



Búsqueda de fármacos para dianas terapéuticas en leucemia

Juan Luis Mendoza Faleiro

Máster universitario en Bioinformática y bioestadística
Bioinformática farmacéutica

Melchor Sánchez Martínez
María Jesús Marco Galindo

30 de junio de 2016



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:	<i>Búsqueda de fármacos para dianas terapéuticas en leucemia.</i>
Nombre del autor:	<i>Juan Luis Mendoza Faleiro</i>
Nombre del consultor/a:	<i>Melchor Sánchez Martínez</i>
Nombre del PRA:	<i>María Jesús Marco Galindo</i>
Fecha de entrega (mm/aaaa):	06/2016
Titulación:	<i>Máster universitario en Bioinformática y bioestadística</i>
Área del Trabajo Final:	<i>Bioinformática farmacéutica</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>Src, leukemia, therapeutic target</i>
Resumen del Trabajo:	
<p>Entre los casos que se registran anualmente de cáncer la leucemia tiene un papel importante. El cáncer es la segunda mayor causa de muerte en el mundo. Esto hace que la identificación de nuevos fármacos con los que tratar la enfermedad sea esencial.</p> <p>En la búsqueda y diseño de fármacos, cuanto más específicos y mayor afinidad exista con la sustancia problema, mejor pronóstico tendrá el paciente. Para ello, lo primero es identificar una diana terapéutica.</p> <p>En el presente trabajo se ha realizado una búsqueda de los principales genes involucrados en leucemia mediante una búsqueda en diversas bases de datos públicas. Se ha creado una red de interacción génica de la que se han podido extraer lo que se presuponen como los nueve genes centrales.</p> <p>Con estos genes como dianas se ha realizado una serie de acoplamientos moleculares con una serie de fármacos. Como resultado se obtuvo que el mejor acoplamiento era entre los fármacos N⁶- benzyl-ADP y RU78783 con la oncoproteína del gen Src. Estos fármacos, que se encuentran en fase experimental, interaccionan con los dominios SH2 y SH3 de la oncoproteína, la cual está constitutivamente activada en un alto porcentaje de células cancerosas. Esta interacción, se cree que podría derivar en una inhibición a nivel catalítico y a nivel estructural de la proteína diana, ejerciendo así una posible acción antitumoral.</p>	

Abstract:

Leukemia constitutes a significant percentage of total cancer cases reported annually. Since cancer is the second leading cause of death worldwide, finding new drugs for its treatment is essential.

Design or search for drugs of high specificity and affinity for the target protein results in an improved prognosis and life expectancy of the patient. However, to find a good drug, it is first necessary to identify possible therapeutic targets.

In the present work, key genes involved in leukemia were selected by searching various public databases. Subsequently, a gene interaction network was created in order to obtain those genes (nine in total) that have a central role in the development and course of the disease.

Assuming these genes as potential therapeutic targets, molecular docking has been made to test a number of drugs. As a result, the best molecular coupling obtained was the one between N⁶-benzyl-ADP and RU78783 with Src oncoprotein. Both drugs, which are in experimental phase, interact with the SH2 and SH3 domains of Src, which appears constitutively activated in a high percentage of cancer cells. These interactions would result both in a catalytic and/or conformational silencing of the target protein, thereby exerting a possible antitumor activity.

Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	2
1.3 Enfoque y método seguido.....	2
1.4 Planificación del Trabajo.....	3
1.5 Breve resumen de productos obtenidos.....	4
1.6 Breve descripción de los otros capítulos de la memoria.....	4
2. 1 Resultados.....	5
2. 2 Discusión.....	11
3. Conclusiones.....	13
4. Glosario.....	14
5. Bibliografía.....	15

Lista de figuras

Figura 1: Diagrama de Gantt con la temporalización del trabajo.....	3
Figura 2: Primera parte de tres de “Script.py”	6
Figura 3: Segunda parte de tres de “Script.py”	7
Figura 4: Tercera parte de tres de “Script.py”	8
Figura 5: Red de interaccion génica completa.	8
Figura 6: Representación de los genes organizados en clústeres	9
Figura 7: Tabla resumen de los resultados obtenidos.....	10
Figura 8: Representación del acoplamiento molecular con 2BDF.....	11

1. Introducción

1.1 Contexto y justificación del Trabajo

El cáncer es una de las enfermedades con mayor mortalidad. En 2012 se estimó que produjo 8.2 millones de muertes y se dieron 14.1 millones de nuevos casos. De estos, la leucemia estuvo involucrada en un 2.5% y en un 3.5% de las muertes [1]. Esto hace que se promueva el interés en la búsqueda de tratamientos para la enfermedad.

En la actualidad, la farmacoterapia es la base clínica del tratamiento de enfermedades y aun así la mayoría de los fármacos son efectivos solo entre un 30% y un 60% de los pacientes [2]. En la terapia del cáncer el principal problema es la quiescencia en las células madre, que da como resultado resistencia a los fármacos [3].

Ante esta resistencia resulta esencial la búsqueda de nuevos fármacos y de nuevas dianas terapéuticas que permitan tratar las enfermedades. Un acercamiento básico para la identificación de dianas son los polimorfismos genéticos que desde hace tiempo se emplean en el descubrimiento de nuevos objetivos para fármacos [4]. En enfermedades como la leucemia donde el componente genético es la base de la enfermedad [5] la búsqueda de genes que dan lugar a la enfermedad es un punto clave para la obtención de un tratamiento.

Actualmente, con el avance de la tecnología y las técnicas bioinformáticas, es posible conocer y trabajar de forma masiva con genes que provocan enfermedades [6] y acceder a ellos desde diferentes bases de datos como pueden ser DisGeNET [7] o ClinVar [8].

En el área de la genómica computacional, técnicas de aprendizaje automático y de minería de datos permiten hacer uso de grandes conjuntos de genes en la búsqueda de patrones de expresión que se identifiquen con la enfermedad [9] o con relaciones entre ellos.

Es por tanto necesario seguir desarrollando tratamientos para la leucemia y una de las posibilidades es mediante la búsqueda de dianas terapéuticas entre los genes que la provocan. En este estudio se han extraído los principales genes involucrados en leucemia. Haciendo uso de sus relaciones y de diferentes *softwares* se han obtenido posibles dianas terapéuticas junto con algunos fármacos que podrían ser empleados en su tratamiento.

1.2 Objetivos del Trabajo

2.1. Objetivos generales:

1. Modelar una red de interacción de los genes implicados en los diferentes tipos de leucemias.
2. Obtener posibles dianas terapéuticas para la leucemia.

2.2. Objetivos específicos:

1. Modelar una red de interacción de los genes implicados en los diferentes tipos de leucemias.
 - 1.1 Crear un *script* que permita el cruce de bases de datos públicas.
 - 1.2 Obtener genes relevantes implicados en la leucemia.
 - 1.3 Integrar la información mediante la herramienta *Cytoscape*.
2. Obtener posibles dianas terapéuticas para la leucemia.
 - 2.1 Seleccionar los nodos comunes en diferentes leucemias.
 - 2.2 Identificar posibles dianas terapéuticas.
 - 2.3 Obtener proteínas con alto valor de “druggability”.
 - 2.4 Realizar “*reprofiling*” obteniendo moléculas con poder terapéutico.

1.3 Enfoque y método seguido

Al inicio del estudio se presentaron dos grandes estrategias posibles. La primera consistía en una mayor inmersión en el campo de la informática. Esta sería realizar una herramienta de software mediante técnicas de “*machine learning*” (aprendizaje automático) y “*data-mining*” (minería de datos). Con esta herramienta propia, se organizaría la información relativa a leucemia obtenida de bases de datos. El otro enfoque es la organización de la información obtenida de bases de datos mediante software ya existente ajustando los parámetros lo máximo posible a las necesidades de este estudio.

La elección de la segunda estrategia, con software de terceros, residió en el factor tiempo, el cual era muy limitante en el estudio. Si bien la primera aproximación podría haber dado un resultado más específico y

personal a lo que se buscaba. No se podría haber llegado más allá de la realización del software.

Quedando lejos del objetivo del trabajo, la obtención de dianas terapéuticas, el mejor enfoque fue servirse de software ya desarrollado.

1.4 Planificación del Trabajo

1.1. Realizar una revisión bibliográfica amplia.

1.2.1. Obtener genes relevantes mediante bases de datos públicas.

1.2.2. Construir una base de datos con la información.

1.3. Integrar la información para la creación de la red de interacción de los genes.

2.1.1. Analizar la red de interacción de los genes.

2.1.2. Seleccionar los nodos principales.

2.2. Comprobar el potencial como dianas terapéuticas.

2.3. Buscar moléculas con un posible poder terapéutico.

3. Creación de la memoria del trabajo.

4. Preparación de la defensa del trabajo.

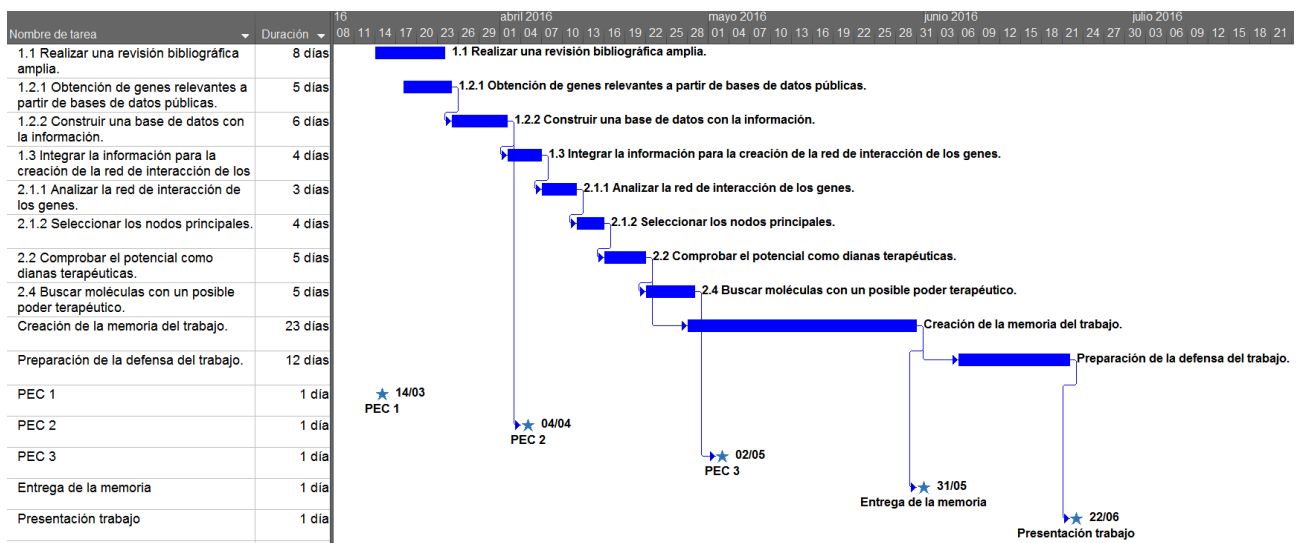


Figura 1: Diagrama de Gantt con la temporalización del trabajo.

1.5 Breve resumen de productos obtenidos

Se obtuvo un *script* escrito en Python mediante el cual se obtiene la filtración de cuatro bases de datos dando como resultado un listado de genes involucrados en leucemia. Los parámetros que recibe son: las bases de datos, bajo un nombre determinado. Y devuelve el archivo, "*genes.txt*", con una columna en la que cada registro es un gen.

Se ha obtenido los nueve genes centrales implicados en leucemia.

Por otro lado, se han encontrado dos fármacos con posible poder terapéutico contra la oncoproteína codificada por el gen SRC, identificada como diana terapéutica.

1.6 Breve descripción de los otros capítulos de la memoria

Los siguientes capítulos de la memoria serán dos. El primero estará constituido por los resultados. En él, se expondrán en orden los datos de interés obtenidos a cada paso de este estudio.

Se mostrará el desarrollo detallado del "script" en el que haciendo uso de la librería "pandas" de Python, se realiza una búsqueda exhaustiva de genes relacionados con la leucemia cruzando las diferentes bases de datos.

Se analizarán los genes y se explicarán los resultados obtenidos mediante el procesamiento de los datos con *Cytoscape*. También se mostrará la red de interacciones de los genes implicados.

Posteriormente se llevará a cabo un análisis de la red, seleccionando los nodos principales y comprobando que genes son posibles dianas terapéuticas.

Con las dianas ya seleccionadas, se medirá la *druggability* para la proteína que codifica. En aquellas que presenten mayor medida de *druggability* se realizará un *reprofiling* para buscar moléculas con un potencial poder terapéutico.

En el segundo capítulo se realizará una interpretación de los resultados aportando información de posible utilidad en el tratamiento de la leucemia.

2. 1 Resultados

Construir una base de datos con la información.

Se desarrolló un *script* escrito en Python 2.7. Haciendo uso de la librería Pandas se filtraron cuatro bases de datos en los campos que describían la implicación del gen. Para ello se usaron palabras clave relativas a leucemia.

Las bases de datos elegidas fueron: DisGeNET, BioMuta, UniprotKB y COSMIC. Se seleccionaron por ser bases de datos de acceso público, así como por poseer registros de nombres de genes y descripciones de enfermedades en los que están implicados, aunque cada una aportó diferente información.

La base de datos DisGeNET, incluye a su vez múltiples bases de datos que relacionan genes y enfermedades [7, 10]. BioMuta proporciona la asociación de mutaciones puntuales en regiones del genoma que ocasionan o están relacionadas con enfermedades [11, 12]. *Universal Protein Resource Knowledgebase* (UniprotKb) es una potente base de datos que incluye proteínas a partir de secuencias del *International Nucleotide Sequence Database Collaboration* (INSDC) así como toda la información facilitada por este organismo [13, 14]. COSMIC es una base de datos que almacena información de las mutaciones somáticas que están relacionadas con cáncer en humanos [15, 16].

Dichas características las hicieron idóneas para el estudio.

Mediante el filtrado se obtuvieron 15 747 genes que estaban involucrados en procesos asociados a leucemia. Donde un 73,93% se obtuvo de la base de datos de COSMIC, un 23,16% de DisGeNET, un 2,83% de UniprotKb y un 0.08% de BioMuta.

Los genes obtenidos en cada base de datos eran únicos dentro de cada una de ellas, pero la redundancia entre las diferentes bases de datos supuso un 42,41%. Este gran número de duplicados, sumado a la dificultad de procesamiento de un volumen tan grande de genes e interacciones, constituyo la razón para que no se buscara en más bases de datos.

En el *script* desarrollado, se inicia con dos funciones, una de búsqueda y otra para separar términos de listas (Figura 2) que se emplearan a lo largo del proceso.

```
import pandas as pd

def buscar_termino(datos, columna):
```

```

"""Busca los términos asociados a Leucemia en una columna de
una base de datos dada"""
terms= ["leukemia", "Leukemia", "leukaemia", "Leukaemia", "CLL",
"ALL", "AML", "CML"]

#Se crea un DataFrame con los valores de las columnas que vamos
a extraer.
db_1= pd.DataFrame(columns=list(datos.columns.values))

#Se extraen los datos filtrados y se incorporan al DataFrame
creado.
for i in terms:
    db_filtrada = datos[datos[columna].str.contains(i)]
    db_1= pd.concat([db_1 , db_filtrada])

return db_1

def separador(lista):
"""Desempaqueta listas dentro de listas"""
genes_sep=[]
for i in lista:
    n = i.split()
    for j in n:
        genes_sep.append(j)
return genes_sep

```

Figura 2: **Primera parte de tres de "Script.py"**. Código comentado de las funciones que se emplearán a lo largo del script en repetidas ocasiones, para cada base de datos que lo precise.

Recibe como argumentos las cuatro bases de datos, cada una con su nombre y formato de origen. Posteriormente hace uso de la función de búsqueda en la que se filtra cada base de datos usando las palabras clave designadas.

Por último, se preparan los datos para el archivo de salida, nuevamente usando una función se separan los nombres de los genes de manera individual y se empaquetan en una lista (Figura 3). Se itera a través de esta última y se escriben en un archivo de texto de salida (*genes.txt*) que será la base de datos de la que partirá el estudio (Figura 4).

```

#Base de datos de DisGenet.
#Carga de la base de datos
db_disgenet = pd.read_table("all_gene_disease_associations.txt",
header=0)

#se obtienen los registros filtrados
db_d1=buscar_termino(db_disgenet,"diseaseName")
#Se crea una lista con los nombres de los genes extraídos de los
registros.
lista_d=list(db_d1['geneSymbol'].unique()) # 3647 genes
genes_d=separador(lista_d)

#Base de datos de BioMuta.
#Carga de las columnas que hacen referencia al nombre del gen y
enfermedad.
db_biomuta = pd.read_csv("BioMuta3.csv",header=None, sep= "\t",
usecols=[2,13])

```

```

#se obtienen los registros filtrados
db_b1=buscar_termino(db_biomuta,13)
#Se crea una lista con los nombres de los genes extraídos de los
registros.
genes_b= list(db_b1[2]) #14 genes

#Base de datos de Uniprot.
#Carga de las columnas que hacen referencia al nombre del gen y
enfermedad.
db_uniprot = pd.read_table("uniprotreviewedHomosapiens.tab",
                          header=0, usecols=[4,5])
#db_uniprot1 = db_uniprot[[4,5]]
uniprot_nonan=db_uniprot.dropna()
#se obtienen los registros filtrados
db_ul=buscar_termino(uniprot_nonan, "Involvement in disease")
#Se crea una lista con los nombres de los genes extraídos de los
registros.
lista_u = list(db_ul["Gene names"])
genes_u=separador(lista_u)# 445 genes

#Base de datos de COSMIC.
#Carga de las columnas que hacen referencia al nombre del gen y
enfermedad.
db_cosmic = pd.read_table("CosmicMutantExport.tsv", header=0,
                          sep= "\t", usecols=[0,12,13,14])

db_c1=buscar_termino(db_cosmic,"Histology subtype 1")
db_c2=buscar_termino(db_cosmic,"Histology subtype 2")
db_c3=buscar_termino(db_cosmic,"Histology subtype 3")

genes_c= list(db_c1["Gene name"].unique())
genes_c+= list(db_c2["Gene name"].unique())
genes_c+= list(db_c3["Gene name"].unique())

#Se establece un contador que haga de iterador a través de la
lista.
c=0
for i in genes_c:
    if '_ENST' in i:
        #Se desempaqueta tomando solo el nombre del gen.
        (genes_c[c],j)=(i.split("_ENST"))
        c+=1

#Se juntan todos los nombres de los genes y se eliminan los
duplicados.
genes= list(set(genes_u+genes_b+genes_d+genes_c))

```

Figura 3: **Segunda parte de tres de “Script.py”**. Código comentado del cuerpo del script, donde se extrae información de cuatro bases de datos y se almacena en una única lista.

La problemática obtenida durante la creación del script obligó a tomar algunas medidas que en él se muestran. El principal problema radicó en la búsqueda y obtención de genes relevantes. Debido al tamaño de las bases de datos, este era excesivamente grande, haciendo que el *hardware* no pudiese procesar correctamente todas ellas. Para ello,

como se puede observar en el *Script*, se procedió a disminuir el volumen de datos que se cargaban en Python. Se identificaron las columnas que contenían la información de interés en cada base de datos y se importaron exclusivamente esos campos. De modo que todos los registros fuesen más cortos y se agilizará el procesamiento de los datos, solucionando así el problema.

```
f = open("genes.txt", "w")

for i in genes:
    f.write(i + "\n")

f.close()
```

Figura 4: Tercera parte de tres de “Script.py”. Código con el final del script para la generación de un archivo de texto que sirva de base de datos con los genes involucrados en leucemia.

Otro problema fue el gran volumen de duplicados que arroja la base de datos COSMIC. Se debía a genes que aparecen repetidas veces con diferentes mutaciones que acaban en el mismo resultado. En cada mutación se les añadía diferentes sufijos. Uno de ellos era el sufijo “_ENSTxxxxx” que hace referencia al código de acceso que posee el transcrito en *ensembl* [17]. Este se pudo eliminar y así reducir duplicados.

El resultado fue un listado de 9 069 genes que respondieron al filtrado.

Puesto que el *script* es sensible a mayúsculas, los términos también fueron introducidos en minúsculas.

Creación y análisis de la red de interacción de los genes.

Para la gestión de la red de interacción génica se empleó el software de código abierto *Cytoscape* en su versión de escritorio 3.3.0 [18].

Con la aplicación interna *ReactomeFI* [19] se incorporó la lista de genes. Esta permitió adjuntar un total de 147 660 interacciones entre ellos y con ello crear la primera red de interacción génica (Figura 5).

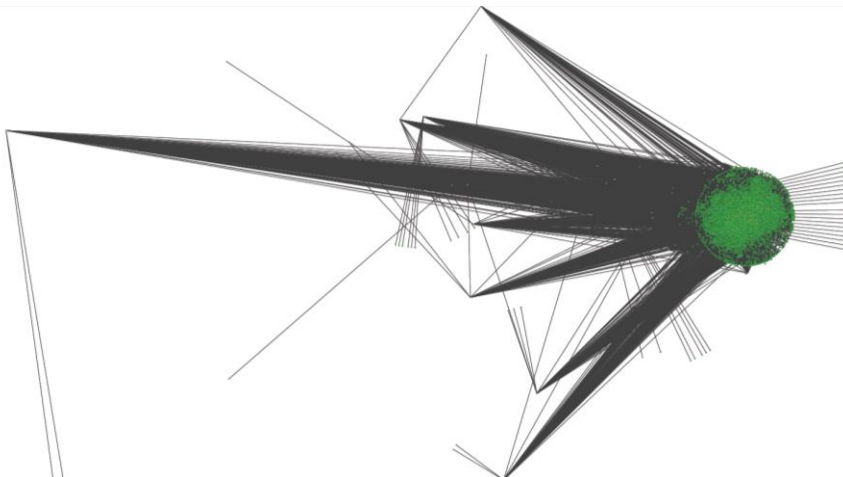


Figura 5: Red de interacción génica completa. En verde se marcan los nodos o genes de la red junto con todas sus conexiones representadas en gris.

Para facilitar la computación de los datos y hacerlos accesibles, se realizó una filtración de los genes en función de su relevancia en la red. Dicho cribado se estableció en el mínimo filtro que permitiera procesar los datos resultantes, este fue que cada gen seleccionado tuviese un mínimo de 1 500 conexiones con otros nodos (genes) en un rango máximo de dos distancias.

Al disminuir el volumen de nodos se pudo llevar a cabo su organización en *clústeres*, nuevamente usando *ReactomeFl*. Se obtuvieron 9 agrupaciones (Figura 6). Se seleccionó cada uno de los *clústeres* y se trabajó independientemente. Con ello se mantuvo la representatividad de todos los sectores de la red durante el resto del estudio.

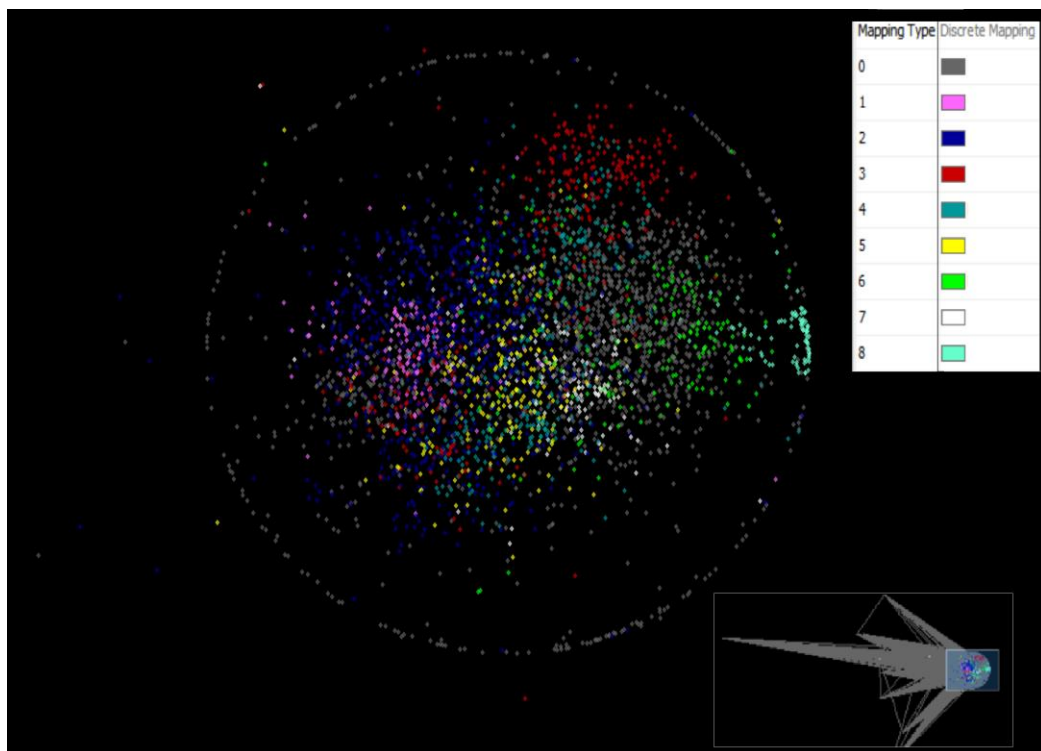


Figura 6: **Representación de los genes organizados en clústeres.** Imagen ampliada de la red original (abajo a la izquierda) en la que solo se muestran los genes sin sus interacciones, coloreados en función del clúster al que pertenecen (clústeres numerados del 0 al 8).

Finalmente, mediante otro análisis de nodos cercanos se realizó nuevamente una filtración, esta vez más exigente. Se limitó el resultado, dejando solamente el gen con más interacciones de cada *clúster* y, por tanto, los nodos principales de cada sección de la red. Se obtuvo un listado de nueve genes que según este estudio serían los principales genes involucrados en leucemia. Estos fueron: UBC, SRC, MAPK1, EED, CDK1, IKBKG, UBB, PSMC2 y NANOG.

Comprobación de dianas terapéuticas y búsqueda de fármacos.

Se empleó la base de datos UniprotKb para identificar el número de acceso (*accession*) de las principales proteínas codificadas por los genes objetivo. De modo que se pudo realizar una comprobación de la

“*druggability*” de cada proteína, empleando la herramienta web de *The European Bioinformatic Institute* (EBI), *DrugEBllity* [20].

Se obtuvieron los dominios más “drugables” de las proteínas encontradas mediante la selección de aquellos con *scores* más elevados según *ensembl*. Este es calculado realizando una ponderación de los valores obtenidos para la unión de 25 ligandos fisicoquímicos conocidos con los sitios de unión generados por 11 modelos de predicción (Support Vector Machine, perceptrón multicapa y arboles decisionales).

Para el gen SRC se seleccionaron dos dominios, uno con el mayor valor de “drugabilidad” (2BDF) y otro con menor valor (2H8H) para emplearlo como grupo control.

Los dominios diana obtenidos se usaron para llevar a cabo pruebas de *docking* con los principales fármacos empleados en genes de leucemia.

Los fármacos fueron obtenidos de la base de datos DGldb (*drug-gene interaction database*). Esta incorpora interacciones de genes con las principales bases de datos de fármacos existentes [21].

Se obtuvo una lista de 77 fármacos. Se seleccionaron aquellos con el mejor valor para el *score* DGldb el cual se basa en el número de fuentes que se han referido a el compuesto como tal fármaco.

Para el acoplamiento molecular se hizo uso del software en línea SwissDock, proporcionado por *Swiss Institute of Bioinformatics* [22]. Las medidas que se obtuvieron están basadas en la energía necesaria para la formación del enlace (Figura 7).

Accession	Gen	Dominio	Fármaco	FullFitness (Kcal/mol)	DGldb score
P12931	SRC	2BDF	DASATINIB	-3123.08	11
			BOSUTINIB	-3033.44	4
			PHENYLPHOSPHATE	--	4
			RU78783	-3339.20	4
			KX2-391	-3068.17	3
			N6-BENZYL ADENOSINE-5'-DIPHOSPHATE	-3349.22	3
			PONATINIB	-3037.59	3
			SARACATINIB	-3068.54	3
		2H8H	DASATINIB	-2407.88	11
		RU78783	-2619.57	4	
P06493	CDK1	4YC6	ALSTERPAULLONE	-2375.99	3
			SU9516	--	3
			HYMENIALDISINE	-2448.90	2
			OLOMOUCINE	-2433.70	2
P28482	MAPK1	1WZY	PHOSPHONOTHREONINE	-2130.75	4

Figura 7: **Tabla resumen de los resultados obtenidos.** Se presentan los datos del acoplamiento molecular para cada fármaco con sus respectivas proteínas y genes, así como el *score* que obtienen en la base de datos DGldb. Los valores “--” muestran acoplamientos fallidos.

Los resultados más favorables se dieron para los fármacos: N⁶-benzyl ADP y RU78783 (Figura 8).

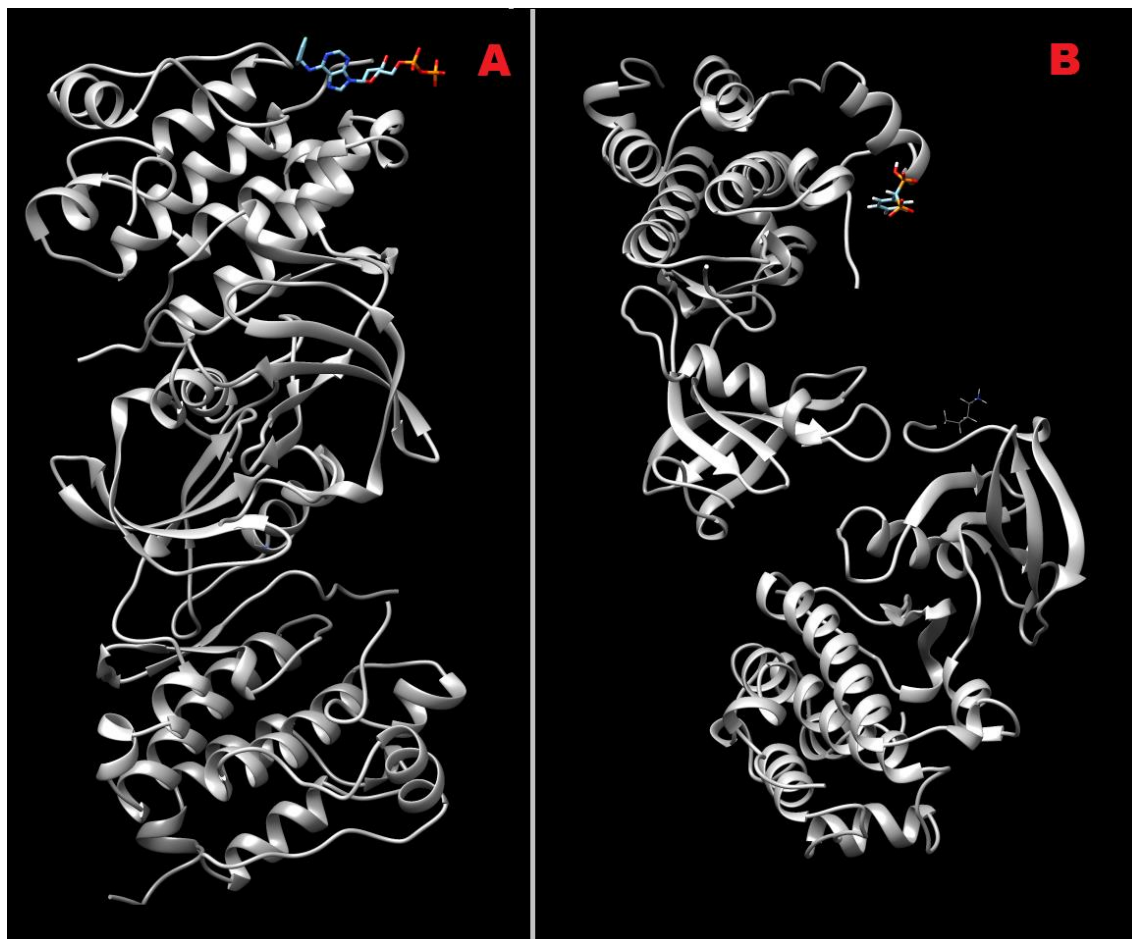


Figura 8: Representación del acoplamiento molecular con 2BDF. En la figura A observamos el docking realizado con N⁶-benzyl ADP y en la figura B con RU78783. Imágenes tomadas con UCSF Chimera [23]

2. 2 Discusión

El volumen de genes obtenidos en primera instancia, muestra la dificultad de trabajar con diferentes bases de datos, las cuales no están estandarizadas y sus diferentes nomenclaturas conllevan al uso reiterado de los mismos casos de estudio. Cuando los genes se llevan a una estandarización, identificando cada uno con su nomenclatura común y eliminando los duplicados, la redundancia que se alcanza no llega a ser alarmante.

La disposición de la red de interacción génica sugiere que los genes poseen conexiones homogéneamente distribuidas exceptuando algún nodo que se queda fuera del área de mayor concentración de genes. Esto podría indicar que, cómo se pretendía en esta primera parte del estudio, estén todos relacionados por algo en común, la leucemia. A su vez la obtención de nueve clústeres indica que existen nueve grupos

relacionales que están involucradas en procesos que dan lugar a leucemia.

De estos nueve genes, *SRC* es el que más fármacos arroja en la búsqueda, de ahí su gran representatividad en las figuras y en las tablas. En la figura 7 se puede observar como ante los mismos fármacos la energía necesaria para llevar a cabo el acoplamiento es mayor en el dominio menos *drugable*. Por tanto, con el dominio proteico control 2H8H (con un *score* de *drugabilidad* menor) se puede afirmar que son más *drugables* aquellos dominios con un *score* más elevado.

Por otro lado, los fármacos con un mejor acoplamiento son N⁶-benzyl adenosine-5'-diphosphate y RU78783, ambos para el gen *SRC*.

La proteína codificada por *SRC* presenta los dominios SH2 y SH3 que están involucrados en el reconocimiento del sustrato y la regulación de la actividad catalítica de la proteína. El dominio SH2 reconoce tanto a la tirosina fosforilada como a los aminoácidos que la flanquean en su extremo C-terminal [24].

Se ha visto que normalmente, *Src* se encuentra en su forma inactiva. El mantenimiento de esta conformación inactiva es debido, en gran parte, por la interacción del dominio SH2 con la fosfo-tirosina 527. En el caso de la oncoproteína, el residuo de tirosina 527 ha sido sustituido por otro residuo que no puede interactuar con el dominio SH2 y, por tanto, la proteína se encuentra activada constitutivamente [24].

Dado que la activación de *Src* promueve la supervivencia, angiogénesis, proliferación e invasión, ésta desempeña un papel fundamental en el crecimiento de los tumores.

Aunque su acción farmacológica no es del todo conocida, el potencial factor terapéutico que podría residir en el fármaco N⁶-benzyl ADP, es su carácter de análogo de ADP.

Se ha visto que la *Src* mutada, incorpora con rapidez análogos del ATP, como el N⁶-benzyl ATP [25], si la quinasa quedase bloqueada con el fármaco, no podría incorporar ATP y su actividad catalítica se vería truncada. Aunque otros estudios [26] indican que la unión del análogo del ADP a la proteína *Src* mutada, no interviene en su función.

Por otro lado, RU78783, es un mimético de la fosfo-tirosina el cual podría simular la fosfo-tirosina 527 sustituida del dominio SH2 de la proteína *Src* [27]. Debido a esto, la quinasa *Src* mutada, presente en las células cancerosas, podría adoptar una conformación cerrada manteniéndose en su forma inactiva.

3. Conclusiones

- A la luz de los resultados es de fácil interpretación la necesidad de grandes cantidades de datos y lo esencial que es la complementariedad de formatos entre diferentes bases de datos.
- Creo que todos los objetivos han sido cubiertos, se han encontrado dianas terapéuticas y se han aportado unos posibles tratamientos identificando fármacos con una acción potencial sobre ellas. Solo se ha podido indagar sobre aquellos con mejor resultado. Complicando la realización de *reprofiling* de los fármacos. Hubiese sido interesante incorporar los usos de los demás fármacos y comprobar aquellos destinados a otras enfermedades.
- La planificación se ha seguido acorde a lo planteado al principio del proyecto, con algunas mínimas variaciones. Se tuvieron que realizar de nuevo los filtrados de los genes y por ello hubo que retroceder, pero la temporalización no sufrió modificaciones.
- La metodología elegida ha sido adecuada, se planteó usar otra quizás más exacta, con programas más complejos. Pero el factor tiempo nuevamente no permitía demoras, por lo que se decidió realizar como aquí se ha mostrado. En la planificación se planteó una revisión bibliográfica únicamente como tarea inicial. Debería haberse planteado como una tarea complementaria a las demás y de misma duración que el proyecto en sí, ya que se ha llevado a cabo en cada tarea.
- Una de las líneas de trabajo que no se han podido abarcar y que resulta muy interesante es profundizar de manera individual en cada uno de los nueve clústeres. Analizar las rutas metabólicas en las que participan. Intentando averiguar si el papel que poseen los nueve genes principales es efectivamente, el de genes centrales en cada una de las posibles rutas. Así como las relaciones entre cada clúster.

4. Glosario

Clustering: agrupación de los datos en formaciones de objetos similares.

Docking: procedimiento por el que se obtiene la conformación más estable de un complejo que resulta de la unión de dos o más moléculas.

Druggability: ensayos para comprobar la afinidad de un compuesto diana frente a un fármaco como ligando.

Nodo: gen central de la red de interacción y del que se presupone derivan otros.

Python: es un lenguaje de programación con la filosofía de favorecer una sintaxis clara y un código legible.

Reprofiling: consiste en emplear un fármaco creado originalmente para una enfermedad, para tratar otra en la que también se obtienen resultados favorables.

Script: programa simple.

5. Bibliografía

1. Ferlay, Jacques et al. 2015. "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012." *International journal of cancer* 136(5): E359–86. <http://www.ncbi.nlm.nih.gov/pubmed/25220842> (Marzo , 2016).
2. Sosa, María Soledad, Paloma Bragado, and Julio A Aguirre-Ghiso. 2014. "Mechanisms of Disseminated Cancer Cell Dormancy: An Awakening Field." *Nature reviews. Cancer* 14(9): 611–22. <http://www.ncbi.nlm.nih.gov/pubmed/25118602> (Marzo , 2016).
3. Spear, B B, M Heath-Chiozzi, and J Huff. 2001. "Clinical Application of Pharmacogenetics." *Trends in molecular medicine* 7(5): 201–4. <http://www.ncbi.nlm.nih.gov/pubmed/11325631> (Marzo , 2016).
4. Evans, W. E. 1999. "Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics." *Science* 286(5439): 487–91. <http://www.sciencemag.org/cgi/doi/10.1126/science.286.5439.487> (Marzo, 2016).
5. Fabbri, Giulia, and Riccardo Dalla-Favera. 2016. "The Molecular Pathogenesis of Chronic Lymphocytic Leukaemia." *Nature Reviews Cancer* 16(3): 145–62. http://www.nature.com/nrc/journal/v16/n3/full/nrc.2016.8.html?WT.ec_id=NRC201603&spMailingID=50777219&spUserID=MTc2NjgwNTE2OAS2&spJobID=863115711&spReportId=ODYzMTE1NzExSO (Marzo, 2016).
6. Kann, Maricel G. 2010. "Advances in Translational Bioinformatics: Computational Approaches for the Hunting of Disease Genes." *Briefings in bioinformatics* 11(1): 96–110. <http://www.ncbi.nlm.nih.gov/pubmed/20007728> (Marzo , 2016).
7. Piñero, Janet et al. 2015. "DisGeNET: A Discovery Platform for the Dynamical Exploration of Human Diseases and Their Genes." *Database : the journal of biological databases and curation* 2015: bav028. <http://www.ncbi.nlm.nih.gov/pubmed/25877637> (Abril, 2016).
8. Landrum, Melissa J et al. 2014. "ClinVar: Public Archive of Relationships among Sequence Variation and Human Phenotype." *Nucleic acids research* 42(Database issue): D980–85. <http://www.ncbi.nlm.nih.gov/pubmed/24234437> (Mayo, 2016).
9. Armstrong, Scott A et al. 2003. "Inhibition of FLT3 in MLL: Validation of a Therapeutic Target Identified by Gene Expression Based Classification." *Cancer Cell* 3(2): 173–83. [http://doi.org/10.1016/S1535-6108\(03\)00003-5](http://doi.org/10.1016/S1535-6108(03)00003-5) (Marzo, 2016)
10. <http://www.disgenet.org/web/DisGeNET/menu> (DisGenet) (Abril, 2016)
11. Wu, Tsung Jung et al. 2014. "A Framework for Organizing Cancer-Related Variations from Existing Databases, Publications and NGS Data Using a High-Performance Integrated Virtual Environment (HIVE)." *Database : the journal of biological databases and curation* 2014: 1–11. <http://doi.org/10.1093/database/bau022> (Marzo, 2016)
12. <https://hive.biochemistry.gwu.edu/tools/biomuta/> (BioMuta) (Abril, 2016)
13. Breuza, Lionel et al. 2016. "The UniProtKB Guide to the Human Proteome." *Database* 2016: bav120.

- <http://database.oxfordjournals.org/lookup/doi/10.1093/database/bav120>
(Marzo, 2016).
14. <http://www.uniprot.org> (UniprotKB) (Abril, 2016)
 15. Forbes, S. A. et al. 2008. "The Catalogue of Somatic Mutations in Cancer (COSMIC)." *Current Protocols in Human Genetics* (SUPPL. 57).
<http://doi.org/10.1002/0471142905.hg1011s57> (Marzo, 2016)
 16. <http://cancer.sanger.ac.uk/census> (COSMIC) (Abril, 2016)
 17. http://www.ensembl.org/info/genome/genebuild/genome_annotation.html
(Mayo, 2016)
 18. Shannon, Paul et al. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome research* 13(11): 2498–2504.
<http://www.ncbi.nlm.nih.gov/pubmed/14597658> (Abril, 2016).
 19. Wu, Guanming et al. 2014. "ReactomeFIViz: A Cytoscape App for Pathway and Network-Based Data Analysis." *F1000Research* 3.
<http://f1000research.com/articles/3-146/v2> (Abril, 2016).
 20. <https://www.ebi.ac.uk/chembl/drugability/> (Mayo, 2016).
 21. Wagner, Alex H et al. 2016. "DGIdb 2.0: Mining Clinically Relevant Drug-Gene Interactions." *Nucleic acids research* 44(D1): D1036–44.
<http://www.ncbi.nlm.nih.gov/pubmed/26531824> (Mayo, 2016).
 22. Grosdidier, Aurélien, Vincent Zoete, and Olivier Michielin. 2011. "SwissDock, a Protein-Small Molecule Docking Web Service Based on EADock DSS." *Nucleic acids research* 39(Web Server issue): W270–77.
<http://www.ncbi.nlm.nih.gov/pubmed/21624888> (Mayo, 2016).
 23. Pettersen, Eric F et al. 2004. "UCSF Chimera--a Visualization System for Exploratory Research and Analysis." *Journal of computational chemistry* 25(13): 1605–12. <http://www.ncbi.nlm.nih.gov/pubmed/15264254> (Mayo, 2016).
 24. Roskoski, Robert. 2004. "Src Protein-Tyrosine Kinase Structure and Regulation." *Biochemical and Biophysical Research Communications* 324(4): 1155–64. <http://doi.org/10.1016/j.bbrc.2004.09.171> (Mayo, 2016)
 25. Liu, Y. et al. 1998. "A Molecular Gate Which Controls Unnatural ATP Analogue Recognition by the Tyrosine Kinase v-Src." *Bioorganic and Medicinal Chemistry* 6(8): 1219–26. [http://doi.org/10.1016/S0968-0896\(98\)00099-6](http://doi.org/10.1016/S0968-0896(98)00099-6) (Mayo, 2016)
 26. Witucki, Laurie A. et al. 2002. "Mutant Tyrosine Kinases with Unnatural Nucleotide Specificity Retain the Structure and Phospho-Acceptor Specificity of the Wild-Type Enzyme." *Chemistry and Biology* 9(1): 25–33. [http://doi.org/10.1016/S1074-5521\(02\)00091-1](http://doi.org/10.1016/S1074-5521(02)00091-1) (Mayo, 2016)
 27. Lange, Gudrun et al. 2003. "Requirements for Specific Binding of Low Affinity Inhibitor Fragments to the SH2 Domain of (pp60)Src Are Identical to Those for High Affinity Binding of Full Length Inhibitors." *Journal of medicinal chemistry* 46(24): 5184–95.
<http://www.ncbi.nlm.nih.gov/pubmed/14613321> (Mayo, 2016).