
    h2("Estudios realizados sobre genes mutados"),
    fluidPage(
      fluidRow(
        tags$head(tags$style(HTML(".small-box {height: 250px}"))),
        valueBox(h2("Al seleccionar un gen se obtiene informacion sobre las mutaciones
encontradas"), width = 6, color = "purple",
          div( selectInput("mutacion", "Selecciona un gen", choices = choicesmutation))
        )
      )
    )
  )
)

```

```

)
),
br(),
add_busy_bar(color = "white"),
fluidRow(
  tabBox(title="Resultados", width = 12,
    tags$div(style= 'cursor:pointer',DT::dataTableOutput("publicacionmut"))
  )
),
tabItem(tabName = "mapa",
  h2("Informacion obtenida de la Enciclopedia de Genes y Genomas de Kioto (KEGG)'),
  tags$div(img(src = "hsa05225.png", height = 890, width = 1060)),
  wellPanel(uiOutput("mapakegg"))
),
tabItem(tabName = "miRData",
  h2("Información correspondiente a miARN"),
  fluidPage(
    fluidRow(
      tags$head(tags$style(HTML(".small-box {height: 250px}"))),
      valueBox(h2("A partir de un gen elegido, se muestra todos los miARN que se dirigen a el"),
width = 6, color = "purple",
      div( selectInput("microgen", "Selecciona un gen", choices = choicesmutation))
    ),
    valueBox(h2("A partir de un miARN, se muestra todos los genes dianas a los que se
dirigen"), width = 6, color = "aqua",

```

```

    div(textInput(inputId = "inputarn", label = "Escribe el miARN", value = "hsa-miR-338-3p"))
  )
)
),
br(),
add_busy_bar(color = "white"),
fluidRow(tabsetPanel(
  tabPanel(title="Resultados ARN", width = 12,
    tags$div(style= 'cursor:pointer',DT::dataTableOutput("publicacionesarn"))),
  tabPanel(title= "Genes Diana", width = 12,
    tags$div(style= 'cursor:pointer',DT::dataTableOutput("publicaciondiana")))
)
)
)
)
)
)
)
)
)
server <- function(input, output, session) {
  #Busqueda Pubmed
  ##Pubmed genes
  PubmedResult<-eventReactive(input$pub1,{
    PM1 <- batch_pubmed_download(
      pubmed_query_string = paste(c(input$pubmeda, " [MH] AND" , new_query),collapse = ""),
      format = "abstract",
      batch_size = 5000,
      dest_file_prefix = "pubmed_"

```

```

)
pmr<-readabs("pubmed_01.txt")
})
output$resultados1<-renderDataTable({
  corpus1<- PubmedResult()
  pmres1<-data.frame(corpus1@PMID, corpus1@Journal)
  colnames(pmres1)<-c("Nro. Identificador", "Revista")
  datatable(pmres1, selection=list(mode="single", selected = c(1)
  ), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))
})
output$panelPublicacionTMP<-renderPrint({
  s = input$resultados1_row_last_clicked
  if (length(s)) {
    corpus1<-readabs("pubmed_01.txt")
    abstract <- corpus1@Abstract[s]
    id <- corpus1@PMID[s]
    link <- a(
      "Ver Publicacion en PubMed"
      ,href= gsub(
        " "
        ,""
      ,paste(
        "https://pubmed.ncbi.nlm.nih.gov/"
        ,corpus1@PMID[s]
      )
    )
  }
})

```

```

    , fixed = TRUE)
    , target="_blank"
  )
  cat(
    paste("<br><p style='color:purple; font-size:24px;'>",corpus1@Journal[s], "</p><br>")
  )
  cat(
    abstract
  )
  cat(
    paste("<br>", link)
  )
}
})
observeEvent(input$pub1,{
  hide("pubmeda")
  hide("pub1")
  hide("panelPublicacion")
  shinyjs::toggle("nb")
})
observeEvent(input$nb,{
  shinyjs::toggle("pubmeda")
  shinyjs::toggle("pub1")
  hide("nb")
  show("panelPublicacion")
})

```

```

output$genes<-renderPrint({

  cat('<h3> Publicaciones sobre <font style="text-trasform:
capitalize";>',input$pubmeda,'</font></h3>')

})

##Pubmed miARN

PubmedResult1<-eventReactive(input$pub2,{

  PM1 <- batch_pubmed_download(

    pubmed_query_string = paste(c(input$pubmedb, " [MH] AND" , new_query),
    collapse = ""),

    format = "abstract",

    batch_size = 5000,

    dest_file_prefix = "pubmed1_"

  )

  pmr1<-readabs("pubmed1_01.txt")

})

output$resultados2<-renderDataTable({

  corpus2<- PubmedResult1()

  pmres2<-data.frame(corpus2@PMID, corpus2@Journal)

  colnames(pmr2)<-c("Nro. Identificador", "Revista")

  datatable(pmr2, selection=list(mode="single", selected = c(1

  ), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))

})

output$panelPublicacionTMPS<-renderPrint({

  s1 = input$resultados2_row_last_clicked

  if (length(s1)) {

```

```

corpus2<-readabs("pubmed1_01.txt")
abstract <- corpus2@Abstract[s1]
id <- corpus2@PMID[s1]
link <- a(
  "Ver publicacion en PubMed"
  ,href= gsub(
    " "
    ,""
    ,paste(
      "https://pubmed.ncbi.nlm.nih.gov/"
      ,corpus2@PMID[s1]
    )
    , fixed = TRUE)
  , target="_blank"
)
cat(
  paste("<br><p style='color:purple; font-size:24px;'>",corpus2@Journal[s1], "</p><br>")
)
cat(
  abstract
)
cat(
  paste("<br>", link)
)
}
})

```

```

observeEvent(input$pub2,{
  hide("pubmedb")
  hide("pub2")
  hide("panelPublicacion1")
  shinyjs::toggle("nb1")
})

observeEvent(input$nb1,{
  shinyjs::toggle("pubmedb")
  shinyjs::toggle("pub2")
  hide("nb1")
  show("panelPublicacion1")
})

output$miarn<-renderPrint({
  cat('<h3> Publicaciones sobre <font style="text-trasform:
capitalize";>',input$pubmedb,</font></h3>')
})

##Pubmed mutaciones

PubmedResult2<-eventReactive(input$pub3,{
  PM2 <- batch_pubmed_download(pubmed_query_string = paste(c("mutation AND" ,
new_query),
                    collapse = ""),
                    format = "abstract",
                    batch_size = 5000,
                    dest_file_prefix = "pubmed2_"
  )

  pmr2<-readabs("pubmed2_01.txt")

```



```

})

output$resultados3<-renderDataTable({
  corpus3<- PubmedResult2()
  pmres3<-data.frame(corpus3@PMID, corpus3@Journal)
  colnames(pmres3)<-c("Nro. Identificador", "Revista")
  datatable(pmres3, selection=list(mode="single", selected = c(1
  ), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))
})

output$panelPublicacion2<-renderPrint({
  s2 = input$resultados3_row_last_clicked
  if (length(s2)) {
    corpus3<-readabs("pubmed2_01.txt")
    abstract <- corpus3@Abstract[s2]
    id <- corpus3@PMID[s2]
    link <- a(
      "Ver publicacion en PubMed"
      ,href= gsub(
        " "
        ,""
        ,paste(
          "https://pubmed.ncbi.nlm.nih.gov/"
          ,corpus3@PMID[s2]
        )
      , fixed = TRUE)
      , target="_blank"

```

```

)
cat(
  paste("<br><p style='color:purple; font-size:24px;'>",corpus3@Journal[s2], "</p><br>")
)
cat(
  abstract
)
cat(
  paste("<br>", link)
)
}
})
#GO
output$resultadosgo<-renderDataTable({input$radio
  x<-as.data.frame(t(getGOInfo(input$radio)))
  colnames(x)<-c("ID GO", "Termino", "Definicion", "IC")
  datatable(x, selection=list(mode="single", selected = c(1)
  ), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))
})
output$referencias<-renderTable({
  y<-select(org.Hs.eg.db,
    keys = choicesgen,
    columns = c("ENTREZID", "SYMBOL"),
    keytype = "SYMBOL")
})

```

```

output$simgen<-renderTable({

  getGeneSim(c(input$input1,input$input2),similarity="OA",similarityTerm="Lin",avg=FALSE,
verbose= FALSE)

})

#KEGG

hiperlink <- a("Ver detalles del mapa en la pagina web de KEGG",href=

  "https://www.genome.jp/dbget-bin/www_bget?pathway+hsa05225", target="_blank")

output$mapakegg<-renderUI({

  tagList("URL link:", hiperlink)

})

#Mutaciones

output$publicacionmut <- renderDataTable({

  mycgds = CGDS("http://www.cbioportal.org/")

  y= getMutationData(mycgds, caseList = "lihc_tcga_all",geneticProfile = "lihc_tcga_mutations",
genes = input$mutacion)

  tabla_mut<-y[,c(-3:-5,-7, -9:-12,-14-15,-18:-22)]

  colnames(tabla_mut)<-c("Entrez ID", "Gen", "Tipo de mutacion", "Cambio en el aa.", "Nro Chr",
"Posicion Inicio", "Posicion Final", "Referencia", "Variante")

  datatable(y, selection=list(mode="single", selected = c(1)

  ), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))

  tabla_mut

})

#miARN

output$publicacionesarn <- renderDataTable({

  example3 <- get_multimir(org = "hsa",

```

```

target = input$microgen,

table = "predicted",

summary = TRUE,

predicted.cutoff = 35,

predicted.cutoff.type = "p",

predicted.site = "all"

)

tablamir<-example3@data[,c(-1,-2,-5:-6)]

colnames(tablamir)<-c("miARN involucrado", "Gen", "Score", "Tipo")

datatable(tablamir, selection=list(mode="single", selected = c(1

), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))

tablamir

})

output$publicaciondiana<-renderDataTable({

genesDiana<- get_multimir(mirna = input$inputarn, summary = TRUE)

tablagen<-genesDiana@data[,c(-2,-5,-6,-8)]

colnames(tablagen)<-c("Base de datos", "miARN involucrado", "Gen", "Tecnica", "ID
Pubmed","Tipo")

datatable(tablagen, selection=list(mode="single", selected = c(1

), options = list(language= list(url = '//cdn.datatables.net/plug-
ins/1.10.11/i18n/Spanish.json'),pageLength = 5))

tablagen

})

}

shinyApp(ui, server)

```

-Codigo R:

```
#Modulo Inicial
```

```
require("easyPubMed")
```

```
require("pubmed.mineR")
```

```
require("lsa")
```

```
##Busqueda en Pubmed
```

```
new_query<-"(hepatocellular carcinoma[TIAB] OR HCC [all fields] OR hepatocarcinoma cells[TIAB])  
AND (microRNA[TIAB] OR miARN[TIAB] OR microRNAs[TIAB]) AND (2000:2020/09[dp])"
```

```
PM1 <- batch_pubmed_download(pubmed_query_string = new_query,  
                             format = "XML",  
                             batch_size = 5000,  
                             dest_file_prefix = "pubmed_"  
                             )
```

```
##Obtencion de corpus primario
```

```
corpus<- xmlreadabs("pubmed_01.txt")
```

```
#Modulo genes-HCC:
```

```
##Obtencion de Corpus secundario(palabras filtradas de interes)
```

```
hcc_gene<-searchabsL(corpus, include = "gene")
```

```
##Tabla.txt con palabras y frecuencias de c/u
```

```
corpus_norm<-xmlword_atomizations(hcc_gene)
```

```
##Tabla.txt con genes y frecuencias de c/u
```

```
corpus_gene<-xmlgene_atomizations(hcc_gene)
```

```
View(corpus_norm)
```

```
##Matriz de los grupos
```

```
terminos<- c("gene", "mir-", "liver", "miarn", "oncogenes", "overexpression", "p53", "apoptosis")
```

```

tdm_gene<-tdm_for_lsa(hcc_gene, terminos)

##Análisis semántico latente

lsa_gene<-lsa(tdm_gene,dims=dimcalc_share())

matrix_gene<-as.textmatrix(lsa_gene)

##Similitud entre terminos

associate_gene<-lapply(terminos, function(x){associate(matrix_gene, x, measure = "cosine",
threshold = "0.7")})

names(associate_gene)<-terminos

cos_sim_calc(matrix_gene)

#GO

#Primero pasar los simbolos de genes a Entrez ID.Se genera un Data Frame

library(GOSim)

library(org.Hs.eg.db)

my.symbols <- c("PTEN", "AFP", "STAT3", "SIRT7", "GC", "XPO5", "HMGA2", "VEGFA", "FOXO1",
"ZEB1", "PDCD4", "GPC3", "EGFR", "SIRT1", "MET", "DNMT1", "IGF1R", "TET1", "CDC42", "KLF5",
"GRN", "FOXA1", "MACC1", "SOCS1", "TP53", "E2F1", "E2F3", "ROCK1", "SP1", "CTGF", "DICER1",
"HGF", "ROCK2", "BMI1", "CDK1", "FUT8", "METTL3", "PIK3R3", "AKAP12", "AURKA", "CTNNB1",
"EGFL7", "HDAC6", "IGF2", "IGF2BP1", "SMAD2", "SOCS3", "UBE2I", "VDAC1", "ZEB2", "ABC1",
"AKT2", "CCNE1", "FBXO11")

entrez_id<-select(org.Hs.eg.db,

  keys = my.symbols,

  columns = c("ENTREZID", "SYMBOL"),

  keytype = "SYMBOL")

View(entrez_id)

getGOInfo(5728)

GOenrichment(5728,cutoff = 0.01,method="elim", allgenes = "174")

#similaridad entre genes

```

```
getGeneSim(c("8614","2852"),similarity="OA",similarityTerm="Lin",avg=FALSE, verbose= FALSE)
```

```
#Modulo miARN-HCC:
```

```
##Obtencion de corpus secundario
```

```
hcc_miarn<-searchabsL(corpus,include = "mir*")
```

```
##Tabla.txt con palabras y frecuencias de c/u
```

```
norm_miarn<-xmlword_atomizations(hcc_miarn)
```

```
##Tabla.txt con genes y frecuencias de c/u
```

```
gene_miarn<-xmlgene_atomizations(hcc_miarn)
```

```
View(norm_miarn)
```

```
##Matriz de los grupos
```

```
terminos<- c("gene", "mir*", "liver", "miarn", "oncogenes", "overexpression", "p53", "apoptosis")
```

```
tdm_miarn<-tdm_for_lsa(hcc_miarn, terminos)
```

```
##Analisis semantico latente
```

```
lsa_miarn<-lsa(tdm_miarn,dims=dimcalc_share())
```

```
matrix_miarn<-as.textmatrix(lsa_miarn)
```

```
##Similitud entre terminos
```

```
associate_miarn<-lapply(terminos, function(x){associate(matrix_miarn, x, measure = "cosine",  
threshold = "0.7")})
```

```
names(associate_miarn)<-terminos
```

```
cos_sim_calc(matrix_miarn)
```

```
library(multiMiR)
```

```
#genes diana. genes validados de un miarn
```

```

genesDiana<- get_multimir(mirna = "hsa-miR-338-3p", summary = TRUE)

tablagen<-genesDiana@data[,c(-2,-5,-6,-8)]

colnames(tablagen)<-c("Base de datos", "miARN involucrado", "Gen", "Tecnica", "ID
Pubmed", "Tipo")

tablagen

#A partir de un gen determinado, buscar que miarn se dirige a este gen

example3 <- get_multimir(org = "hsa",
                          target = "PTEN",
                          table = "predicted",
                          summary = TRUE,
                          predicted.cutoff = 35,
                          predicted.cutoff.type = "p",
                          predicted.site = "all")

tablamir<-example3@data[,c(-1,-2,-5:-6)]

colnames(tablamir)<-c("miARN involucrado", "Gen", "Score", "Tipo")

tablamir

#miarn relacionados a una enfermedad

example2<-get_multimir(disease.drug = 'hepatocellular carcinoma', table = 'disease.drug')

example2@data

#Modulo mutaciones-HCC:

##Obtencion de Corpus secundario

hcc_mutation<-searchabsL(corpus, include = "mutation")

##Tabla.txt con palabras y frecuencias de c/u

norm_mutation<-xmlword_atomizations(hcc_mutation)

View(norm_mutation)

##Tabla.txt con genes y frecuencias de c/u

```



```

gene_mutation<-xmlgene_atomizations(hcc_mutation)

View(gene_mutation)

##Matriz de los grupos

terminos<- c("gene", "mir-", "liver", "miarn", "oncogenes","overexpression", "p53", "apoptosis")

terminos<- c("mir-122", "mir-433", "mir-224", "mir-124", "mir-145", "mir-199a", "mir-33a3p",
"mir-9", "mir-98-5p", "mir-18a", "mir-128-3p", "mir-125b", "mir-34a", "mir-410-3p")

tdm_mutation<-tdm_for_lsa(hcc_mutation, terminos)

##Análisis semántico latente

lsa_mutation<-lsa(tdm_mutation,dims=dimcalc_share())

matrix_mutation<-as.textmatrix(lsa_mutation)

##Similitud entre terminos

associate_mutation<-lapply(terminos, function(x){associate(matrix_mutation, x, measure =
"cosine", threshold = "0.7")})

names(associate_mutation)<-terminos

cos_sim_calc(matrix_mutation)

library(cgdsr)

#Estudios sobre HCC

choicesmutation<-c(YTHDF2,AFP,DICER1, GNMT, CREB1, ADI1, CYP2C9, SIRT7, UGT2B4, METTL3,
BMI1, CTNNB1, PPP1R15B, PTEN, TP53, UGT2B7, TERT, XPO5, AGA, CAV1, CEP55, PBX3, ADAMTS4,
CCNE1, CDC25A, FGF19, ZC3H13, CSMD3, CTHRC1, E2F7, IGF2BP1, MET, NRAS, PCNA, RHEB,
RICTOR, TG, TSC1, TSC2, VEGFA)

mycgds = CGDS("http://www.cbioportal.org/")

x= getMutationData(mycgds, caseList = "lihc_tcga_all",geneticProfile = "lihc_tcga_mutations",
genes = 'PTEN')

tabla_mut<-x[,c(-3:-5,-7, -9:-12,-14-15,-18:-22)]

```

```
colnames(tabla_mut)<-c("Entrez ID", "Gen", "Tipo de mutacion", "Cambio en el aa.", "Nro Chr",  
"Posicion Inicio", "Posicion Final", "Referencia", "Variante")  
  
df.pri<-getProfileData(mycgds, genes= "PTEN", geneticProfiles = "lihc_tcga_rna_median_zscores",  
caseList = )  
  
getCaseLists(mycgds,'lihc_tcga')[,c(1:2)]  
  
  
plot(mycgds, "lihc_tcga_pan_can_atlas_2018", c("PTEN"), "lihc_tcga_mrna" ,"lihc_tcga_all")  
plot(mycgds, "gbm_tcga", c("PTEN","TP53"), "lihc_tcga_rna_seq_v2_mrna" ,"lihc_tcga_all")  
  
df = getProfileData(mycgds, "NF1", c("gbm_tcga_gistic","gbm_tcga_mrna"), "gbm_tcga_all")
```