



Identificación de posibles dianas terapéuticas para el tratamiento de la leucemia

Mario Colombo Canut
Máster en Bioinformática y Bioestadística

Melchor Sánchez Martínez

13/01/2016

GNU Free Documentation License (GNU FDL)

Copyright © 2016 MARIO COLOMBO
CANUT

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

FICHA DEL TRABAJO FINAL

Título del trabajo:	Identificación de posibles dianas terapéuticas para el tratamiento de la leucemia
Nombre del autor:	Mario Colombo Canut
Nombre del consultor:	Melchor Sánchez Martínez
Fecha de entrega (mm/aaaa):	01/2016
Área del Trabajo Final:	Bioinformática farmacéutica
Titulación:	<i>Máster en Bioinformática y Bioestadística</i>
Resumen del Trabajo (máximo 250 palabras):	
<p>En el presente trabajo se presenta un <i>workflow</i> para identificar posibles dianas terapéuticas para una determinada patología. utilizando las bases de datos públicas. En un primer paso se seleccionan los genes implicados, extrayéndolos mediante un <i>bash script</i> de las bases de datos BioMuta y DisGenet. Con estos genes se construye una red de interacción en Cytoscape, permitiendo extraer los nodos centrales con una gran conectividad (2600 vecinos grado 2).</p> <p>Con estos 68 genes centrales en la red cargamos todas las proteínas relacionadas y sus interacciones gracias a la ayuda de Biogrid y Reactome, con lo que obtenemos una nueva red. De esta red obtenemos también las proteínas centrales (2500 vecinos grado 2) obteniendo 16 candidatas a dianas terapéuticas.</p> <p>Testamos la viabilidad de dichas dianas, calculando por una lado su druggability, y por otro realizando una búsqueda en Drug Gene Interaction Database, donde obtendremos un índice de acuerdo con posibles fármacos para la misma.</p> <p>De esta forma obtenemos que la mejor diana terapéutica para la leucemia sería la cadena alfa de la casein kinasa 2 (CSNK2A1) ya que presenta una druggability muy alta del 92%. Para está diana existen posibles fármacos candidatos, el ácido elágico (<i>Ellagic Acid</i>) presente en frutas y verduras como las fresas o el resveratrol presente en la piel de la uva. Ambos compuestos están siendo estudiados para otras patologías.</p>	

Abstract (in English, 250 words or less):

This project presents a workflow to identify potential therapeutic targets for a given pathology using public databases. As first step, the genes involved are selected, pulling them through a bash script from BioMuta bases and DisGenet data. With these genes, an interaction network on Cytoscape is built, allowing the central nodes to be extracted with great connectivity (2600 neighbors grade 2).

Along with these 68 core genes in the network, all related proteins and their interactions are loaded in the mesh, with the help of Biogrid and Reactome, and we get a new network. From this network we will also get the core proteins (2500 neighbors grade 2) obtaining 16 candidates for therapeutic targets.

We tested the viability of such targets by calculating on one hand their druggability and on the other by searching in Drug Gene Interaction Database, where we will obtain an index of potential drugs to deal with it.

As a consequence we obtain that the best therapeutic target for leukemia would be the alpha chain of casein kinase 2 (CSNK2A1) because it has a very high druggability, of 92%. For this target there are potential drug candidates, such as the Ellagic Acid found in fruits and vegetables as strawberries or the resveratrol, found in grape skins. Both compounds are being studied for other pathologies.

Palabras clave (entre 4 y 8):

leukemia, CSNK2A1, casein kinase 2, ellagic acid, druggability, gene target, drug repositioning, resveratrol

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.....	3
1.6 Breve descripción de los otros capítulos de la memoria.....	3
2. Investigación.....	4
2.1 Identificación de genes o biomoléculas relevantes en la literatura.....	4
2.2 Diseño de un 'script' para la obtención de genes relevantes a partir de las bases de datos.....	4
2.3 Integración de la información obtenida.....	5
2.4 Selección de proteínas centrales con la ayuda de Cytoscape, biogrid y reactome.....	6
2.5 Identificación de genes o biomoléculas relevantes en la literatura.....	7
3. Conclusiones.....	10
4. Glosario.....	11
5. Bibliografía.....	12

Lista de figuras

Figura 1 Diagrama de Gantt con la planificación del proyecto.....	3
Figura 2 Bash script para la búsqueda de genes relacionados con una patología en DisGenet y BioMuta.....	4
Figura 3 Cytoscape con la red correspondiente a los genes relacionados con la leucemia.....	5
Figura 4 Representación de los 68 genes centrales en Cytoscape.....	6
Figura 5 Interactoma para los nodo centrales obtenidos en el apartado 2.4.....	6
Figura 6 Red con las 16 proteínas principales.....	7
Figura 7 Red con las 16 proteínas principales representadas en forma circular para una mejor visualización.....	8

1. Introducción

1.1 Contexto y justificación del Trabajo

El inicio del siglo XXI ha traído consigo una revolución, la revolución genómica. Tras más de una década de trabajo, se publicaba en 2001^{[1],[2]} el borrador de la secuencia del genoma humano. A pesar de ser un gran hito en la historia de la biología, no fue más que el inicio de la gran revolución por venir. A pesar de su importancia, disponer de la secuencia del genoma, no implica conocer su funcionamiento. Si realizamos una analogía con una lengua, disponer del genoma humano implicaría conocer el alfabeto, sin embargo queda un largo trayecto hasta escribir un libro como “Cien años de soledad” de Gabriel García Márquez.

Podemos considerar la publicación de estos trabajos como el pistoletazo de salida para la nueva biología. Ya en aquel momento se especulaba sobre la importancia de la genómica para el desarrollo de fármacos^[3] o para entender la complejidad de los sistemas biológicos^[4]. Pero la eclosión de nuevos campos y técnicas ha superado las expectativas. Gracias a los microarrays hemos conocido mejor la expresión génica ante factores diferenciales^[5], y a día de hoy las nuevas técnicas de secuenciación permiten generar ingentes cantidades de información^[6], siendo por tanto imprescindible desarrollar técnicas estadísticas avanzadas y algoritmos informáticos^[7].

Dentro este contexto de revolución genómica en el que estamos inmersos, la búsqueda de la medicina personalizada y de dianas terapéuticas para la obtención de nuevos fármacos es un campo en pleno desarrollo, en el cual se enmarca este trabajo^{[8],[9],[10]}.

No cabe duda que el cáncer es una de las grandes epidemias de la actualidad^[11], generando una gran cantidad de esfuerzo y información a su alrededor. La leucemia en concreto, es el cáncer más común en niños, con una gran mortalidad, del alrededor del 30% para la leucemia mieloide aguda (AML), del 15% para leucemia linfocítica aguda (ALL) o del 50% para leucemia mielomonocítica juvenil (JMML)^[12]. Cabe destacar que aunque la leucemia presenta diversos tipos como los mencionados anteriormente, para este proyecto la vamos a abordar como una única patología, ya que las rutas metabólica implicadas en todas ellas son similares^[13].

Como hemos comentado, en la actualidad se genera una gran cantidad de información genómica, siendo esta además de dominio público a través de bases de datos disponibles en internet. El presente trabajo pretende extraer información sobre la leucemia de estas bases de datos, integrarla y tratar de identificar unas posibles dianas terapéuticas para esta patología. Una vez identificadas, realizaremos un análisis de su

viabilidad y buscaremos posibles fármacos para ellas, a través de la técnica del *drug repositioning*^[8].

1.2 Objetivos del Trabajo

1.2.1. Objetivos generales:

- 1 Identificar posibles dianas terapéuticas para la leucemia.
- 2 Desarrollar herramientas de trabajo para la extracción de información relevante de las bases de datos públicas.

1.2.2. Objetivos específicos:

- 1 Identificar posibles dianas terapéuticas para la leucemia.
 - 1.1 Extraer genes o biomoléculas relevantes en las rutas metabólicas de la leucemia de la bibliografía.
 - 1.2 Integrar la información con diversos métodos para la selección de las dianas.
 - 1.3 Testear la viabilidad de estas moléculas mediante el cálculo de su "*druggability*".
- 2 Desarrollar y utilizar herramientas de trabajo para la extracción de información relevante de las bases de datos públicas.
 - 2.1 Comparar diversos métodos para la extracción de la información de las bases de datos públicas.
 - 2.2 Crear una herramienta de software que permita cruzar la información obtenida en la búsqueda bibliográfica con las bases de datos existentes.
 - 2.3 Utilizar herramientas de integración de la información como Cytoscape o Scan.
 - 2.4 Diseñar de un algoritmo para la selección de las posibles dianas terapéuticas.
 - 2.5 Utilizar y/o desarrollar herramientas para testar la viabilidad de las dianas ("*druggability*")

1.3 Enfoque y método seguido

Para la realización de este proyecto vamos a utilizar la información genómica disponible en bases de datos públicas. Cada día se generan cantidades ingentes de información, y uno de los grandes retos de la actualidad es analizarla e integrarla para sacar conocimiento útil. En este proyecto, utilizaremos una metodología conocida. Extraemos información relevante a la leucemia de las bases de datos públicas ('data-mining'), la integramos gracias a un software con sus interrelaciones, a partir de estas relaciones escogemos nodulos (genes) centrales basados en algoritmos internos del programa relacionado con la teoría de grafos. A partir de estos genes, cargamos toda la red de proteínas relacionadas, y a partir del mismo principio seleccionaremos las proteínas centrales que serán nuestras candidatas probables.

A continuación, vemos cual es su *druggability* y si existen posibles fármacos para intercatuar con ella (*drug repositioning*).

Se trata de una metodología muy utilizada, siendo una de las tendencias más importantes del mercado farmacéutico, ya que supone una mayor efectividad en el diseño de fármacos.

1.4 Planificación del Trabajo

Identificamos las siguientes tareas, secuenciada en un diagrama de Gantt (Figuras 1):

- 1.1.1 Revisión bibliográfica de libros y artículos científicos.
- 1.1.2 Identificación de genes o biomoléculas relevantes en la literatura.
- 1.2.1 Integración de la información obtenida.
- 2.1.1 Comparación de los resultados obtenidos por diversos métodos.
- 2.2.1 Diseño de un 'script' para la obtención de genes relevantes a partir de las bases de datos.
- 2.3.1 Integración de información con Cytoscape
- 2.4.1 Selección de proteínas centrales con la ayuda de Cytoscape, biogrid y reactome.
- 2.5.1 Utilización de herramientas informáticas que permitan testar la viabilidad de las dianas identificadas.
3. Redacción de la memoria.
4. Preparación de la defensa del trabajo.

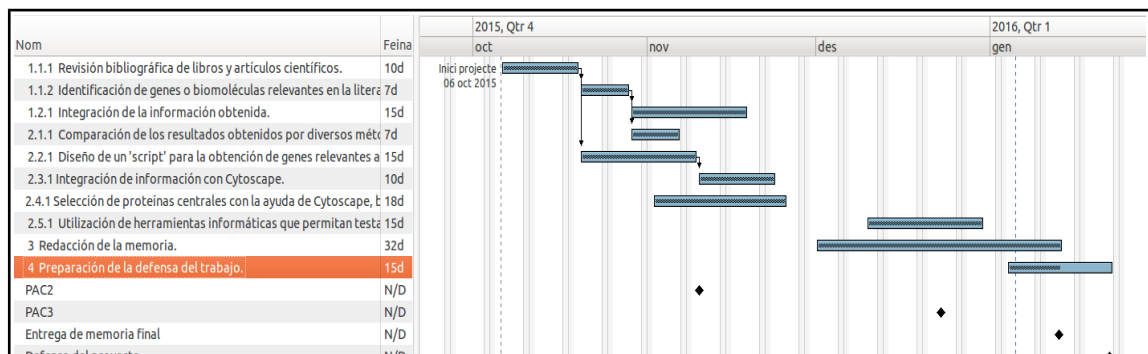


Figura 1. Diagrama de Gantt con la planificación del proyecto

1.5 Breve resumen de productos obtenidos

Se ha diseñado un bash script que permite combinar BioMuta y DisGenet, obteniendo una lista de genes implicados en una cierta patología.

Se han obtenido 16 posibles dianas terapéuticas para la leucemia, ordenadas por su *druggability* y con posibles fármacos para ellas.

1.6 Breve descripción de los otros capítulos de la memoria

La memoria se estructura siguiendo el desarrollo de las tareas planificadas, de forma secuencial, y alterando un poco el orden establecido de las mismas. Comentando para cada una de ellas los resultados obtenidos y las decisiones realizadas.

2. Investigación

2.1 Identificación de genes o biomoléculas relevantes en la literatura.

Se ha procecido tanto a una revisión general sobre la leucemia en textos científicos^[13] como a una revisión de bases de datos científicas públicas. Se han visitado las siguientes bases de datos:

- UniprotkB^[14]: UniprotkB es la base de datos de referencia para la anotación de proteínas.
- ClinVar^[15]: ClinVar recoge información sobre variaciones genéticas y su relación con la salud humana.
- BioMuta^[16]: BioMuta integra la información de diversas bases de datos, publicaciones y análisis NGS relacionadas con el cáncer en una sola plataforma.
- DisGenet^[17]: DisGenet integra la información de diversas bases de datos, relacionadas con las enfermedades.
- Ensembl^[18]: Ensembl es la base de datos de referencia sobre el genoma.
- Nectar^[19]: Es una plataforma que integra la información de diversas bases de datos como Uniprot o Ensembl.

Para el desarrollo del presente trabajo se ha considerado que la bases de datos más oportunas para trabajar son BioMuta y DisGenet. Ambas integran información de otras bases de datos como UniprotkB o ClinVar, y además relacionan la información con enfermedades en general (DisGenet) o con el cáncer específicamente. Nectar también se considera una base de datos interesante, però no permite la descarga de la base de datos integrada, sino de las anotaciones individuales. También se ha utilizado una herramienta de 'text-mining' on-line, PubTator^[20], pero sus resultados no me han sido prácticos a la hora de trabajar, ya no he logrado sacar una lista con genes únicos a partir de la información obtenida.

Se han extraído 8290 genes relacionados con la leucemia combinando BioMuta y DisGenet. Estos genes se han de integrar en una red para después seleccionar los nodos centrales.

2.2 Diseño de un 'script' para la obtención de genes relevantes a partir de las bases de datos.

En este *script* (Figura 2), el usuario debe situarse en la carpeta donde tiene las bases de datos de Biomuta y DisGenet descargas, ya sea en formato .txt o .csv. El script recibe 3 argumentos, los dos primeros son el nombre de los archivos con las bases de datos pertinentes, el orden no importa. El tercer argumento es la patología que nos interesa. El *script* arroja como resultado un archivo de texto (output.txt) con una lista de los genes relacionados con la patología buscada.

```
#!/bin/bash
# Script para unir genes de bases de datos BioMuta y Disgenet en
base a una patología
# El usuario debe estar situado en la carpeta con los archivos
pertinentes
# El script recibe tres argumentos:
# los dos primeros archivos de las bases de datos en formato .txt
o .csv
# y el tercero la enfermedad
# Mario Colombo 2015
args=("$@")
cat ${args[0]}| grep ${args[2]}| awk '{print $2}'| sort | uniq >
tmp1.txt
cat ${args[1]}| grep ${args[2]} | awk '{print $2}'| sort | uniq >
tmp2.txt
cat tmp1.txt tmp2.txt | sort | uniq > output.txt
echo "El número de genes únicos encontrados es:"
wc -l output.txt
exit;
```

Figura 2. Bash script para la búsqueda de genes relacionados con una patología en DisGenet y BioMuta

2.3 Integración de la información obtenida.

A partir de los genes hallados en las bases de datos, se ha procedido a su integración con Cytoscape^[21].

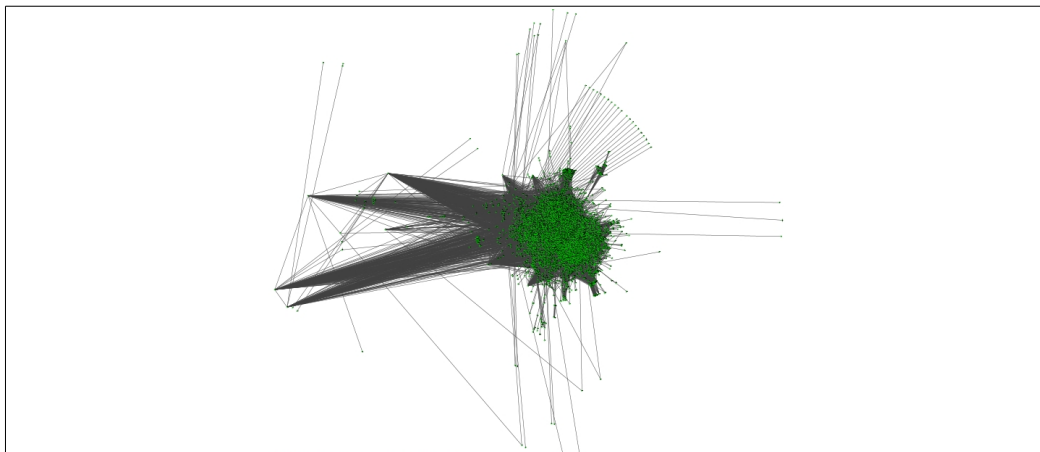


Figura 3. Cytoscape con la red correspondiente a los genes relacionados con la leucemia.

Comentar, que a pesar de cargar en el programa 8290 nombres de genes, solo se muestran 4602 nodos (Figura 3), creo que se debe a la utilización de sinónimos para los genes en ciertas bases de datos, produciendo que en lugar de 8290 genes, realmente el número real sea muy inferior.

A partir de aquí se realiza un filtrado para seleccionar nodos centrales. De acuerdo con la teoría de grafos, consideramos como nodos centrales aquellos que presentan una mayor conectividad. Para ello, utilizamos los algoritmos internos de Cytoscape, y filtramos aquellos nodos que presenten al menos 2600 vecinos con una distancia de grado 2. Obteniendo 68 genes centrales en la red que podemos observar en la siguiente figura (Figura 4).

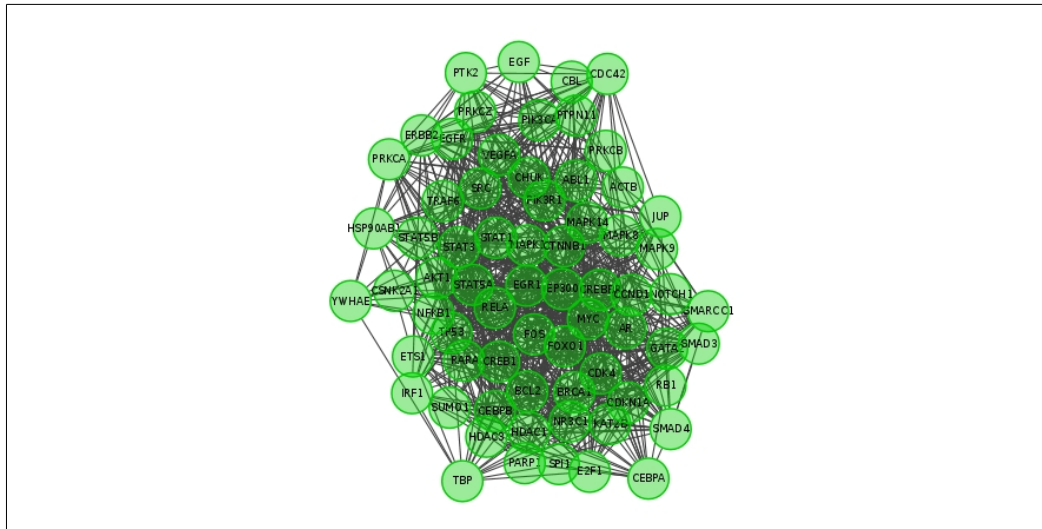


Figura 4. Representación de los 68 genes centrales en Cytoscape.

2.4 Selección de proteínas centrales con la ayuda de Cytoscape, biogrid y reactome.

A partir de estos 68 genes, vamos a cargar todas el interactoma, cargando las proteínas asociadas y sus relaciones con la ayuda de Biogrid^[22] y Reactome^[23], generando la siguiente red (Figura 5):

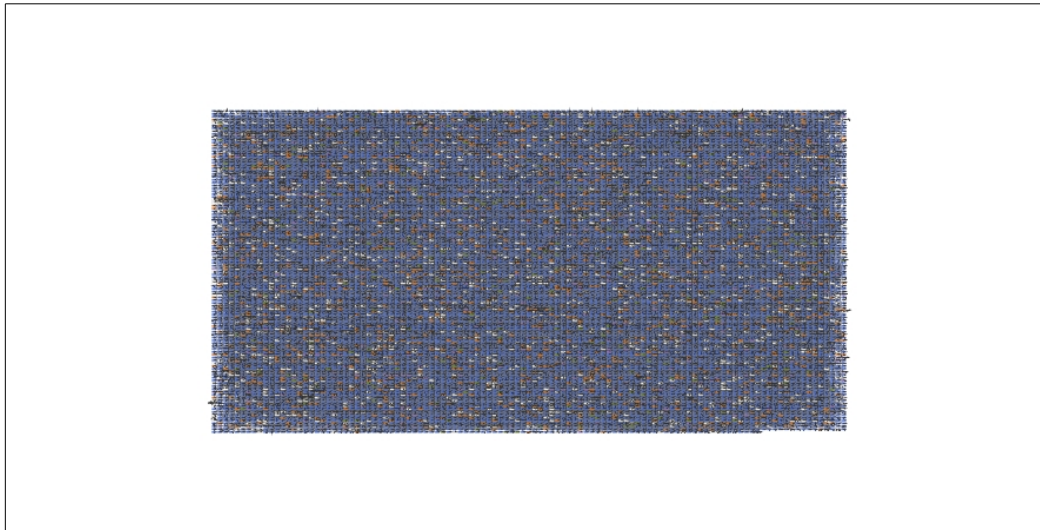


Figura 5. Interactoma para los nodo centrales obtenidos en el apartado 2.4

Una red con 12528 nodos. De nuevo vamos a realizar un filtrado para seleccionar los nodos centrales, en este caso con 3500 vecinos con una distancia de grado 2. Y obtenemos 16 nodos principales (Figura 6):

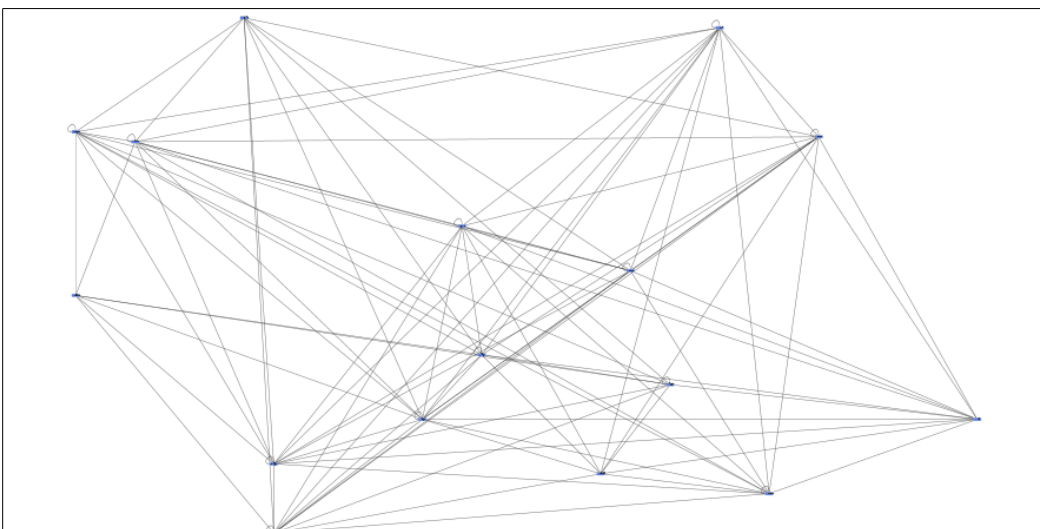


Figura 6. Red con las 16 proteínas principales

Si mostramos, el gráfico en forma circular (Figura 7):

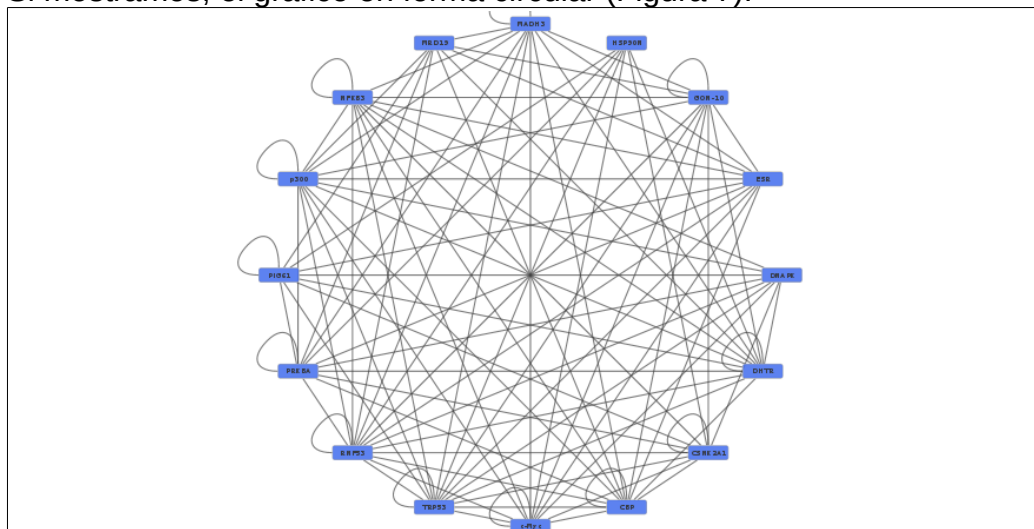


Figura 7. Red con las 16 proteínas principales representadas en forma circular para una mejor visualización

Obtenemos por tanto las siguientes posibles dianas terapéuticas, adjuntando el nombre y su identificador en UniprotkB^[14]

1. DHTR : Androgen receptor. P10275
2. c-Myc : Myc proto-oncogene protein. P01106
3. p300 : Histone acetyltransferase p300. Q09472
4. HSP90N : Heat shock protein. P07900
5. CBP : CREB-binding protein. Q92793
6. DNAPK : DNA-dependent protein kinase catalytic subunit. P78527
7. MADH3 : Mothers against decapentaplegic homolog 3. P84022
8. PRKBA : RAC-alpha serine/threonine-protein kinase. P31749
9. GON-10 : Histone deacetylase 1. Q13547
10. NFKB3 : Transcription factor p65. Q04206
11. ESR1 : Estrogen receptor. P03372
12. MRD19 : Catenin beta-1. P35222
13. PIG61 : Epidermal growth factor receptor. P00533
14. CSNK2A1 : Casein kinase II subunit alpha. P68400
15. RNF53 : Breast cancer type 1 susceptibility protein. P38398
16. TRP53 : Thiamine-triphosphatase. Q9BU02

2.5 Estudio de viabilidad de las dianas identificadas

Para ver la viabilidad de las dianas encontradas vamos a utilizar dos herramientas web, por un lado DrugEBlity del EMBL-EBI^[24] nos permitirá calcular la *druggability* de la diana, es decir, evalúa el promedio de la idoneidad del sitio de unión para moléculas pequeñas bajo regla de 5 Lippinsky^[25]. Utilizaremos también la base de datos DGIdb^{[26],[27][28]} para la minería de datos de las interacciones entre genes y fármacos. Obteniendo los siguientes resultados:

Diana	Druggability _[24]	Score DGldb _[24]
CSNK2A1	92%	3
PRKBA [AKT1]	64%	6
ESR1	55%	14
DHTR [AR]	53%	11
HSP90N [HSP90AA1]	51%	6
PIG61 [EGFR]	50%	7
p300	44%	0
MRD19	42%	0
NFKB3	25%	0
CBP	12%	0
MADH3	20%	0
RNF53 [BRCA1]	1%	1
TRP53	0%	1
C-MYC	0%	1
GON-10 [HDAC1]	-	6
DNAPK	-	1

Observamos que la mejor candidata a diana terapéutica para la leucemia sería la cadena alfa de la casein kinasa 2 (CSNK2A1) ya que presenta una druggability muy alta del 92%. Para esta diana existen posibles fármacos candidatos, el ácido elágico_[29] presente en frutas y verduras como las fresas o arándanos. Este compuesto se encuentra en fase de estudio clínicos para otras patologías como el linfoma folicular o la protección cardiovascular en casos de obesidad y presenta un índice 3 en DGldb. O el resveratrol_[30] con un índice de 1 DGldb se trata de un polifenol presente en la piel de la uva que se está estudiando su uso como tratamiento para el herpes labial y como posible anticancerígeno.

Otra posible candidata es la RAC-alpha serine/threonine-protein kinase (PRKBA), presenta la segunda mayor *druggability*, un 64% y con un índice de 6 en DGldb para el trióxido de arsénico_[31] y que de hecho ya se utiliza para el tratamiento de la leucemia. Además del trióxido de arsénico, la lista de fármacos candidatos es enorme, estando muchos de ellos en investigación actualmente (ver archivo anexo).

Destacar también como posibles candidatas los receptores hormonales, tanto de andrógenos (DHTR) como de estrógenos (ESR1), ambas presentan una druggability un poco superior al 50% y con los mayores índices para DGldb 11 y 14 respectivamente. Tienen fármacos anticancerígenos conocidos como el torimefano_[32] para el cáncer de mama (ESR1) o la nilutamida_[33] para el cáncer de próstata.

También presentan una druggability interesante, superior al 50% la *Heat shock protein* HSP90N y el receptor para el crecimiento epitelial PIG61. HSP90N tiene un índice DGldb de 6 con el Vorinostat_[34], un fármaco que se está investigando para el tratamiento de una clase del linfoma de células T cutáneas. PIG61 presenta también un índice DGldb de 6 con el Vandetanib fármaco aprobado para tratar una gran cantidad de tumores.

El resto de candidatas, las podríamos descartar dada su baja druggability y su bajo índice DGldb.

En conclusión, hemos identificado posibles dianas terapéuticas, con potencial para ser tratadas. Como ya se ha comentado, nuestra mejor diana la cadena alfa de la casein kinasa 2 (CSNK2A1), que además hemos encontrado sustancias candidatas para el tratamiento que no están establecidas (ácido elágico y el resveratrol).

3. Conclusiones

3.1 Resultado del estudio

- A partir de la información disponible en bases de datos públicas es posible identificar posibles dianas terapéuticas para una determinada enfermedad.
- Se han identificado posibles dianas terapéutica y fármacos candidatos para ellas, no obstante hace falta constatar estos datos con más estudios tanto a nivel teórico, como llegado el caso con estudios clínicos en sus fases más preliminares.

3.2 Desarrollo del proyecto

- Los objetivos del proyecto se han cumplido parcialmente. Por un lado, se ha asolido el objetivo principal de identificar unas posibles dianas terapéuticas para la leucemia, pero por otro lado no se han desarrollado herramamientas propias, si no que se han utilizado las ya existentes.
- El plan del proyecto se han seguido de acuerdo a lo indicado, intentando cumplir todas las tareas para el plazo indicado, y adaptando objetivos y tareas de acuerdo con las indicaciones del tutor para la una correcta consecución del proyecto.

4. Glosario

AML: leucemia mieloide aguda.

ALL: leucemia linfocítica aguda.

BioMuta: Base de datos que integra la información de diversas bases de datos, publicaciones y análisis NGS relacionadas con el cáncer en una sola plataforma.

ClinVar: Base de datos que recoge información sobre variaciones genéticas y su relación con la salud humana.

Cystoscape: plataforma de software de código abierto para la visualización de redes complejas y la integración de éstas con datos genómicos

DisGenet: DisGenet integra la información de diversas bases de datos, relacionadas con las enfermedades.

DGIdb: Drug-gene interaction database. Base de datos para el data-mining de fármacos para targets genómicos.

Druggabilty: es un término usado en el descubrimiento de fármacos para describir una diana biológica) que se sabe que o se predice que se pueden unir a a un fármaco. Además, por definición, la unión del fármaco a una diana debe alterar la función con un beneficio terapéutico para el paciente.

Drug repositioning: es la técnica que busca utilizar fármacos conocidos para nuevas indicaciones.

Ensembl: Ensembl es la base de datos de referencia sobre el genoma.

JMML: leucemia mielomonocítica juvenil

Nectar: Es una plataforma que integra la información de diversas bases de datos como Uniprot o Ensembl.

UniprotkB: UniprotkB es la base de datos de referencia para la anotación de proteínas.

5. Bibliografía

- [1] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Szustakowki, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. <http://doi.org/10.1038/35057062>
- [2] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Zhu, X. (2001). *The sequence of the human genome*. *Science* (New York, N.Y.), 291(5507), 1304–51. <http://doi.org/10.1126/science.1058040>
- [3] [Ohlstein EH](#), [Ruffolo RR Jr](#), [Elliott JD](#) (2000). Drug discovery in the next millennium. *Annu Rev Pharmacol Toxicol*. 2000;40:177-91.
- [4] Kitano, H. (2002). Systems biology: a brief overview. *Science* (New York, N.Y.), 295(5560), 1662–4. <http://doi.org/10.1126/science.1069492>
- [5] Allison, D. B., Cui, X., Page, G. P., & Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews. Genetics*, 7(1), 55–65. <http://doi.org/10.1038/nrg1749>
- [6] Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–45. <http://doi.org/10.1038/nbt1486>
- [7] De Brevern, A. G., Meyniel, J.-P., Fairhead, C., Neuvéglise, C., & Malpertuy, A. (2015). Trends in IT Innovation to Build a Next Generation Bioinformatics Solution to Manage and Analyse Biological Big Data Produced by NGS Technologies. *BioMed Research International*, 2015, 904541. <http://doi.org/10.1155/2015/904541>
- [8] Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews. Drug Discovery*, 3(8), 673–83. <http://doi.org/10.1038/nrd1468>
- [9] Driscoll, C. O. (2004). A virtual space odyssey. *Nature Reviews. Drug Discovery*, (May), 1–4.
- [10] Schenone, M., Dančák, V., Wagner, B. K., & Clemons, P. a. (2013). Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology*, 9(4), 232–240. <http://doi.org/10.1038/nchembio.1199>
- [11] Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., & Thun, M. J. (2008). Cancer Statistics, 2008. *CA: A Cancer Journal for Clinicians*, 58(2), 71–96. <http://doi.org/10.3322/CA.2007.0010>
- [12] <http://www.cancer.org/espanol/cancer/leucemiaenninos/guiadetallada/leucemia-en-ninos-early-survival-rates> [20/12/2015]
- [13] Weinberg, R.A. *Biology of Cancer*. Ed. Garland Science. 2n Edition. 2015
- [14] <http://www.uniprot.org/help/uniprotkb> [15/10/2015]
- [15] <http://www.ncbi.nlm.nih.gov/clinvar/intro/> [15/10/2015]
- [16] Wu, T.-J., Shamsaddini, A., Pan, Y., Smith, K., Crichton, D. J., Simonyan, V., & Mazumder, R. (2014). A framework for organizing cancer-related variations from existing databases, publications and NGS data using a High-performance Integrated Virtual Environment (HIVE). *Database*, 2014(0), bau022–bau022. <http://doi.org/10.1093/database/bau022>
- [17] Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., ... Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, 2015(0), bav028–bav028. <http://doi.org/10.1093/database/bav028>
- [18] <http://www.ensembl.org/index.html> [16/10/2015]
- [19] <http://nectarmutation.org/main> [20/10/2015]

- [20] <http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/PubTator/index.cgi?user=User705811436> [20/10/2015]
- [21] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Ideker, T. (2003). Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks, (Karp 2001), 2498–2504. <http://doi.org/10.1101/gr.1239303.metabolite>
- [22] <http://thebiogrid.org/> [23/11/2015]
- [23] <http://www.reactome.org/> [23/11/2015]
- [24] <https://www.ebi.ac.uk/chembl/drugability/gene> [23/12/2015]
- [25] Lipinski, C.A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. Drug Discovery Today: Technologies **1** (4): 337:341. [doi:10.1016/j.ddtec.2004.11.007](http://doi.org/10.1016/j.ddtec.2004.11.007)
- [26] <http://dgidb.genome.wustl.edu/> [23/12/2015]
- [27] Spies, N. C., Koval, J., Das, I., Callaway, M. B., Kumar, R. D., Bose, R., ... David, E. (2014). DGldb - Mining the druggable genome, 10(12), 1–14. <http://doi.org/10.1038/nmeth.2689.DGldb>
- [28] Wagner, A. H., Coffman, A. C., Ainscough, B. J., Spies, N. C., Skidmore, Z. L., Campbell, K. M., ... Griffith, O. L. (2015). DGldb 2.0: mining clinically relevant drug–gene interactions. Nucleic Acids Research, gkv1165. <http://doi.org/10.1093/nar/gkv1165>
- [29] <http://www.drugbank.ca/drugs/DB08846> [9/1/2016]
- [30] <http://www.drugbank.ca/drugs/DB02709> [9/1/2016]
- [31] <http://www.drugbank.ca/drugs/DB01169> [9/1/2016]
- [32] <http://www.drugbank.ca/drugs/DB00539> [9/1/2016]
- [33] <http://www.drugbank.ca/drugs/DB00665> [9/1/2016]
- [34] <http://www.drugbank.ca/drugs/DB02546> [9/1/2016]

