

UNIVERSIDAD DE COSTA RICA  
SISTEMA DE ESTUDIOS DE POSGRADO

PLATAFORMA BIOINFORMÁTICA PARA LA IDENTIFICACIÓN DE  
NUEVAS INFERENCIAS DEL FENÓMENO DE COMPENSACIÓN DE DOSIS  
GÉNICA EN EL PANEL NCI-60

Tesis sometida a la consideración de la Comisión del Programa de Posgrado  
en Ciencias Biomédicas para optar al grado y título de Maestría Académica en Bioinformática y  
Biología de Sistemas

JORGE TORRES CALVO

Ciudad Universitaria Rodrigo Facio, Costa Rica

2021

“Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Ciencias Biomédicas de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Bioinformática y Biología de Sistemas”

---

PhD. José Molina Mora  
**Representante de la Decana  
Sistema de Estudios de Posgrado**

---

PhD. Rodrigo Mora Rodríguez  
**Profesor Guía**

---

PhD. José Guevara Coto  
**Lector**

---

M.Sc. Man Sai Acón Chan  
**Lector**

---

Dra. Mariela Eugenia Arias Hidalgo  
**Directora Programa de Posgrado en  
Ciencias Biomédicas**

---

Jorge Torres Calvo  
Sustentante

## Tabla de contenido

Hoja de aprobación.....	ii
Tabla de contenido .....	iii
1. Resumen .....	iv
2. Lista de figuras .....	v
3. Lista de tablas .....	vi
4. Introducción.....	1
5. Problema.....	7
5.1. Justificación.....	7
5.2. Hipótesis.....	10
5.3. Objetivo general .....	10
5.4. Objetivos específicos.....	10
6. Métodos.....	12
6.1. Desarrollo de la plataforma bioinformática BioNetUCR.....	12
6.1.1. Componentes visuales de la plataforma bioinformática.....	13
6.1.2. Comandos soportados.....	16
6.1.3. Características no funcionales .....	16
6.1.4. Relaciones soportadas por BioNetUCR .....	17
6.2. Incorporación de bases de datos a la plataforma desarrollada .....	17
6.3. Construcción de la topología de la red y comparación con la de (Acón, 2018).....	17
6.4. Construcción del modelo mejorado de biología de sistemas .....	18
6.5. Validación <i>in silico</i> de la hipótesis e inferencia de nuevos conocimientos .....	18
6.5.1. Modelo mínimo incorporando el número de copias para todas las especies.....	18
6.5.2. Modelo mínimo extendido hacia genes del metabolismo .....	18
6.5.3. Compensación de dosis génica de MYC y STAT3 con datos de CCLE.....	19
7. Resultados .....	20
8. Discusión.....	35
9. Conclusión.....	42
10. Anexos.....	44
11. Bibliografía.....	46

## 1. Resumen

Las células generan transcritos en respuesta a estímulos internos y externos, siendo algunos codificantes de proteínas y otros no codificantes. Los transcritos que codifican proteínas son regulados por varios elementos, ya sean de tipo proteico como los factores de transcripción, o no codificante como los microARNs. Los factores de transcripción y microARNs forman redes complejas de interacción que deben ser estudiadas para comprender los mecanismos de regulación génica, su relevancia en el desarrollo de patogénesis y en la investigación para identificar nuevos biomarcadores. Una de las técnicas que se utiliza en la actualidad para estudiar las redes de regulación es el modelaje *in silico* de las mismas, lo que permite hacer simulaciones, validar hipótesis y entender cuáles son los elementos relevantes de una determinada red.

Una investigación previa (Acón, 2018) utilizó los perfiles de expresión del panel NCI-60 de células cancerosas para encontrar una red mínima con compensación génica. Estas células tienen la característica de que presentan múltiples copias de los cromosomas, pero los niveles de algunas proteínas se mantienen en rangos normales. Partiendo de esta investigación se planteó la hipótesis de que el desarrollo de una plataforma informática de fácil uso permitiría encontrar nuevas interacciones entre los genes, factores de transcripción y microARNs involucrados en el fenómeno de compensación de dosis génica. Se desarrolló la plataforma planteada (BioNetUCR), dotándola de facilidades de integración con Cytoscape (Shannon et al., 2003), herramienta especializada en visualización de redes y con COPASI (COMplex PATHway SIMulator; Hoops et al., 2006), especializada en simulación y análisis de redes bioquímicas.

Utilizando la plataforma desarrollada se generó una red a partir de los elementos de la red del modelo mínimo encontrado en la investigación anterior, extendiendo un nivel hacia todos los genes regulados por cada nodo de la red mínima. Posteriormente se filtró la red dejando únicamente los genes contenidos en la base de datos del metabolismo humano llamada Virtual Metabolic Human. Se generó un modelo en COPASI con todos los genes del metabolismo y adicionalmente se estudiaron los genes del metabolismo regulados por al menos tres especies de la red mínima. Además se hicieron análisis de modelos en COPASI utilizando otros datos de expresión en células cancerosas, para analizar los resultados con respecto a los obtenidos con los datos de NCI-60.

## 2. Lista de figuras

Figura 1: Diagrama de la plataforma desarrollada y la interacción con otras aplicaciones.....	13
Figura 2: Opciones de Archivos y Preferencias en BioNetUCR .....	14
Figura 3: Opción de Herramientas en BioNetUCR .....	15
Figura 4: Secciones de BioNetUCR. ....	20
Figura 5: Red generada en investigación previa y la generada por BioNetUCR .....	22
Figura 6: Modelo mínimo en COPASI, generado en la plataforma previa y en BioNetUCR ....	23
Figura 7: Reacción de degradación para MYC en el modelo mínimo creado con la plataforma previa y BioNetUCR .....	23
Figura 8: Reacción de síntesis para MIR17 en el modelo mínimo creado con la plataforma previa y BioNetUCR .....	24
Figura 9: Modelo en COPASI para MYC y STAT3 generado en la plataforma previa y BioNetUCR .....	24
Figura 10: Comparativo de un parameter scan para el CNV de MYC y STAT3 en modelo con CNV diploide versus modelo con CNV de CellMiner .....	25
Figura 11: Red de genes del metabolismo regulados por las especies del modelo mínimo .....	26
Figura 12: Red de genes del metabolismo con más interacciones con FTs y microARNs del modelo mínimo. ....	29
Figura 13 Amplificaciones de MYC y STAT3 en datos de CCLE.....	32
Figura 14 Comparativo de datos experimentales de CCLE con datos simulados .....	32
Figura 15 Compensación de MYC y STAT3 en datos de CCLE .....	33
Figura 16 Resultados de expresión para MYC y STAT3 según número de copias en CCLE.....	33
Figura 17 Niveles de expresión para MYC y STAT3 para distintas líneas celulares con datos de CCLE .....	34
Figura 18 Índice de compensación para MYC y STAT3 en distintos tipos de cáncer con datos de CCLE .....	34
Figura 19: Configuración de preferencias en BioNetUCR.....	37

### 3. Lista de tablas

Tabla 1 Comparativo de características de la plataforma previa y BioNetUCR .....	21
Tabla 2: Comandos para crear red compatible con la desarrollada en plataforma previa .....	40
Tabla 3: Funciones de síntesis y degradación usados en genes, microARNs y FTs. ....	40



**Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.**

Yo, Jorge Torres Calvo, con cédula de identidad 204070547, en mi condición de autor del TFG titulado Plataforma Bioinformática para la identificación de nuevas inferencias del fenómeno de compensación de dosis génica en el panel NCI-60.

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI  NO \*

\*En caso de la negativa favor indicar el tiempo de restricción: \_\_\_\_\_ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

**FIRMA ESTUDIANTE**

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

## 4. Introducción

Las células habitan en un ambiente complejo y pueden percibir múltiples señales, tales como: temperatura, presión osmótica, concentración de nutrientes o químicos citotóxicos. También toman en cuenta la información intracelular, como el nivel de metabolitos o daño al ADN (Brown & Jackson, 2015). Como respuesta a esas señales las células activan varios mecanismos, entre ellos la síntesis de proteínas apropiadas, según las señales percibidas (Alon, 2006). Por ejemplo, el fumado provoca el aumento en la síntesis de proteínas específicas que facilitan el ingreso del virus SARS-CoV-2 en las células (Aliee et al., 2020). La regulación apropiada de la síntesis de proteínas es crítica para el crecimiento, proliferación y muerte celular (Kourtis & Tavernarakis, 2009).

Los organismos dedican una parte importante de sus genomas a producir proteínas y otros transcritos que regulan cuándo la expresión de cada gen se debe encender o apagar (Li & Johnson, 2010). El proceso de transcripción de genes es regulado principalmente por factores de transcripción (FT), los cuales se unen a los promotores de los genes blanco y activan o reprimen su transcripción (Hershberg & Margalit, 2006). Los factores de transcripción reconocen secuencias específicas de ADN para controlar la cromatina y la transcripción, formando un sistema complejo que guía la expresión del genoma. Están involucrados en distintos aspectos de la fisiología humana, así como enfermedades, por lo que el estudio de cómo se interrelacionan con otros elementos en la expresión génica, tiene mucha relevancia (Lambert et al., 2018). Desde la primera publicación en 1978 en que se mapea detalladamente cómo un factor de transcripción se adhiere a una secuencia de ADN (Tjian, 1978), el número de factores de transcripción identificados ha venido aumentando. Al 2018 se habían identificado más de 1600 factores de transcripción en el genoma humano (Lambert et al., 2018).

Existe además un segundo tipo de regulador de la expresión génica llamado microARN. Estos son ARNs no codificantes de alrededor de 22 nucleótidos de longitud (H. Zhang et al., 2015). Fueron identificados por primera vez en los años 90 en *C. elegans* (Wightman et al., 1994) y a inicios de la década del 2000 en mamíferos (Lagos-Quintana et al., 2002). Hay varios mecanismos por los cuales los microARNs reducen la expresión de proteínas específicas, algunos resultan en niveles reducidos de ARN mensajero (mARN), mientras que otros actúan únicamente para reducir la



expresión de proteínas (Mott & Mohr, 2015). Los microARNs regulan la expresión guiando un complejo de silenciamiento inducido por ARN llamado RISC, hacia la región 3' no traducida de un mRNA blanco (Nakanishi, 2016). En el caso de los microARNs, la versión 22 de la base de datos identificada como MirBase registra cerca de 2600 microARNs maduros (Kozomara et al., 2019; Kozomara & Griffiths-Jones, 2014). Aunque no todos han sido validados experimentalmente (Alles et al., 2019), se debe agregar el hecho de que en muchos casos un mismo microARN puede regular muchos genes (Medina & Slack, 2008). Adicionalmente, experimentos clínicos han concluido que los microARNs están involucrados en el desarrollo de varias enfermedades como el cáncer y el HIV (Zhao et al., 2016). Al 2021 la base de datos miR2Disease tiene identificadas más de 150 enfermedades relacionadas con más de 300 microARNs (Jiang et al., 2009).

Investigaciones realizadas en los últimos años sugieren que la mayoría de los genes humanos pueden ser blanco de los microARNs (Flynt & Lai, 2008). Muchos microARNs además tienen como blanco FTs, los cuales a su vez regulan la expresión de los microARN, formando una intrincada red de regulación (Tsang et al., 2007). Por ejemplo, la proto-oncoproteína c-MYC simultáneamente regula la expresión del factor de transcripción E2F1 y de un grupo de seis microARNs, que reprimen la expresión de E2F1. Esta coregulación positiva de E2F1 y sus microARNs reguladores, permiten que la señalización en la célula esté muy controlada. Adicionalmente, los microARNs algunas veces autoregulan su propia expresión por medio de ciclos que incluyen factores de transcripción específicos (Y. Guo et al., 2016).

Las redes regulatorias de expresión génica en eucariotas tienen varias propiedades organizacionales: 1- son modulares, permitiendo que múltiples genes sean simultáneamente regulados a través de los mismos mecanismos, 2- genes individuales son a menudo regulados por múltiples factores de transcripción, que de manera combinatoria se unen a distintas regiones regulatorias, incluyendo los promotores de genes (Roy et al., 2013). La topología de redes regulatorias especifica entradas hacia el sistema regulatorio de cada gen involucrado y en los casos en donde el gen codifica un factor de transcripción, define su salida hacia genes blanco en la próxima capa de la red (de-Leon & Davidson, 2009). Las redes biológicas a menudo contienen pequeñas subredes de topología recurrente denominadas motivos, entre los que se encuentran los

de ciclos autoregulatorios y los de retroalimentación. Se ha demostrado que redes que tienen algunos de estos ciclos pueden presentar compensación génica y adaptarse a las fluctuaciones en el número de copias de genes (Bleris et al., 2011). Esta característica hace relevante el estudio de redes en donde intervienen los microARNs y factores de transcripción.

Los métodos actuales en biología de sistemas tratan de entender los sistemas biológicos, tales como células o tejidos y cómo sus funciones resultan de la interacción de sus componentes. Esto involucra el estudio de su funcionamiento como un todo, más que sus partes individuales (Najafi et al., 2014). Las tecnologías “ómicas” como la genómica, proteómica, transcriptómica y metabolómica, facilitan el análisis simultáneo de genes, transcritos, proteínas y metabolitos, generando enormes cantidades de datos. La posibilidad de acceder a estos datos hace necesario el desarrollo de plataformas bioinformáticas que puedan modelar redes biomoleculares (Oulas et al., 2017). El estudio de redes de regulación transcripcional o de interacción proteína-proteína permiten entender muchas funciones celulares (Padi & Quackenbush, 2015). Las redes regulatorias identificadas ya sea experimentalmente o mediante predicciones tienen típicamente miles de nodos y miles o hasta millones de arcos, por lo que el uso de herramientas informáticas se hace indispensable (Q. Song et al., 2017).

Los modelos matemáticos permiten a los investigadores estudiar cómo los procesos regulatorios complejos están conectados y cómo disrupciones en esos procesos pueden contribuir al desarrollo de una enfermedad (Fischer, 2008). El interés en el análisis de redes es interdisciplinario, alcanzando varias campos entre los que están las matemáticas, física, biología, ciencias de la computación y estadística (Emmert-Streib & Dehmer, 2011). Las redes se describen en forma de grafos estáticos. Matemáticamente un grafo  $G = (N,E)$  se define como un par de conjuntos donde  $N$  es un conjunto de  $n$  vértices y  $E$  es un conjunto de  $m$  arcos, cada uno de los cuales conecta dos nodos (vértices). Algunas propiedades de los grafos, tales como el grado de un nodo, si es bidireccional o en una sola dirección, los ciclos, las rutas y otras más, son comúnmente usadas cuando se trabaja con redes biológicas (Najafi et al., 2014). En el contexto de una célula o tejido, redes particulares pueden corresponder a varios procesos similares, rutas metabólicas y funciones, en los cuales sus componentes (nodos) exhiben un alto grado de interacción, en comparación con otros que no están incluidos en el proceso (Rahiminejad et al., 2019).

El impacto de una infección de *H. pylori* en las rutas regulatorias de la mucosa gástrica (Carbo et al., 2013) e investigaciones de inmunología como lo es la activación de las células T (Eftimie et al., 2016), son algunos ejemplos donde se han utilizado los modelos matemáticos. En Costa Rica, la tesis titulada "Plataforma de biología de sistemas para la predicción de microARNs involucrados en redes de la regulación de dosis génica en cáncer aneuploide" también utilizó modelos matemáticos. En esta tesis se demostró la utilidad de tener una plataforma informática para analizar redes de regulación involucrando microARNs, genes y factores de transcripción. Se estudió una característica de las aneuploidías en cáncer, que consiste en que el número de cromosomas no corresponde al esperado. En el caso de los humanos lo normal es tener 23 pares de cromosomas, pero en casos de cáncer es común encontrar tres o más copias de un mismo cromosoma (Acón, 2018).

Usualmente las aneuploidías en autosomas no hacen viable la vida por periodos prolongados, debido a que interfieren con el crecimiento y desarrollo de un organismo (Torres et al., 2008). El síndrome de Down es una excepción pues es una trisomía en el cromosoma 21 y las personas con este síndrome pueden vivir hasta más de 80 años en condiciones ambientales adecuadas (Karmiloff-Smith et al., 2016). Más de un 90% de los tumores sólidos y de un 75% de los tumores hematológicos presentan aneuploidías (Sheltzer & Amon, 2011). La hiperdiploidía (un mayor número de cromosomas que lo normal) se presenta más a menudo en las células cancerosas que la hipodiploidía (un menor número de cromosomas que lo normal) (Kojima & Cimini, 2019). Si en términos generales las aneuploidías suelen ser letales, surge la pregunta de cómo es que las células cancerosas logran sobrevivir teniendo múltiples copias de un mismo cromosoma. Algún tipo de mecanismo parece regular la expresión génica en estos casos.

En la tesis mencionada previamente, se planteó la hipótesis de que el control de la red de regulación que media el mecanismo de compensación de dosis génica en células tumorales aneuploides está localizado en pocos microARNs, cuya identificación permitiría plantear el desarrollo de una terapia contra el cáncer (Acón, 2018). En la investigación se utilizó el panel NCI-60, que es un conjunto de 60 líneas celulares de nueve distintos tipos de cáncer: pulmón, colon, mama, melanoma, renal, próstata, ovario, sistema nervioso central y leucemia. El panel NCI-60 son las líneas tumorales más estudiadas en la investigación del cáncer, proveyendo una plataforma única

para plantear hipótesis enfocadas a entender la biología del cáncer (Abaan et al., 2013). Descubrimientos a partir del panel NCI-60 han permitido el desarrollo de medicamentos aprobados en Estados Unidos por la Administración de Alimentos y Medicamentos (FDA, por sus siglas en inglés) para el tratamiento del cáncer de colon, tal es el caso del medicamento oxaliplatin (T. Guo et al., 2019).

La investigación previa determinó *in silico* que existe una red de regulación alrededor de los genes MYC, STAT3, STAT5A y STAT5B con 18 microARNs y 41 factores de transcripción, entre los que se destacan MYC y STAT3. Los análisis realizados en la herramienta COPASI demostraron que esta red presenta el comportamiento de compensación de dosis génica, estando la mayoría del control concentrado en tres microARNs (mir17, mir19A y mir20A). Perturbaciones en la red mostraron que la inhibición de estos tres microARNs generan un alto incremento de más de 45 veces en las concentraciones de MYC, de casi 8 veces en STAT3, así como más de 50 veces en la concentración total del modelo, por lo que el efecto de esos microARNs es mantener la expresión de los genes en niveles bajos (Acón, 2018). MYC es un onco-gen que forma parte de muchas rutas metabólicas de transducción de señales y es un gen de respuesta temprana en muchos complejos de receptores (Dang, 2012). Por su parte STAT3 es un gen muy importante en la transducción de señales de las tirosinas quinasas que se activan en las células cancerosas, además de un factor de transcripción que regula la expresión de numerosos genes que contribuyen al progreso de tumores (Lee et al., 2019).

Dado que los niveles de aumento en las concentraciones de MYC y STAT3 después de inhibir los microARNs mir17, mir19A y mir20A fueron muy altos, se realizaron incrementos lineales en el número de copias de MYC y de STAT3 para estudiar los efectos en la expresión de dichos genes, y se observó que no se produjo un incremento lineal en la expresión de estos genes y por el contrario se observó un incremento en las concentraciones de los tres microARNs blanco, esto confirma la existencia de un mecanismo de compensación de dosis génica (Acón, 2018). Otro estudio más reciente también confirma la existencia de un mecanismo de compensación de dosis génica en células cancerosas (Brennan et al., 2019). Lo descubierto evidencia la utilidad de contar con una plataforma bioinformática general y amigable para construir modelos de regulación y datos de expresión del panel NCI-60.

Usando como base la plataforma bioinformática desarrollada en la tesis mencionada, se desarrolló una nueva plataforma (BioNetUCR), incorporando nuevas fuentes de datos y mejoras funcionales significativas. Aun estando en proceso de desarrollo, la misma fue utilizada por otros estudiantes de la maestría para sus respectivas tesis. La facilidad de uso se demostró con el hecho de que estos estudiantes pudieron usarla guiándose con una breve documentación. BioNetUCR permite a investigadores sin conocimientos avanzados en informática estudiar de manera ágil redes de regulación no solo a partir del panel NCI-60, sino también a partir de otras fuentes de datos como The Cancer Cell Genome Atlas (TCGA) (Tomczak et al., 2015) o Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012). Estas otras bases de datos no se incluyen en la plataforma, pero quienes trabajan con ellas pueden incorporarlas realizando cambios en la configuración de BioNetUCR. Los comandos y opciones de menú incluidos proveen flexibilidad al usuario para incluir estas u otras nuevas fuentes de datos, así como para editar redes, visualizarlas y generar modelos matemáticos en COPASI utilizando datos experimentales.

## 5. Problema

### 5.1. Justificación

Las redes de regulación pueden ser muy complejas y sin las herramientas adecuadas su análisis puede resultar inviable. Es imposible para una persona analizar de manera efectiva las distintas interacciones presentes en cientos de miles de datos de experimentos. Actualmente hay muchos repositorios desde los que se puede acceder a estas bases de datos para su análisis y generación de modelos. De no contarse con herramientas especializadas, resulta muy probable que se omitan elementos importantes en la red. Esto debido a la dificultad de determinar interacciones indirectas entre dos moléculas, como es el caso de ciclos o relaciones que no se dan de manera directa entre ellas.

La Universidad de Costa Rica ya comenzó a dar sus primeros pasos en el modelaje de sistemas biológicos usando conceptos matemáticos. Parte de este esfuerzo incluye el desarrollo de aplicaciones informáticas dirigidas a apoyar al investigador en la creación y análisis de modelos, pero es un área muy amplia y hay mucho camino por recorrer. Un trabajo de tesis previo sirve como punto de partida para este nuevo proyecto y fue la base para una publicación realizada en CONCAPAN 2016 (Acón et al., 2016). Sin embargo, este no es el único en Costa Rica que involucra modelos de regulación relacionados con genes, microARNs y factores de transcripción. Hay otros que fueron publicados en IWOBI 2018, desarrolladas por estudiantes de la maestría de Bioinformática y Biología de Sistemas. También hay otros trabajos de tesis en marcha que utilizan modelos matemáticos. Esto refleja una oportunidad para publicaciones futuras de otros proyectos de investigación. Algunos de los proyectos publicados son:

1. A transcriptional regulatory network model reveals miR-34a as a potential regulator of proliferation in tumor cell lines.
2. A Systems Biology approach to identify candidate targets to downregulate angiogenic gene expression in Cancer.

### 3. A Systems Biology Approach to Investigate Control Targets of Intracellular Calcium Transients induced by *Clostridium perfringens* Phospholipase C.

Los trabajos publicados en IWOBI 2018 tuvieron un elemento en común: a partir de una lista de genes, el autor de la tesis mencionada construyó la red y el modelo matemático usando la plataforma bioinformática desarrollada en su investigación. La construcción automatizada del modelo les permitió a estos estudiantes avanzar mucho más rápido en comparación con los estudiantes de los cursos de biología de sistemas de una generación atrás. Estos al no tener la plataforma tuvieron que construir las redes y los modelos de sus proyectos manualmente, logrando en el mejor de los casos ajustar el modelo con los datos experimentales y realizar algunos análisis, pero no les fue posible escribir un artículo con calidad publicable.

Sin embargo, la plataforma bioinformática desarrollada en (Acón, 2018) no fue conceptualizada para ser usada por otras personas, ni como una plataforma de uso general. La misma carece de funcionalidades para ese propósito, tal es el caso de poder hacer consultas de relaciones entre genes y transcritos o editar con facilidad la red en estudio, pues no eran los objetivos del proyecto. De lo discutido anteriormente, resulta clara la utilidad de tener una nueva plataforma mejorada, no solo para buscar nuevas inferencias del modelo de compensación de dosis génica, sino también para nuevas investigaciones utilizando modelos matemáticos. El que permita utilizar otras fuentes de datos además del panel NCI-60 y no exija que el investigador tenga conocimientos avanzados en informática, son características muy deseables en una nueva plataforma.

Las principales limitaciones que tiene la plataforma bioinformática del trabajo anterior son:

1. Se modela solamente un nivel de interacciones a partir de cada gen.

La red se construye tomando en cuenta únicamente las especies que interactúan directamente con alguno de los genes de interés, lo que podría dejar por fuera especies relevantes que están a más de una interacción de dichos genes.

2. Está enfocada en las especies que regulan un gen en particular, no considera el caso de las especies que son reguladas por un regulador de interés.

Esta limitación obliga al investigador a identificar los genes regulados por un regulador y proveer estos genes regulados a la plataforma.

3. No se considera el número de copias de los genes que codifican microARNs y FTs.

Por no tener disponibles los datos y para no tener demasiadas variables en el modelo, se utilizó el valor diploide en el número de copias de los genes codificantes de microARNs y FTs.

4. Incorporación y editaje de nuevas relaciones.

No es posible agregar de manera dinámica relaciones que el investigador considera que pueden aportar valor o remover algunas que considera son irrelevantes para el caso en estudio.

5. Consultas de interacciones entre genes, FTs y microARNs.

No es posible realizar consultas de si dos genes, FTs o microARNs están relacionados.

6. Se debe escoger entre interacciones validadas experimentalmente o putativas solamente.

No permite combinar interacciones validadas con interacciones putativas en un mismo modelo.

7. La interfaz y operación del programa depende de una persona con amplio conocimiento de la plataforma y experiencia en programación.

8. No posee un ambiente integrado diseñado para el usuario no informático.

9. Lenguaje de programación utilizado.

Fue escrita en el lenguaje Delphi, que actualmente no es tan usado como el lenguaje C#, por lo que darle mantenimiento futuro por parte de nuevos estudiantes resultaría



problemático, además, dicho lenguaje no posee un ambiente integrado de desarrollo que sea gratuito, como sí sucede con C#.

Por ser un área que ha tenido auge apenas en los últimos años, aún no hay plataformas bioinformáticas disponibles con un ambiente integrado desde donde se pueda partir de algunos genes de interés y llegar hasta modelar las redes con otros genes, microARN y FTs relacionados. Existen algunas iniciativas pero con limitaciones importantes, por ejemplo, hay un plugin llamado CyTransfinder para la plataforma de visualización de redes Cytoscape, que permite ir creando una red a partir de ciertos genes, pero es muy lenta, no tiene ningún tipo de integración con plataformas para ajustar modelos y posee pocas funcionalidades (Politano et al., 2016). Dado lo expuesto anteriormente, tener una plataforma amigable que permita a los investigadores no expertos en informática, modelar redes de regulación podría ser de mucho beneficio para futuras investigaciones. La posibilidad de crear, editar y visualizar las redes de interacciones entre genes, microARNs y FTs representa un gran ahorro de tiempo para el investigador y le permite enfocarse en el estudio del problema que investiga.

## 5.2. Hipótesis

El desarrollo de una plataforma bioinformática mejorada permitirá la identificación de nuevas inferencias del fenómeno de compensación de dosis génica a partir de información del panel NCI-60.

## 5.3. Objetivo general

Generar nuevos conocimientos sobre el fenómeno de compensación de dosis génica a partir de una plataforma bioinformática mejorada, desarrollada para el modelado de redes complejas de regulación génica por microARNs y factores de transcripción.

## 5.4. Objetivos específicos

5.4.1. Desarrollar un ambiente interactivo que permita la construcción de redes complejas de genes, microARNs y FTs, incorporando datos validados experimentalmente y putativos.

5.4.2. Incorporar en la plataforma la función de importar datos de expresión de ARN en el modelo para ser usados en la tarea de estimación de parámetros de COPASI, estudiando alternativas al método usado en (Acón, 2018), con el fin de que el proceso resulte más sencillo para el usuario.

5.4.3. Construir la topología de red y compararla con la desarrollada en (Acón, 2018).

Usando los genes MYC y STAT3 de la sección *Fase 3: modelos con ciclo sensor* de (Acón, 2018), construir la topología de la red con la plataforma desarrollada en dicha investigación y con la nueva plataforma, para verificar que ambas redes contienen los mismos nodos y arcos.

5.4.4. Usar las nuevas funciones de la plataforma para construir un modelo mejorado, incorporando el número de copias de microARNs y FTs.

5.4.5. Analizar el modelo construido para inferir nuevos conocimientos sobre el fenómeno de compensación de dosis génica.

## 6. Métodos

### 6.1. Desarrollo de la plataforma bioinformática BioNetUCR

Se utilizó el lenguaje C# y Visual Studio versión 2019 para el desarrollo. Para la visualización de redes desde C# se evaluaron distintas herramientas: GraphViz, Microsoft Automatic Graph Layout y YWorks, pero las dos primeras presentaron algunos problemas técnicos para los fines deseados y YWorks es de licenciamiento comercial de alto costo. Debido a esto se evaluaron lenguajes y herramientas alternativas a lo planteado inicialmente (C#), entre ellos Electron y Node.js. Posteriormente se determinó que es posible comunicar aplicaciones desarrolladas en C# con la poderosa y popular herramienta Cytoscape, mediante el estilo de arquitectura REST. Se continuó entonces con el plan original de desarrollar la herramienta en el lenguaje C#, con Cytoscape como visualizador de las redes.

Para el almacenamiento de las bases de datos se analizaron Neo4J y ArangoDB, ambas herramientas orientadas al manejo de grafos. Sin embargo, se descartaron por no tener un API robusto para la integración con aplicaciones desarrolladas en C# o por demandar un tiempo significativamente mayor en cada llamado al API, en comparación con llamados a código desarrollado directamente en C#. Adicionalmente, el proceso de instalación para el usuario final no informático resulta significativamente más sencillo si no se requieren instalar administradores de bases de datos de terceros. Por lo anterior, el manejo de bases de datos se desarrolló como parte de la plataforma, creando los respectivos algoritmos para la manipulación de los datos en memoria.

Para la generación del modelo COPASI se utilizó el API que éste ofrece por medio de DLLs. El API permite entre otras cosas crear especies, parámetros, funciones e incorporar datos experimentales. Su uso ofrece una mayor compatibilidad con futuras versiones de COPASI, que el esquema que se utilizó en la plataforma anterior, en donde se generaban directamente los archivos de texto con cada modelo. Seguidamente se muestra un diagrama general de la interacción de BioNetUCR con Cytoscape y COPASI.

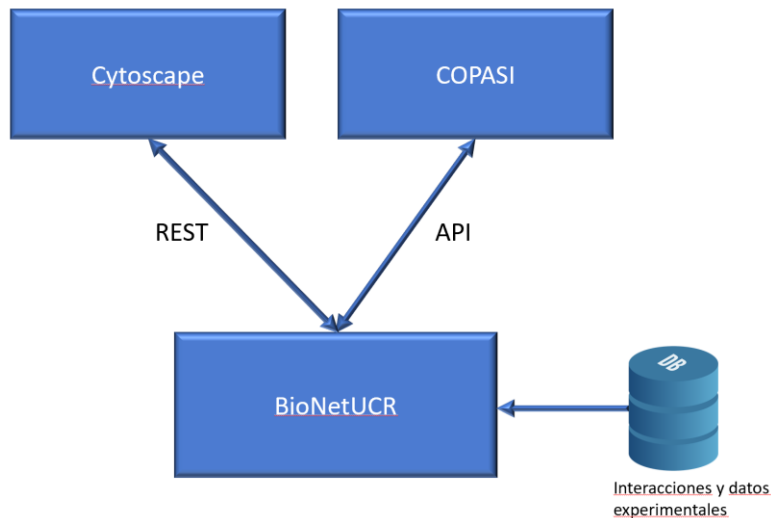


Figura 1: Diagrama de la plataforma desarrollada y la interacción con otras aplicaciones.

En vista de que cada nodo tiene usualmente pocos arcos en comparación con la cantidad total de nodos, el almacenamiento de los grafos en una matriz tradicional no es adecuado, por lo cual se utilizó una matriz dispersa. Se usa como estructura principal dos vectores, uno en donde cada elemento tiene el nodo origen y destino con el tipo de regulación que el origen ejerce sobre el destino y otro con índices a ese vector para mayor velocidad de acceso. Se creó una clase para representar un grafo. Esta se utiliza para representar la red diseñada por el usuario para el experimento en estudio y para la red principal con todas las relaciones entre los genes, microARNs y factores de transcripción. Esta red fue construida a partir de bases de datos especializadas.

#### 6.1.1. Componentes visuales de la plataforma bioinformática

BioNetUCR cuenta con cuatro áreas principales que permiten al usuario interactuar con ella. Las principales funcionalidades son:

1. Línea de ejecución de comandos. Permite ejecutar cualquiera de los comandos descritos en la sección de anexos.
2. Ejecución de bloques de comandos (scripts). Desde esta sección se puede ejecutar un grupo de comandos de manera secuencial, por ejemplo, cargar una red, agregarle relaciones a esa

red y almacenarla con otro nombre. Esta área es para desarrollar scripts que pueden ser almacenados, cargados nuevamente, editados y ejecutados. De esta forma se podrán reutilizar en nuevos proyectos de investigación scripts que se hayan desarrollado previamente.

3. Área con estadísticas. Aquí se presentan estadísticas de la red del usuario, específicamente se indican cuántos nodos y arcos hay, así como cuántos de esos nodos son de tipo gen, FT o microARN.
4. Sección de resultados. Mantiene un histórico de los comandos que se van ejecutando y los resultados de cada comando, en caso de que los haya.
5. Visualización de la red. Si se ejecuta el comando VIEW NETWORK, se muestra una imagen de la red de regulación, pudiéndose identificar cuáles nodos representan genes, factores de transcripción o microARNs, así como las relaciones entre ellos. Esta opción requiere que Cytoscape se esté ejecutando y es útil para tener una visión general de la red. Para analizar visualmente la red en más detalle, se debe utilizar la opción del Menú: Tools-View in Cytoscape, que permitirá visualizar y manipular la red directamente en Cytoscape.
6. Opciones de archivos y preferencias en menú principal:

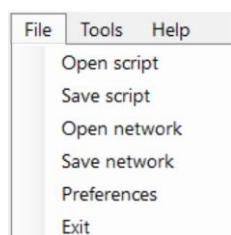


Figura 2: Opciones de Archivos y Preferencias en BioNetUCR

- *Open Script*: carga un script con comandos soportados por la plataforma. Los archivos con scripts tienen formato texto plano, conteniendo un comando por línea. La extensión por defecto es btn.
- *Save Script*: guarda un script con comandos soportados por la plataforma.

- *Open network*: carga una red previamente almacenada. Los archivos con redes tienen formato texto plano, conteniendo una relación por línea. Usan la extensión tab, pues los términos en cada línea se separan por un tabulador.
- *Save Network*: guarda en un archivo la red que se tenga en la plataforma.
- *Preferences*: permite definir cuál es el archivo con las interacciones entre genes, FTs y microARNs y el archivo con los datos experimentales. Adicionalmente tiene una prevista para uso futuro, en donde se podrá indicar que los datos experimentales están en un servicio REST.
- *Exit*: termina la ejecución de la plataforma.

## 7. Opciones de herramientas en menú principal:

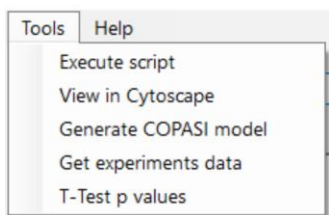


Figura 3: Opción de Herramientas en BioNetUCR

- *Execute script*: ejecuta los comandos que estén en la ventana de bloques de comandos.
- *View in Cytoscape*: exporta la red actual a Cytoscape para su visualización.
- *Generate COPASI Model*: genera el modelo en COPASI utilizando la red actual y los datos experimentales existentes en el archivo *Experiments.tab*.
- *Get experiments data*: extrae del archivo de datos experimentales indicado en *Preferences*, los datos que corresponden a los nodos de la red y los guarda en el archivo *Experiments.tab*. Esto con el fin de no tener datos experimentales innecesarios en el modelo.
- *T-test p values*: aplica el t-test para medias pareadas a un archivo que genera COPASI con los valores experimentales y estimados para cada especie, dando como resultado los respectivos valores de p. La prueba t-test, también conocida como t de Student, es una prueba estadística ampliamente usada para comparar la media de dos grupos de datos, se puede usar tanto para medias entre grupos independientes como para grupos

que son dependientes entre sí (Kim, 2015). Para que COPASI genere este archivo, se debe habilitar en la opción Output Assistant el gráfico Parameter Estimation Results y una vez generado el gráfico tras finalizar el ajuste, se debe usar la opción Save Data e indicar el nombre de archivo donde se almacenarán los datos a ser usados en la opción t-test p values.

### 6.1.2. Comandos soportados

Se implementaron 23 comandos en BioNetUCR que pueden ser ejecutados de forma individual desde la línea de ejecución de comandos o en grupos mediante scripts. Los scripts se almacenan como archivos de texto y pueden editarse con editores convencionales para luego ser cargados y ejecutados en BioNetUCR. Se implementó la funcionalidad de autocompletar los comandos de modo que con solo escribir los primeros caracteres aparezcan los comandos que coinciden, lo cual facilita su escritura. Los comandos son principalmente de tres tipos y una lista completa con su descripción se encuentra en la sección de anexos:

1. Informativos: Se trata de comandos que muestran información, por ejemplo, LIST BY SOURCE que despliega los arcos de la red ordenados por el origen.
2. Modifican la red: Son comandos que agregan o remueven elementos de la red, como el comando DELETE NODE *node* que borra un nodo de la red.
3. De configuración: Estos permiten definir aspectos como los genes de interés al generar el modelo en COPASI o el nombre del archivo donde almacenar los datos experimentales del modelo.

### 6.1.3. Características no funcionales

BioNetUCR tiene bajos requerimientos de CPU y memoria, pudiéndose ejecutar en una PC con el sistema operativo Windows 10, con procesador i3 y 4GB de RAM, sin embargo, dependiendo del tamaño de la red y en especial para los procesos de generación del modelo y exportación a COPASI, sí es recomendable una PC con mayor CPU y memoria.

#### 6.1.4. Relaciones soportadas por BioNetUCR

FT → microARN: un factor de transcripción regula un microARN.

FT → gen: un factor de transcripción regula un gen.

FT → FT: un factor de transcripción regula otro factor de transcripción.

microARN → FT: un microARN regula un factor de transcripción.

microARN → gen: un microARN regula un gen.

#### 6.2. Incorporación de bases de datos a la plataforma desarrollada

Una vez desarrollada la plataforma bioinformática, su funcionamiento fue validado inicialmente con datos de prueba y luego se procedió a incorporar las mismas bases de datos utilizadas en (Acón, 2018). Adicionalmente se descargaron los datos de número de copias de todos los genes, microARN y FTs disponibles en CellMiner (Reinhold et al., 2012) y se extrajeron todos los que están presentes en la base de datos de interacciones de (Acón, 2018). Posteriormente con estos datos se actualizaron los valores de número de copias que estaban con un valor diploide.

#### 6.3. Construcción de la topología de la red y comparación con la de (Acón, 2018)

Utilizando los comandos de BioNetUCR mostrados en la sección de anexos se construyó la topología de la red del modelo mínimo creado en (Acón, 2018). Posteriormente se exportó a Cytoscape con la opción de menú “View in Cytoscape” y se le aplicó el estilo “Degree Sorted Circle Layout” y se ajustaron los colores de los nodos para facilitar la comparación. Con la opción “Get experiments data” se extrajeron los datos experimentales para los nodos de la red y utilizando la opción “Generate COPASI model” se generó el modelo en COPASI para su comparación con el generado en (Acón, 2018). Se verificó que las especies, parámetros, reacciones y funciones fueran equivalentes entre ambos modelos, esto se hizo utilizando la información que COPASI provee y herramientas de uso general como Excel, Notepad++ y WinMerge. Para una comparación de un modelo más complejo se generó en ambas plataformas un modelo con los genes de interés



MYC y STAT3 y se validó su equivalencia en COPASI de la misma forma que con el modelo mínimo.

#### 6.4. Construcción del modelo mejorado de biología de sistemas

Una vez validada la topología de la red se generó un modelo matemático a partir de esta. Se utilizaron las mismas especies, relaciones y tipo de reacciones que en (Acón, 2018), así como ecuaciones diferenciales ordinarias, pero se incorporaron los valores de las copias de genes que codifican los FTs y microARNs disponibles en CellMiner (Reinhold et al., 2012). Mediante las facilidades de estimación de parámetros de COPASI, se ajustó el modelo generado.

#### 6.5. Validación *in silico* de la hipótesis e inferencia de nuevos conocimientos

##### 6.5.1. Modelo mínimo incorporando el número de copias para todas las especies

Se procedió a ajustar el modelo mínimo descrito en el punto 6.4, utilizando el método NL2SOL y los resultados obtenidos fueron similares al ajuste del modelo mínimo original que utiliza valores diploides para los microARNs. NL2SOL es un nuevo algoritmo incorporado en COPASI, logrando ajustes para los casos en estudio en menor tiempo que los demás algoritmos ofrecidos por la aplicación. Posteriormente se hizo un Parameter Scan del número de copias de MYC y STAT3 para comparar los resultados con los del modelo mínimo original.

##### 6.5.2. Modelo mínimo extendido hacia genes del metabolismo

Para validar el nivel de funcionalidad que ofrece BioNetUCR y el tipo de redes que se pueden estudiar con su uso, se utilizaron distintos comandos para generar un nuevo modelo extendido a partir del mínimo. Posteriormente se filtró la red generada para dejar únicamente los genes de metabolismo (Noronha et al., 2019) que son regulados por los elementos del modelo mínimo. El modelo generado tiene 439 especies y se ajustó en COPASI tras casi 3 semanas de ejecución en un equipo con procesador i7 y 16GB de memoria. Posteriormente se determinaron los genes del metabolismo regulados por al menos tres especies del modelo mínimo.

### 6.5.3. Compensación de dosis génica de MYC y STAT3 con datos de CCLE

Aunque inicialmente se pretendía trabajar solo con los datos de NCI60, dada la posibilidad de BioNetUCR de utilizar otras fuentes de datos, se evaluó el alcance de la compensación de dosis génica de MYC y STAT3 en diferentes tipos de cáncer con un conjunto de datos más grande. Se aplicaron los mismos criterios usados en la investigación con NCI60 para identificar genes candidatos bajo compensación de dosis para 898 líneas celulares de Cancer Cell Line Encyclopedia (CCLE), para los que se tienen datos de expresión y de número de copias. Los datos de expresión se normalizaron con respecto al promedio de las líneas celulares diploides para cada gen, y tanto la expresión normalizada como el número absoluto de copias se transformaron a logaritmo en base 2 y se centraron en 0 (diploidía) para una comparación directa. Asimismo, se calcularon las desviaciones estándar de ambos datos transformados a logaritmo en base 2 para buscar grupos de genes candidatos putativos bajo compensación de dosis génica. Posteriormente se utilizó BioNetUCR para generar distintos modelos con todos los datos de CCLE, así como con datos individuales de tipos de cáncer específicos.

## 7. Resultados

La *Figura 4* muestra las distintas secciones de BioNetUCR. El comando *View Network* mostrado en la sección 1 hace que BioNetUCR utilice las funcionalidades de Cytoscape para graficar la red actual y mostrarla en la sección 5. Los dos primeros comandos en la sección 2 generan la red de nodos que regulan MYC y STAT3 y todas las interrelaciones entre ellos. Los demás comandos permiten que el modelo generado en COPASI a partir de esta red sea compatible con el generado con la plataforma anterior, en donde varias reglas están implementadas a nivel de código y por lo tanto son fijas. La sección 3 muestra estadísticas de la red, entre ellas el total de nodos y arcos. Los comandos ejecutados y resultados de estos se muestran en la sección 4.

The screenshot displays the BioNetUCR interface with the following sections:

- Section 1:** A text input field containing the command `VIEW NETWORK`.
- Section 2:** A list of commands in a script, including `ADD BACKWARD MYC`, `ADD BACKWARD STAT3`, `DELETE AUTOREGULATED EDGES`, `TAG MYC`, `TAG STAT3`, `DELETE SOURCE NODES`, `DELETE EDGE E2F1 MIR449C`, `DELETE EDGE ETS1 HOXB4`, `DELETE EDGE ETS1 NME2`, `DFI FTF FVVF QAT29 HVVRA`, and `VIEW NETWORK`.
- Section 3:** Network statistics showing: Total Nodes of type gen: 0, Total Nodes of type TF: 42, Total Nodes of type microRNA: 17, Total Nodes: 59, and Total Edges: 391.
- Section 4:** A log of execution results, listing the same commands as in Section 2 with the prefix "Command executed:".
- Section 5:** A network graph visualization showing numerous nodes (represented by yellow circles) and a dense web of edges (represented by grey lines) connecting them.

*Figura 4: Secciones de BioNetUCR.*

La plataforma previa fue construida para ayudar a determinar la existencia de un mecanismo de compensación de dosis génica, por lo que la generalidad y facilidad de uso no fueron considerados. Dada la utilidad mostrada por dicha plataforma, BioNetUCR se desarrolló incorporando los

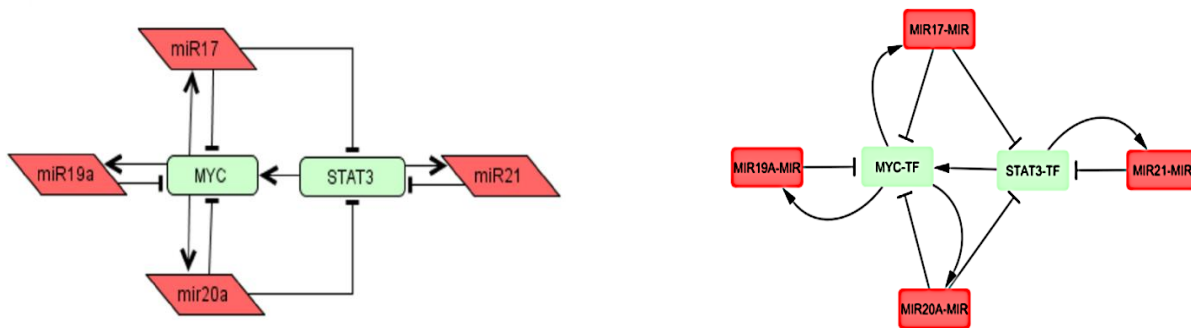
mismos beneficios, pero extendiéndolos de forma significativa. La *Tabla 1* muestra las principales características que presenta la BioNetUCR en comparación con la plataforma previa.

<b>Características</b>	<b>Plataforma previa</b>	<b>BioNetUCR</b>
No requiere conocimientos avanzados en informática.		✓
Comandos para crear y editar redes de regulación génica		✓
Ejecución de scripts de comandos		✓
Integración con herramienta de graficación de redes		✓
Integración con API de COPASI		✓
Funciones para almacenamiento y lectura de archivos de redes y scripts		✓
Generación de modelos en formato COPASI	✓	✓
Desarrollado en lenguaje de programación popular en universidades		✓
Permite agregar a la red microARNs y FTs que regulan un gen en particular	✓	✓
Permite agregar a la red genes que son regulados por un microARN o FT particular		✓
Incorpora número de copias tanto para microARNs como para FTs		✓
Cálculo de t-test		✓
Ayuda integrada		✓

*Tabla 1 Comparativo de características de la plataforma previa y BioNetUCR*

En la *Figura 5* se aprecia gráficamente que la red correspondiente al modelo mínimo generada por BioNetUCR y graficada en Cytoscape (imagen derecha), corresponde a la misma red construida en (Acón, 2018). Al exportar cualquier red desde BioNetUCR hacia Cytoscape por defecto se usa el *Circular Layout*, sin embargo, una vez en Cytoscape es posible cambiar el layout o los estilos

de nodos y arcos. Para facilitar la comparación en la *Figura 5* se usó el *Degree Sorted Circle Layout* y posteriormente se cambiaron los colores y estilos para tener una red fácilmente comparable con la creada por la plataforma anterior. Resulta claro que en ambos casos los nodos y arcos, así como el tipo de relaciones entre las especies son equivalentes.



*Figura 5: Red generada en investigación previa y la generada por BioNetUCR (derecha).*

El modelo mínimo generado en COPASI con BioNetUCR también resulta equivalente al desarrollado con la anterior plataforma. Las *Figura 6, 7 y 8* incluyen imágenes tomadas de COPASI con ambos modelos, mostrando que los resultados son los mismos salvo diferencias en nomenclatura de nombres. En la *Figura 6* ambos casos tienen las mismas 6 especies y las mismas 12 reacciones de degradación y síntesis para cada una de ellas. La diferencia en Global Quantities de 24 versus 26 corresponde a que BioNetUCR realiza una optimización de parámetros dejando solo los que se utilizan en el modelo. Para el caso específico del modelo mínimo, las dos que no se incluyen son los parámetros de represión para MYC y STAT3, pues estos dos factores de transcripción actúan como activadores. Finalmente, la cantidad de funciones también es la misma y la equivalencia fue verificada extrayendo el detalle de cada una y realizando la comparación con herramientas externas.

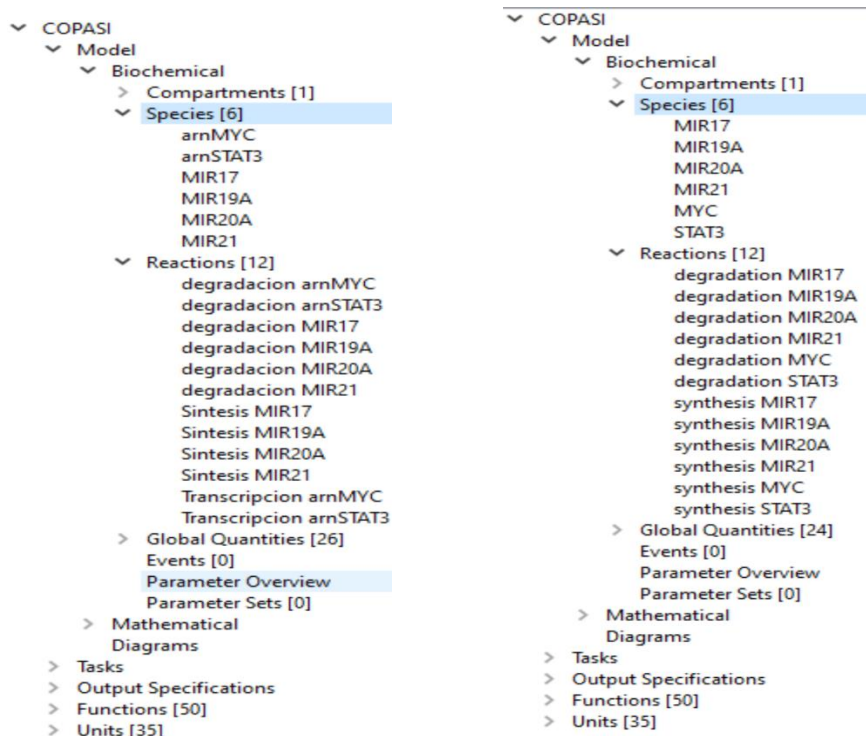


Figura 6: Modelo mínimo en COPASI, generado en la plataforma previa y en BioNetUCR (derecha)

La Figura 7 detalla a manera de ejemplo las reacciones de degradación para MYC en ambas plataformas y la Figura 8 las de síntesis para MIR17. Los parámetros que inician con *ks* representan síntesis, *kr* represión, *kd* degradación y *ka* activación. Los “modifiers” corresponden a especies que tienen algún tipo de regulación sobre el sustrato.

The figure shows two screenshots of the COPASI software interface, displaying the details of a reaction. The left screenshot shows the reaction details in the previous platform, and the right screenshot shows the same reaction details in BioNetUCR.

**Left Screenshot (Previous Platform):**

Reaction: degradacion arnMYC

Reaction:  $\text{arnMYC} \rightarrow \text{MIR17 MIR20A MIR19A}$

Rate Law: Function for degradacion arnMYC

Rate Law Unit:  Default  mol/s  mol/(m<sup>3</sup>s) default

Role	Name	Mapping
Parameter	KdArn_MYC	KdArn_MYC
Modifier	MIR17	MIR17
Modifier	MIR19A	MIR19A
Modifier	MIR20A	MIR20A
Substrate	arnMYC	arnMYC
Volume	default	default
Parameter	kr_MIR17	kr_MIR17
Parameter	kr_MIR19A	kr_MIR19A
Parameter	kr_MIR20A	kr_MIR20A

**Right Screenshot (BioNetUCR):**

Reaction: degradacion MYC

Reaction:  $\text{MYC} \rightarrow \text{MIR17 MIR20A MIR19A}$

Rate Law: Function for degradacion MYC

Rate Law Unit:  Default  nmol/s  nmol/(μl\*s) Compar

Role	Name	Mapping
Parameter	"kd_MYC"	kd_MYC
Substrate	"MYC"	MYC
Parameter	"kr_MIR17"	kr_MIR17
Modifier	"MIR17"	MIR17
Parameter	"kr_MIR20A"	kr_MIR20A
Modifier	"MIR20A"	MIR20A
Parameter	"kr_MIR19A"	kr_MIR19A
Modifier	"MIR19A"	MIR19A

Figura 7: Reacción de degradación para MYC en el modelo mínimo creado con la plataforma previa y BioNetUCR (derecha)

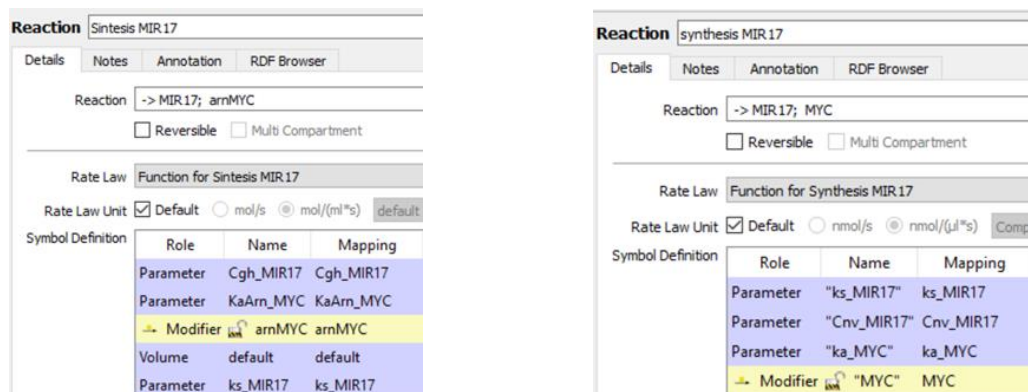


Figura 8: Reacción de síntesis para MIR17 en el modelo mínimo creado con la plataforma previa y BioNetUCR (derecha)

En la *Figura 9* se muestra la comparación de un modelo más complejo generado alrededor de los genes MYC y STAT3, cuenta con 59 especies y también resulta equivalente en ambas plataformas. Para generarlo en BioNetUCR se usaron los mismos comandos mostrados en la Tabla 2, pero agregando los comandos ADD BACKWARD STAT3 y TAG STAT3

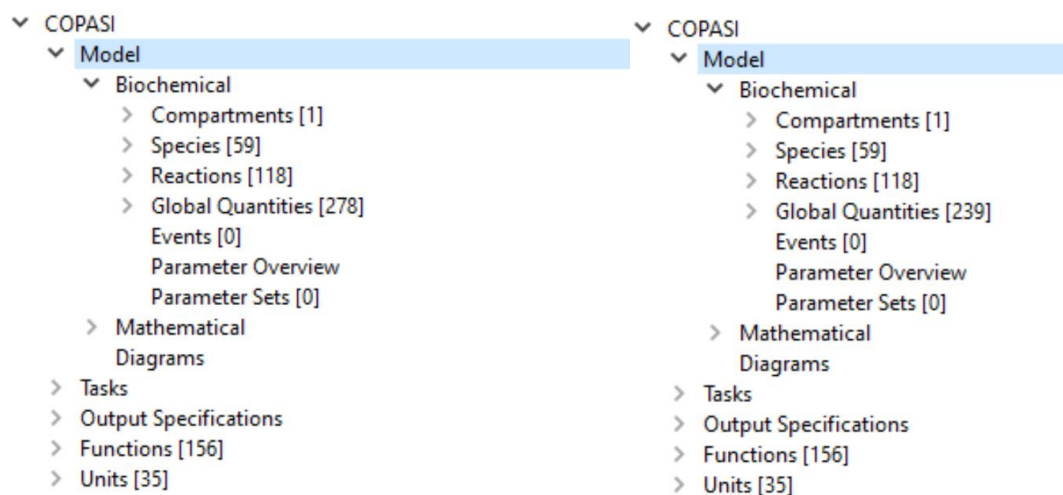


Figura 9: Modelo en COPASI para MYC y STAT3 generado en la plataforma previa y BioNetUCR (derecha)

La *Figura 10* muestra el resultado de realizar un Parameter Scan en COPASI para el número de copias tanto de MYC como de STAT3. Las gráficas de la izquierda corresponden al modelo mínimo de la plataforma anterior, usando el número de copias diploide para los microARNs y las de la derecha al nuevo modelo con números de copias tomados de CellMiner y redondeados al entero más cercano. Los cambios entre ambos casos son mínimos y se atribuyen más a efectos del redondeo, pues los valores de números de copias en CellMiner de los microARNs involucrados son precisamente el valor diploide. Para la única especie que cambia el número de copias es MYC, en el nuevo modelo tiene un valor de 2 y en el anterior de 1, sin embargo, esto no tiene mayor efecto en las proporciones del número de copias versus valores de expresión que se aprecian en ambas gráficas.

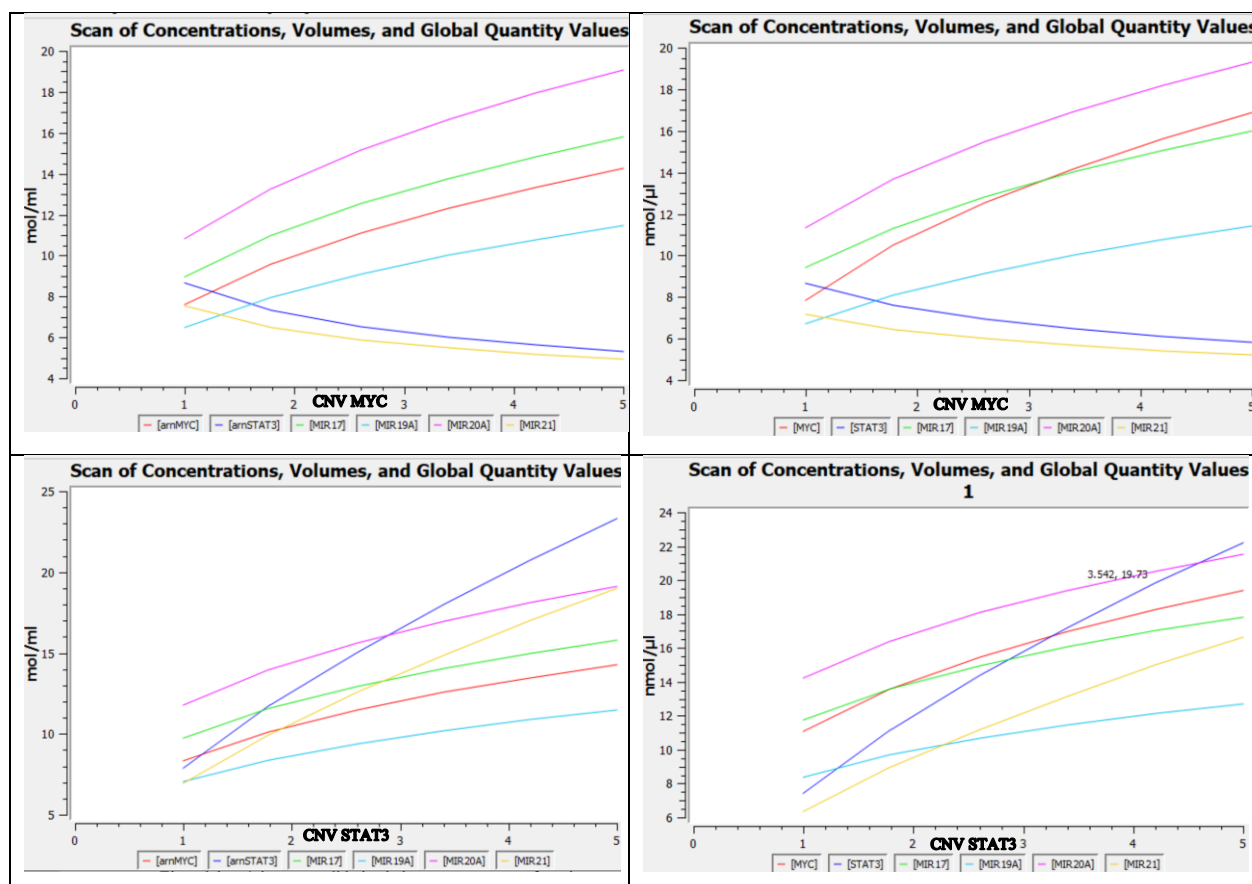
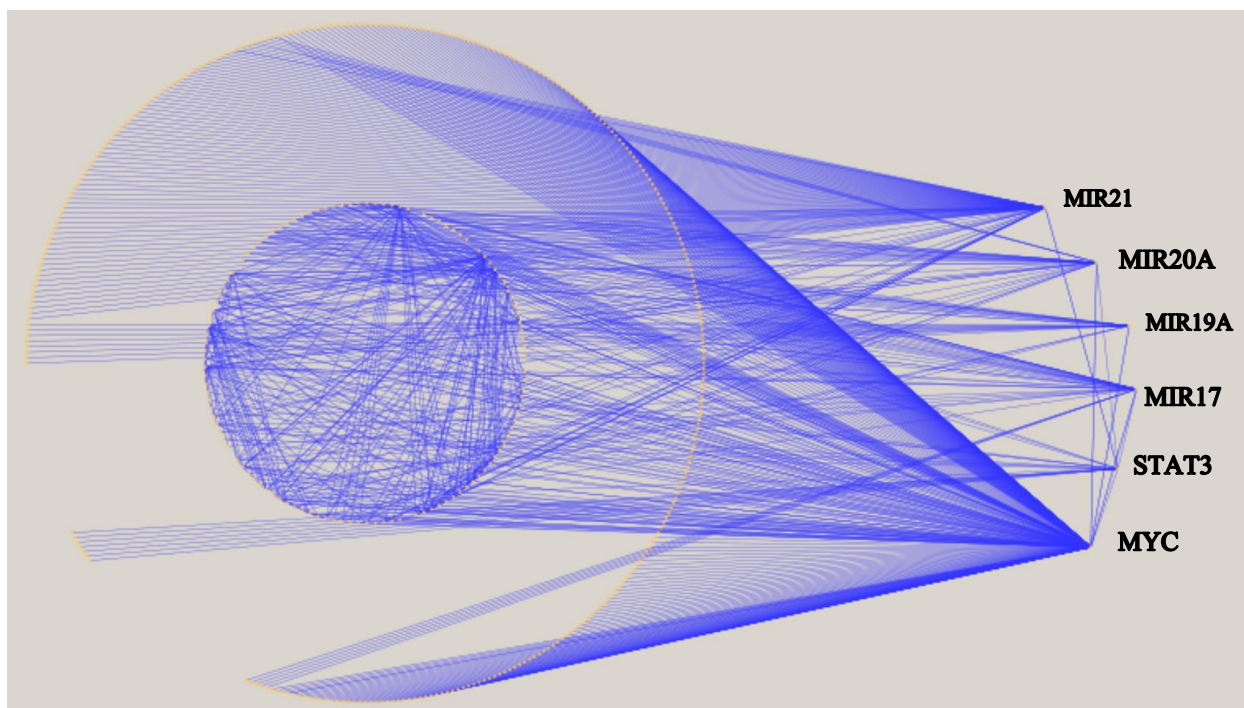


Figura 10: Comparativo de un parameter scan para el CNV de MYC y STAT3 en modelo con CNV diploide versus modelo con CNV de CellMiner (derecha)



**Relación del modelo mínimo con los genes del metabolismo:**

Al extender la red del modelo mínimo hacia los genes que son regulados por esta y posteriormente filtrarla con los genes del metabolismo, obtenemos una red con 816 arcos y 439 nodos, de los cuales 316 son genes, 98 factores de transcripción y 25 microARNs. La red resultante destacando las especies del modelo mínimo se muestran en la *Figura 11*. MYC regula 320 genes, siendo la especie que mayor influencia tiene sobre toda la red. MIR21 regula 96 genes, MIR17 regula 40, MIR20A 37, MIR19A 27 y STAT3 24. Los restantes genes que no son regulados directamente por las especies del modelo mínimo son regulados por otros factores de transcripción y microARNs que surgen de las interacciones al extender la red del modelo mínimo.



*Figura 11: Red de genes del metabolismo regulados por las especies del modelo mínimo*

Filtrando en Cytoscape la red de la *Figura 11* para dejar solo los genes del metabolismo que tienen 3 o más interacciones con los factores de transcripción y microARNs, resulta la red de la *Figura 12*. Se incluyen también los genes que son regulados por MYC y STAT3, aunque tengan menos de 3 interacciones, siendo HMOX1 el único que cumple con este criterio. Ninguno de los genes del metabolismo aparece regulado directamente por más de 5 especies de la red del modelo mínimo. Hay un gen regulado por cinco de estas especies, dos genes por cuatro especies y dos genes por tres. Su relación con las especies del modelo mínimo se muestran en la *Figura 12* y se ha identificado que están relacionados con el cáncer y varias rutas metabólicas importantes GeneCards (Safran et al., 2002):

#### PTEN

Regulado por MYC, MIR17, MIR19A, MIR20A y MIR21. Está localizado en la posición 10q23.31 y su pérdida está asociada con un incremento en la inestabilidad en el número de cromosomas (Vidotto et al., 2018). PTEN está identificado como un gen supresor de tumores que se encuentra mutado en una cantidad importante de tumores, entre ellos próstata, pulmón y mama (Safran et al., 2002). Regula de forma negativa la ruta metabólica PI3K, la cual es muy importante en casos de cáncer (Wikman et al., 2012). Esta regulación negativa es mediada por MYC que provoca una sobreexpresión de PTEN (Kaur & Cole, 2013). Por su parte la sobreexpresión de MIR17 directamente suprime el mRNA y la expresión de la proteína PTEN, pudiendo además incrementar la radioresistencia de los carcinomas nasofaríngeos a través de la ruta metabólica PTEN/AKT (Hu et al., 2019, p. 17). En el caso de MIR19A la sobreexpresión de este contribuye a la quimioresistencia de múltiples tipos de cáncer. Por ejemplo, al ser PTEN blanco de MIR19A, una reducción en la expresión de PTEN aumenta la resistencia al medicamento oxaliplatino en casos de cáncer colorectal (Y. Zhang et al., 2020). MIR20A actúa como un regulador negativo de PTEN y media la proliferación, migración y apoptosis de múltiples mielomas. La sobreexpresión de MIR20A promueve la actividad de la ruta metabólica PI3K/AKT (Gao et al., 2019). En el caso de MIR21, se ha encontrado que está elevado en cáncer de ovario y que la supresión de MIR21 aumenta la expresión de PTEN, inhibe la actividad de PI3K/AKT, promueve la apoptosis y reduce la proliferación celular (Liu et al., 2019, p. 21). Resulta entonces que MYC, MIR17, MIR19A, MIR20A, MIR21 y PTEN conforman una red reguladora que tiene una fuerte influencia sobre las rutas metabólicas PI3K-AKT, cuya alteración es común en distintos tipos de cáncer. Nuevos

proyectos de investigación podrían profundizar en el modelaje de esta red y tratar de determinar si alguno de los microARNs tiene mayor relevancia que los demás en casos de cáncer.

### DNMT1

Regulado por MIR17, MIR19A Y MIR20A. Está localizado en el cromosoma humano 19p13.2 y codifica una proteína que contiene 1632 aminoácidos, la cual puede estar implicada en la ocurrencia, progresión y pronóstico del cáncer (Li et al., 2017, p. 1). La metilación del ADN es la modificación epigenética más común en el genoma de los mamíferos, patrones de metilación erróneos pueden encontrarse en tumores humanos y en muchas anomalías. El gen DNMT1 es importante para mantener la metilación del ADN y participa en la oncogénesis causando hipermetilación o hipometilación (Ma et al., 2017, p. 1). Se ha determinado que en procesos de quimioresistencia en glioblastomas, DNMT1 está negativamente correlacionado con la expresión de MIR20A (Zhou et al., 2015, p. 1). Estudios previos que establezcan la relación entre MIR17, MIR19A y DNMT1 son muy escasos, en especial alrededor del cáncer, sin embargo, sí se ha determinado que la disminución en la expresión de DNMT1 mediada por MIR17 provoca metilación defectuosa en la enfermedad de Parkinson (H.-Q. Zhang et al., 2021).

### TGFBR2

Regulado por MIR17, MIR19A, MIR20A y MIR21. Está localizado en el cromosoma 3p (Velapasamy et al., 2018). TGFBR2 es un supresor de cáncer y un blanco terapéutico con potencial en el tratamiento del cáncer de cérvix (Yang et al., 2017). A este gen se le asocia con rutas metabólicas relacionadas con distintos tipos de cáncer, sin embargo, aún hay poca literatura en esta área y en particular en su relación con los microARNs indicados. Uno de los casos en donde sí se ha establecido una relación entre MIR20A y TGFBR2 es en la fibrosis del hígado, proceso patológico esencial que puede deteriorarse y convertirse en cirrosis y cáncer del hígado (Fu et al., 2020). En el caso de MIR21, análisis de cáncer de próstata mostraron que el incremento en la expresión de MIR21 junto con receptores de andrógeno (AR), reducen significativamente la expresión de TGFBR2 (Mishra et al., 2014).

### UBE2C

Regulado por MYC, MIR17 y MIR20A. Está localizado en el cromosoma 20q13 y la proteína expresada por este gen es casi indetectable en tejidos normales (R. Wang et al., 2017, p. 2). Está

sobreexpresado en más de 20 tipos de cáncer y su sobreexpresión está correlacionada con un bajo pronóstico de supervivencia, lo cual sugiere su involucramiento en el avance e invasión de tumores (Dastsooz et al., 2019). El incremento de UBE2C en cáncer es a menudo visto en niveles similares a los de MYC (Hao et al., 2012). Con respecto a MIR17 y MIR20A se ha demostrado que en casos de cáncer gástrico hay sobreexpresión de estos dos microARNs, así como su relación directa entre ellos y UBE2C (Y. Zhang et al., 2015).

Adicionalmente se encuentra el caso del gen HMOX1, localizado en el cromosoma 22q12 (Nitti et al., 2018). A diferencia de los anteriores no aparece regulado por ningún microARN, pero sí por MYC y STAT3. Es el único gen del metabolismo que se encuentra en la red que tiene estas características. La sobreexpresión de HMOX1 ha sido observada en varios tipos de cáncer, incluyendo tumores cerebrales, melanomas, leucemia mieloide crónica y linfoma, lo cual sugiere una posible contribución de HMOX1 al crecimiento tumoral por medio de angiogénesis, metástasis y proliferación (Tan et al., 2015). HMOX1 es un importante mediador de la activación de STAT3, la sobreexpresión de HMOX1 potenciada por la glucoproteína Interleucina 6 produce la activación de STAT3 y una disminución en la expresión de HMOX1 produce resultados opuestos (J. Song et al., 2019, p. 1). También está el gen BMPR2, al cual se le liga con el cáncer (S. Wang et al., 2017), sin embargo, su relación con los microARNs MIR17, MIR19A, MIR20A y MIR21 que lo regulan, no está claramente establecida.

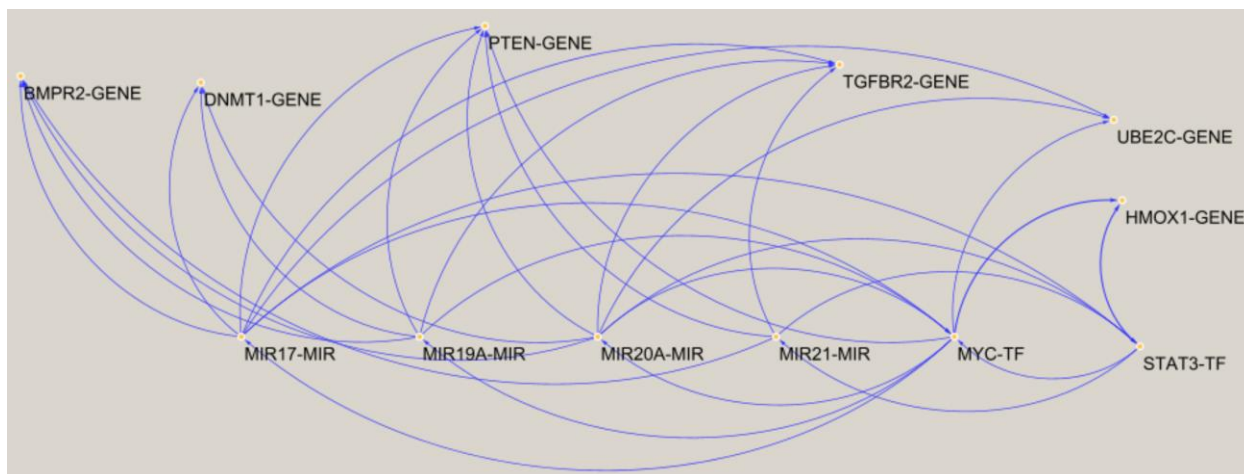


Figura 12: Red de genes del metabolismo con más interacciones con FTs y microARNs del modelo mínimo.

### **Estudio de la red mínima con datos de CCLE:**

Haciendo un análisis similar al realizado en la investigación anterior con NCI60, se estudiaron los datos de CCLE para evaluar compensación alrededor de la red mínima. Una revisión detallada de los datos individuales de MYC y STAT3 en CCLE muestra que MYC tiene muchas amplificaciones con un alto número de copias, todas ellas probablemente compensadas. Por el contrario, STAT3 tiene pocas amplificaciones compensadas de solo una copia adicional (*Figura 13*). También hay varios genes con mayor variación en CNV en comparación con MYC y entre ellos está TBC1D3B con una variación particularmente baja en expresión. Este gen codifica una proteína involucrada en la señalización de las proteínas RAB GTPase y el tráfico vesicular GeneCards (Safran et al., 2002), está ubicado en el cromosoma 7q, una ubicación cromosómica compartida con algunos de los candidatos obtenidos con NCI60, incluidos STAT3, STAT5A y STAT5B. Estos hallazgos confirmaron que MYC es uno de los genes compensados y amplificados con mayor frecuencia en CCLE y hay otros candidatos bajo una sólida compensación de dosis génica.

Para investigar si este comportamiento de aparente compensación de dosis de MYC y STAT3 podría explicarse por los mismos circuitos propuestos para los datos de NCI60, se utilizó BioNetUCR para estudiar estos circuitos en 3 niveles diferentes. Primero intentando ajustar un modelo matemático pan-CCLE usando todas las líneas celulares para ajustar el modelo mínimo de compensación de dosis de MYC y STAT3. Aunque no fue posible ajustar el modelo para las diferentes especies (excepto STAT3), se puede observar una tendencia cuando el conjunto de datos simulados se compara con los datos experimentales (*Figura 14*) y el modelo resultante repite el comportamiento de la compensación de la dosis de genes para ambos genes. (*Figura 15*). En segundo lugar, bajo la hipótesis de que la alta heterogeneidad de CCLE impide un ajuste correcto del modelo mínimo, se procedió a ajustar el modelo mínimo a conjuntos de datos específicos por tipo de cáncer. De esta forma sí se logró un ajuste completo para todas las especies de la mayoría de los modelos de compensación de dosis génica, excepto para el cáncer de pulmón y de mama (*Figura 16*, panel superior). Luego se utilizó la función parameter scan de COPASI en el número de copias de MYC/STAT3 para los distintos modelos por tipo de cáncer y se pudo observar una compensación de dosis de MYC generalizada en los tipos de tumores con cierto grado de

heterogeneidad, excepto para el modelo de cáncer de próstata que no mostró compensación de dosis génica. (*Figura 16*, panel central). Por el contrario, hubo una heterogeneidad mucho mayor en la compensación de la dosis génica de STAT3 entre los tipos de cáncer. Ocho modelos específicos de tipo de cáncer no pudieron del todo compensar la dosis de STAT3 (*Figura 16*, panel inferior).

Con el fin de caracterizar aún más esta heterogeneidad en la compensación de la dosis de genes, se utilizó el modelo como una herramienta para realizar la estimación de parámetros individuales de líneas celulares específicas. Con los modelos de un solo experimento, se realizaron exploraciones de parámetros para calcular el aumento en la expresión génica in silico en estado estable al aumentar el número de copias de genes de 1 a 5. Los valores se normalizaron a la expresión en  $CNV = 1$  para efectos de comparación. Se pudo observar que la mayoría de los modelos específicos de líneas celulares tienen la capacidad de compensación de dosis de MYC en sus configuraciones actuales (determinadas por sus valores de parámetros) con solo unos pocos valores atípicos, como lo muestran los aumentos en la expresión génica al aumentar CNV (por ejemplo, un aumento en un número de copias de 1 a 5 conduce a un aumento promedio de solo 2.3 veces en la expresión (*Figura 17*, panel izquierdo). Por el contrario, la compensación de la dosis de STAT3 es mucho más débil para la mayoría de los modelos específicos de la línea celular con un aumento promedio de 4 veces con un aumento de CNV de 1 a 5 (*Figura 17*, panel derecho). Se estableció un índice de compensación para la comparación, en donde un índice de compensación cercano a 1 indica una compensación muy débil y un índice de compensación cercano a 0 indica una fuerte compensación. Al trazar los índices de compensación de los modelos de las líneas celular particulares en los distintos tipos de cáncer, se observa que sus configuraciones de parámetros permiten una compensación significativa de MYC entre los tipos de cáncer, mientras que STAT3 está débilmente compensado (*Figura 18*). Estos resultados indican que la compensación de dosis génica de MYC está habilitada en la mayoría de los tipos de cáncer a pesar del alto nivel de heterogeneidad observado en el cáncer en general. Aunque esta heterogeneidad y la alta complejidad del conjunto de datos CCLE impide el ajuste de modelos matemáticos más grandes de compensación de dosis génica, representa una herramienta muy importante para validar modelos mínimos de regulación génica y cuantificar el grado de heterogeneidad en propiedades emergentes de sistemas biológicos como la compensación de dosis génica.

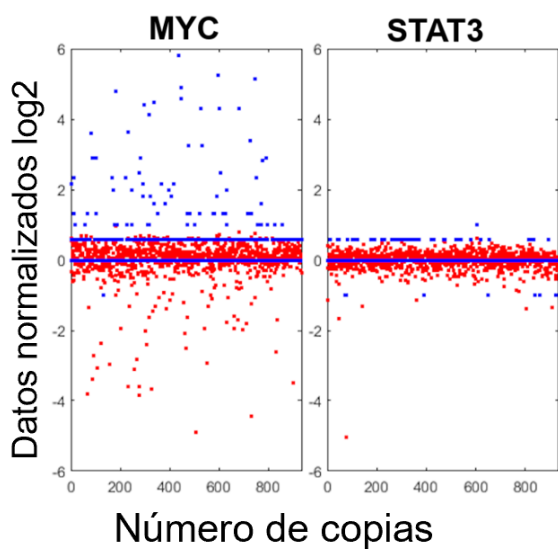


Figura 13 ampliaciones de MYC y STAT3 en datos de CCLE

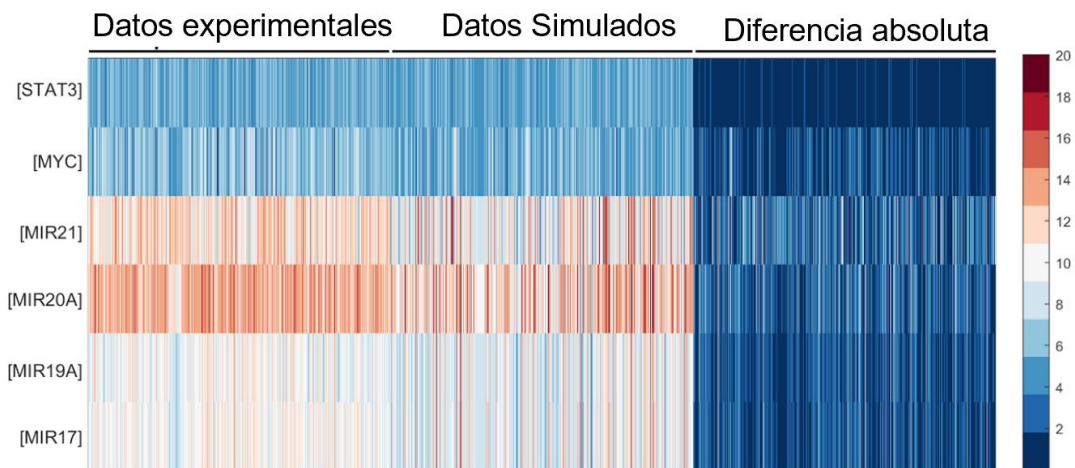


Figura 14 Comparativo de datos experimentales de CCLE con datos simulados

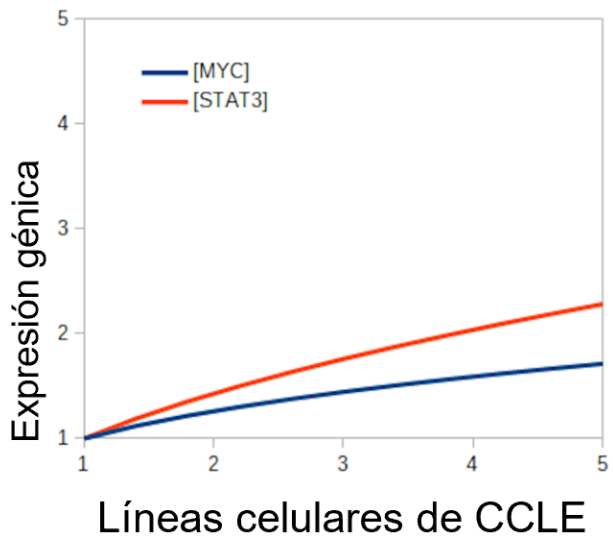


Figura 15 Compensación de MYC y STAT3 en datos de CCLE

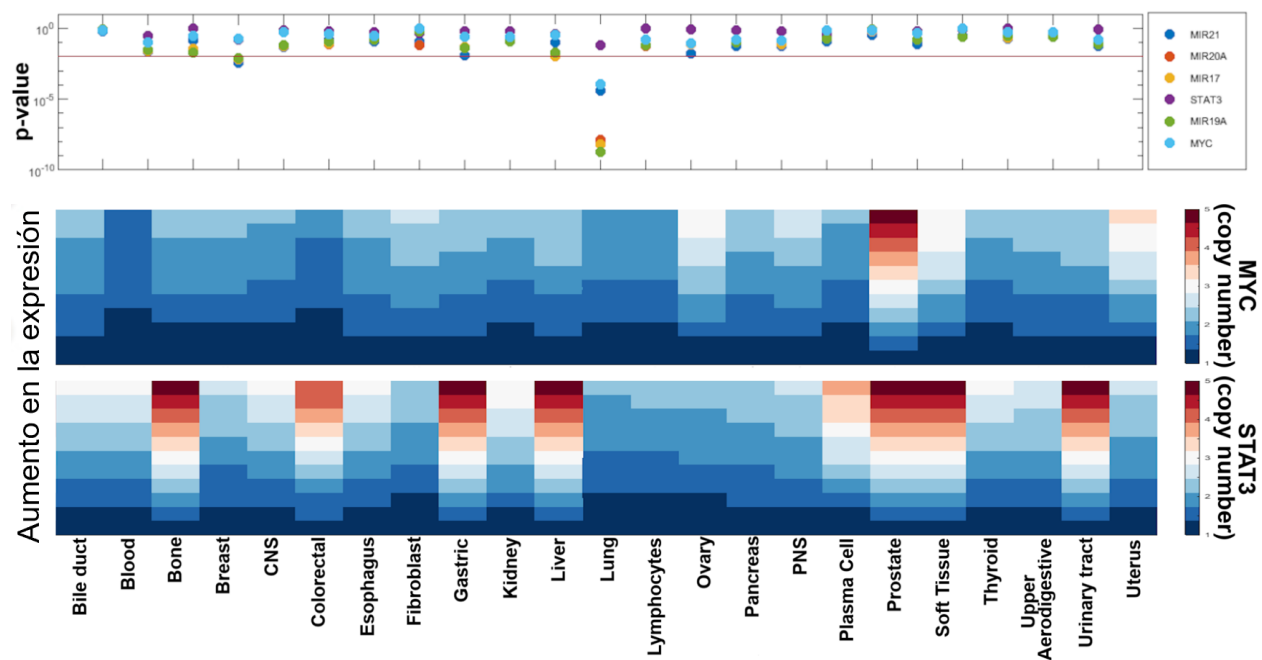


Figura 16 Resultados de expresión para MYC y STAT3 según número de copias en CCLE



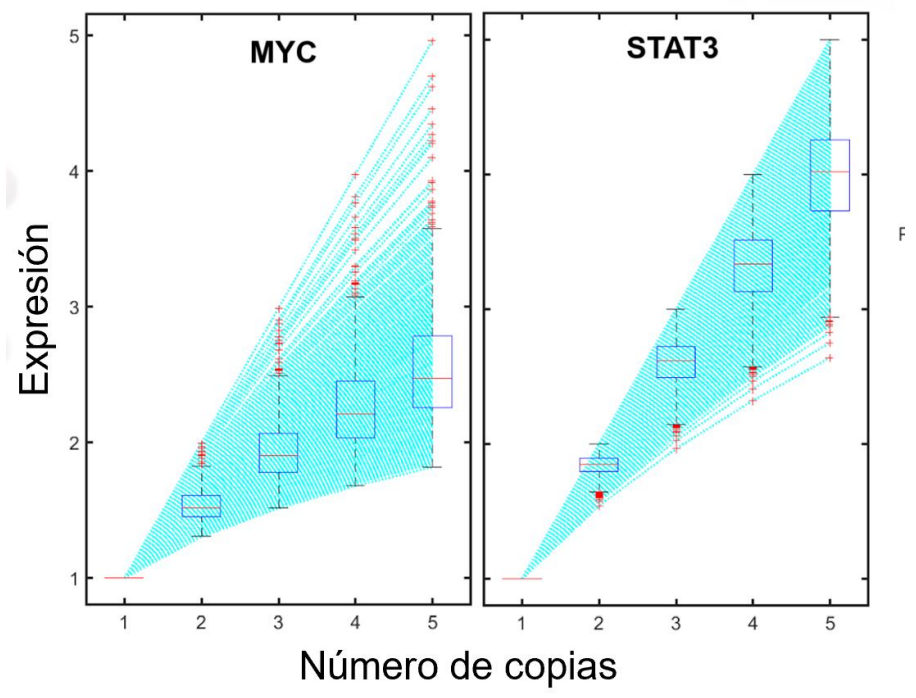


Figura 17 Niveles de expresión para MYC y STAT3 para distintas líneas celulares con datos de CCLE

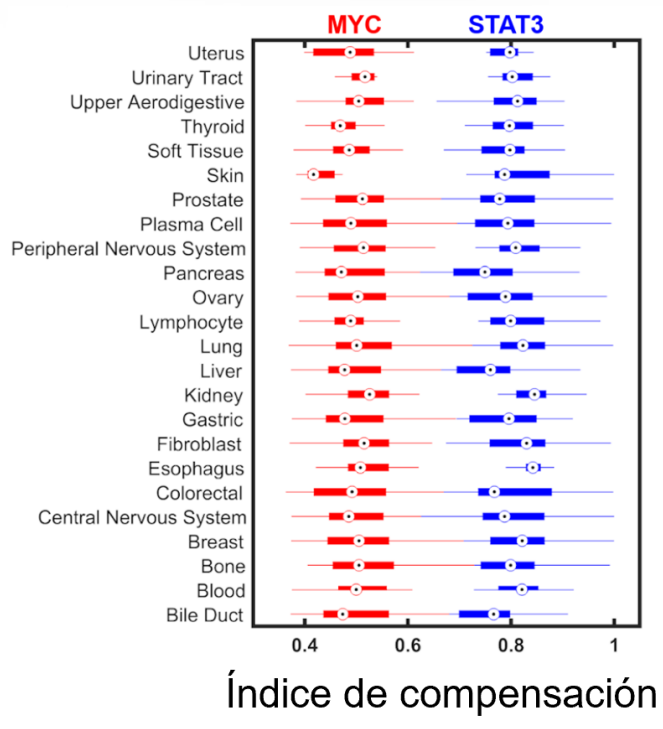


Figura 18 Índice de compensación para MYC y STAT3 en distintos tipos de cáncer con datos de CCLE

## 8. Discusión

La plataforma desarrollada permite generar rápidamente redes de interacción de genes, microARNs y factores de transcripción. Dichas redes pueden ser visualizadas en Cytoscape y los respectivos modelos generados en COPASI, reduciendo de manera muy significativa el tiempo requerido. Adicionalmente disminuyen los riesgos de introducir errores, lo cual es común cuando se realiza el proceso de forma manual. En minutos se pueden tener resultados que manualmente tomarían hasta semanas o meses y realizar cambios en la red para generar un nuevo modelo es un proceso sencillo y rápido.

La incorporación de un comando para reutilizar los valores de parámetros estimados previamente, también tiene un gran impacto en el tiempo requerido para tener nuevas versiones de un modelo. Por ejemplo, en la red del modelo mínimo extendido hacia los genes del metabolismo, el primer ajuste de parámetros tomó 20 días en un equipo con 16GB y un procesador Intel i7. Si por alguna razón fuera necesario interrumpir el proceso y retomarlo, se pueden utilizar los parámetros estimados y así no tener que reiniciar el proceso desde el principio. De igual forma si se requiere generar un segundo modelo eliminando alguna especie que se desee descartar y volver a generar y ajustar un nuevo modelo, se pueden utilizar los parámetros ajustados del modelo anterior. Eliminar algunas especies y generar un nuevo modelo es parte de la dinámica de las investigaciones que utilizan modelos, pues no todos los nodos podrían ser de interés, o los datos experimentales podrían no ser suficientes o consistentes con el resto de las especies. Por esto tener la posibilidad de reutilizar los parámetros de un modelo anterior resulta de mucha utilidad.

Uno de los criterios que se pueden utilizar para descartar especies de un modelo es aplicar un t-test para medias pareadas de los datos experimentales y los simulados para cada una de las especies. Si el  $p$  es menor a un determinado valor de significancia, es porque no se tuvo un buen ajuste para esa especie y por lo tanto no es adecuada para el modelo con los datos experimentales que se disponen. Este proceso de aplicar el t-test cuando se trabaja con miles de datos y muchas especies resulta complicado realizarlo manualmente. La opción incorporada a BioNetUCR para calcular los t-test a partir de los datos de ajuste que genera COPASI, también representa un significativo ahorro de tiempo. Las funcionalidades de reutilizar los valores de los parámetros estimados y de aplicar el t-test son adiciones que no estaban consideradas inicialmente, pero

durante el desarrollo de la tesis se determinó que realizan un importante apoyo al investigador en la creación y validación de los modelos.

En lo referente a las limitaciones identificadas inicialmente en la plataforma anterior, estas se superaron de la siguiente forma:

1. Se modela solamente un nivel de interacciones a partir de cada gen.

Los comandos ADD BACKWARD y ADD FORWARD permiten aumentar un nivel de interacción para cada nodo en las respectivas direcciones. Se puede repetir este proceso con cualquier nodo y cuantas veces se desee, sin embargo, el crecimiento de relaciones es exponencial cada vez que se usan los comandos. Por esta razón se recomienda revisar los efectos en el tamaño de la red cada vez que se expande la red agregando un nivel de interacción a alguna de las especies.

2. Está enfocada en las especies que regulan un gen en particular, no considera el caso de las especies que son reguladas por un regulador de interés.

Esta limitación fue eliminada con la creación del comando ADD FORWARD.

3. No se considera el número de copias de los genes que codifican microARNs y FTs.

Se obtuvo de CellMiner (Reinhold et al., 2012) un listado actualizado de genes con los respectivos números de copias en NCI60, lo cual hace más realista el ajuste de los modelos. En los datos utilizados en (Acón, 2018) las especies para las que se tuvo el número de copias fue muy limitado en especial para los microARNs, por lo que en muchas de las especies se usaron los valores diploides, pero ahora se dispone de los números de copias para muchas de ellos. Esta información queda disponible en el archivo NCI60Experiments.txt y puede ser usada en BioNetUCR por medio de la opción Preferences, tal y como se muestra en la *Figura 19*.

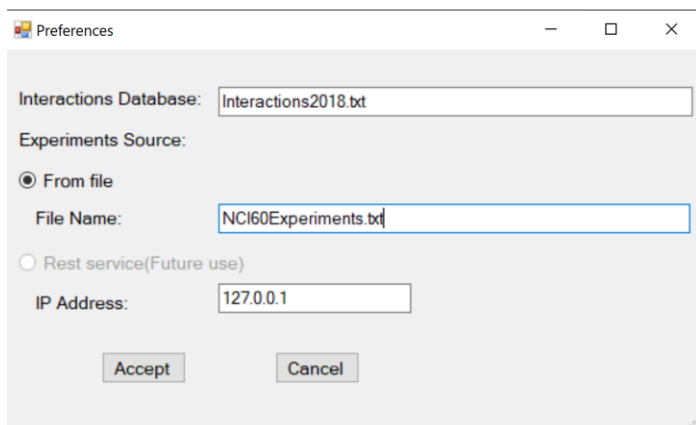


Figura 19: Configuración de preferencias en BioNetUCR

En el caso específico del modelo mínimo, el utilizar los números de copias en lugar del valor diploide y proceder a ajustar el modelo, los resultados no variaron de manera significativa. Para analizar en mayor detalle los efectos del número de copias en el modelo mínimo, se procedió a realizar un Parameter Scan en COPASI para ver el efecto de distintos valores del número de copias tanto para MYC como para STAT3. Este proceso se realizó tanto para el modelo mínimo creado con la plataforma anterior, como con BioNetUCR, los resultados se pueden ver en la *Figura 10*. Los resultados son muy similares y se nota que hay compensación, pues los niveles de expresión no crecen en la misma proporción que el número de copias.

#### 4. Incorporación y editaje de nuevas relaciones.

Se implementaron comandos que facilitan el editaje de la red, agregando o eliminando elementos de ésta, lo que permite generar distintas versiones de la red con sus respectivos modelos. La sección de anexos muestra los distintos comandos soportados por BioNetUCR. Las redes se almacenan como archivos de texto plano, esto también permite la edición de las mismas en editores de terceros, en cuyo caso se debe respetar el mismo formato que se utiliza en el comando ADD EDGE y separando cada elemento con un TAB.

5. Consultas de interacciones entre genes, FTs y microARNs.

Se crearon varios comandos para este fin. El comando PATH permite ver las diferentes rutas que hay entre dos nodos, hasta una distancia máxima de tres arcos y en ambas direcciones. SOURCE permite ver en qué arcos un nodo particular es el origen y TARGET en qué arcos es destino.

Si por ejemplo en la red del modelo mínimo usamos el comando PATH MIR19A STAT3, el resultado será el siguiente:

```
MIR19A-->MYC-->MIR17-->STAT3
MIR19A-->MYC-->MIR20A-->STAT3
STAT3-->MYC-->MIR19A
```

Pero el comando PATH MIR19A MIR21 solo retornará la ruta:

```
MIR21-->STAT3-->MYC-->MIR19A
```

La razón es que se trata de la única ruta entre las especies MIR19A y MIR21 que está a un máximo de tres arcos. Este máximo de tres está fijo en BioNetUCR, sin embargo, si en el futuro se considera útil se podría aumentar o hacer configurable por el usuario.

6. Se debe escoger entre interacciones validadas experimentalmente o putativas solamente, no permite combinar interacciones validadas con interacciones putativas en un mismo modelo.

Si se desean incorporar interacciones putativas se puede hacer creando un nuevo archivo de interacciones, agregando los arcos respectivos. Como el archivo de interacciones es configurable, se pueden tener distintos archivos, ya sea con o sin interacciones putativas.

El uso de interacciones putativas sin embargo perdió relevancia debido a que se dispone adicionalmente de una nueva versión de interacciones, con un total de 437569 interacciones validadas, que podrían utilizarse en investigaciones futuras en lugar de la versión utilizada en (Acón, 2018) y en este trabajo, la cual tiene 61286.

7. La interfaz y operación del programa depende de una persona con amplio conocimiento de la plataforma y experiencia en programación.

La plataforma fue desarrollada pensando en que sea utilizada por usuarios no especializados en informática. Teniendo instalado COPASI y Cytoscape, basta con copiar en una carpeta los distintos archivos contenidos en un archivo zip y ejecutar BioNetUCR para poder utilizarla. Posee la funcionalidad de autocompletar los comandos una vez que estos comienzan a ser digitados e incluye ayuda con un listado de todos los comandos disponibles.

8. No posee un ambiente integrado diseñado para el usuario no informático.

El esquema de usar una línea de comandos, un área para resultados y la posibilidad de crear scripts es bastante utilizado en herramientas de tipo científico, de modo que lo esperable es que el uso de la herramienta resulte natural.

La visualización de la red en Cytoscape, así como la creación y visualización del modelo en COPASI, operan de forma automática a partir de la selección de la respectiva opción del menú.

9. Lenguaje de programación utilizado.

Se desarrolló en C#, que es un lenguaje muy popular y posee un ambiente integrado amigable (Visual Studio 2019), que es de uso gratuito para ambientes universitarios.

La disponibilidad de comandos que se pueden ejecutar en distinto orden agrega mucha flexibilidad a la plataforma desarrollada. Un ejemplo es la funcionalidad de la plataforma anterior, la cual estaba creada para un fin específico y se puede simular con un conjunto de comandos en BioNetUCR. Generar un modelo de todas las especies que regulan MYC, así como todas las interacciones entre ellas aplicando las reglas que utiliza a nivel de código la plataforma anterior, se puede lograr con la plataforma nueva usando los comandos que se muestran en la *Tabla 2*.

El principal comando es ADD BACKWARD MYC, el cual busca todos los microARNs y FTs que regulan MYC, así como todas las relaciones entre ellos. Posteriormente se procede a ejecutar otros comandos para eliminar nodos autoregulados o que no son regulados por otros, así como los nodos que representan especies para las cuales no hay datos experimentales. Finalmente se utiliza el comando TAG para indicar que MYC es un factor de transcripción de interés especial. Esto significa que se le deben aplicar las fórmulas de síntesis y degradación para genes indicadas en la *Tabla 3*, la cual muestra las fórmulas de síntesis y degradación utilizadas en ambas plataformas.

```

ADD BACKWARD MYC
DELETE AUTOREGULATED EDGES
DELETE SOURCE NODES
DELETE EDGE E2F1 MIR449C
DELETE EDGE ETS1 HOXB4
DELETE EDGE ETS1 NME2
DELETE EDGE GATA2 HOXB4
DELETE EDGE HOXB4 MYC
DELETE EDGE MIR449C MYC
DELETE EDGE MYB NME2
DELETE EDGE MYC NME2
DELETE EDGE NME2 MYC
TAG MYC

```

*Tabla 2: Comandos para crear red compatible con la desarrollada en plataforma previa*

$$Synthesis\ gene = CN_{gene} \cdot KS_{gene} \cdot \frac{1 + KA_{TF A1} \cdot [TF A1] + \dots}{1 + KR_{TF R1} \cdot [TF R1] + \dots}$$

$$Degradation\ gene\ RNA = KD_{gene\ RNA} \cdot [Gene\ RNA] \cdot (1 + KR_{mi\ RNA 1} \cdot [mi\ RNA 1] + \dots)$$

$$Synthesis\ TF = CN_{TF} \cdot KS_{TF} \cdot \frac{1 + KA_{TF A1} \cdot [TF A1] + \dots}{1 + KR_{TF R1} \cdot [TF R1] + \dots + KR_{mi\ RNA 1} \cdot [mi\ RNA 1] + \dots}$$

$$Degradation\ TF = KD_{TF} \cdot [TF]$$

$$Synthesis\ mi\ RNA = CN_{mi\ RNA} \cdot KS_{mi\ RNA} \cdot \frac{1 + KA_{TF A1} \cdot [TF A1] + \dots}{1 + KR_{TF R1} \cdot [TF R1] + \dots}$$

$$Degradation\ mi\ RNA = KD_{mi\ RNA} \cdot [mi\ RNA]$$

*Tabla 3: Funciones de síntesis y degradación usados en genes, microARNs y FTs.*

Los modelos computacionales realizados tanto con datos de NCI60 como de CCLE sugieren que la identificación de blancos que puedan bloquear la compensación de dosis génica podrían conducir a una estrategia en la que la sobreexpresión de estos genes, y otros bajo su influencia, permitan combatir ciertos tipos de cáncer. No obstante, la posible aplicación terapéutica de una estrategia dirigida a la compensación de dosis génica en las células cancerosas requiere una evaluación sistemática del alcance de las alteraciones del número de copias y su compensación en grandes conjuntos de datos, como CCLE y TCGA, para la identificación de más genes candidatos. El análisis adicional de los modelos actuales de compensación de dosis génica o incluso de otros genes podría revelar blancos adicionales en el cáncer en general o en tipos específicos de cáncer. La búsqueda futura de los determinantes moleculares de la compensación, por ejemplo, mutaciones en las secuencias de interacción de los microRNA blanco, podría conducir a la identificación de biomarcadores para dirigir estrategias de medicina de precisión.

Adicionalmente, las fuertes interacciones resultantes entre las especies de la red mínima y algunos genes del metabolismo, representan una oportunidad para realizar futuras investigaciones que incluyan simulaciones en COPASI con datos experimentales de distintos tipos de cáncer y ver los efectos de perturbaciones en algunas de las especies. Para estos estudios se pueden usar las funcionalidades de BioNetUCR que permiten utilizar distintas fuentes de datos, de modo que las investigaciones no estén limitadas a NCI60, sino que se utilicen también fuentes de datos como TCGA y CCLE.



## 9. Conclusión

La plataforma desarrollada (BioNetUCR) permite generar fácilmente modelos matemáticos de las relaciones entre genes, factores de transcripción y microARNs, incorporando datos experimentales. BioNetUCR disminuye de forma muy significativa el tiempo necesario para crear modelos de este tipo, pudiendo tener en minutos modelos que manualmente podrían tomar semanas o hasta meses. El uso de la plataforma reduce además la posibilidad de introducir errores en la elaboración de modelos.

Para validar el funcionamiento de BioNetUCR se realizó una comparación con un modelo mínimo de compensación de dosis génica generado en una investigación previa, obteniendo los mismos resultados. Para demostrar las capacidades de la plataforma desarrollada, se utilizaron distintos comandos de ésta para generar un nuevo modelo extendido a partir del mínimo, mostrando las relaciones de genes del metabolismo que son regulados por las especies del modelo mínimo. Posteriormente el modelo fue ajustado con datos experimentales en la herramienta COPASI. Adicionalmente con BioNetUCR se generaron varios modelos utilizando datos de CCLE para distintos tipos de cáncer, pudiendo determinarse que hay compensación de dosis génica de MYC en la mayoría de los tipos de cáncer a pesar del alto nivel de heterogeneidad observado en el cáncer en general. Esto confirma los hallazgos de la investigación previa con los datos de NCI60.

Cada uno de los objetivos específicos se pudieron cumplir, la red generada a partir del modelo mínimo y su influencia hacia los genes del metabolismo, así como los estudios con los datos de CCLE validan la hipótesis y objetivo general planteados. En el caso de los genes del metabolismo, la red generada y el respectivo modelo comprenden 439 nodos y 816 arcos. Tras casi 3 semanas de ejecución de COPASI se logró realizar un ajuste del modelo. Posteriormente usando BioNetUCR se generaron los  $p$  que permiten evaluar si algunas de las especies se deben descartar por no tener un ajuste adecuado con los datos experimentales que se disponen.

De la red generada se analizaron los genes del metabolismo que son regulados por al menos tres especies de la red del modelo mínimo. Los genes de interés resultantes fueron PTEN, DNMT1, TGFBR2, UBE2C y BMPR2, además está HMOX1 que es el único gen del metabolismo regulado por STAT3 y MYC, todos ellos según la literatura tienen un papel importante en el cáncer. Está

fuera del alcance de esta tesis la investigación en más profundidad del modelo generado con todos los genes del metabolismo, pero los resultados sirven de base para nuevos proyectos de investigación de otros estudiantes de la maestría en Bioinformática y Biología de Sistemas.

La plataforma incorpora funcionalidades adicionales a las previamente planeadas, como la generación de valores  $p$  a partir de aplicar un t-test para muestras pareadas. También se incluyó una prevista para en el futuro poder acceder a datos experimentales en un servicio de tipo REST, en el cual no solo se tengan datos de NCI60, sino también TCGA y CCLE. Sus distintas funcionalidades y facilidad de uso la hacen de clara utilidad para investigadores y estudiantes en estudios de redes de interacción entre genes, factores de transcripción y microARNs. La plataforma facilita el análisis de los efectos de distintas especies sobre otras, haciendo viable la utilización de modelos matemáticos sin requerir conocimientos profundos en el área de informática.

## 10. Anexos

### Comandos soportados por BioNetUCR

- **ADD FORWARD *node***: agrega todos los nodos regulados por el nodo *node* y las interrelaciones entre ellos y los nodos existentes en la red del usuario.
- **ADD BACKWARD *node***: Busca todos los nodos que regulan el nodo *node* y las interrelaciones entre ellos y los nodos existentes en la red del usuario.
- **ADD EDGE *source\_node source\_type target\_node target\_type action***: agrega un nuevo arco del nodo *source\_node* al nodo *target\_type*. El parámetro *target\_node* puede ser *GENE* (representa un gen), *MIR* (representa un microARN) o *TF* (representa un factor de transcripción). *Action* puede ser *A* (representa activación) o *R* (representa represión).
- **CLEAR ALL**: limpia todo el ambiente definido dentro de la plataforma, es similar a cerrar y volver a abrir la plataforma.
- **CLEAR RESULTS**: limpia la ventana de resultados de la plataforma.
- **DELETE EDGE *source\_node target\_node***: borra el arco entre los nodos indicados.
- **DELETE NODE *node***: borra el nodo indicado y todos los arcos con *node* como origen o destino.
- **DELETE NODES IN FILE *file\_name***: borra todos los nodos en el archivo especificado. El archivo debe ser texto plano y contener un nodo por línea, por ejemplo:  

```
ETS1
Gata2
```
- **DELETE NODES NOT IN FILE *file\_name***: borra todos los nodos que no estén en el archivo especificado. El formato del archivo es el mismo que en el comando anterior.
- **DELETE AUTOREGULATED EDGES**: borra todos los arcos que van de un nodo a sí mismo.
- **DELETE SOURCE NODES**: borra todos los nodos que no son regulados por otros nodos.
- **DELETE SINK NODES**: borra todos los nodos que no regulan a ningún otro nodo.
- **EXPERIMENTS FILE *file\_name***: por defecto la plataforma espera que los datos experimentales estén en el archivo llamado *experiments.tab*, este comando permite indicar que use otro archivo al momento de generar el modelo en COPASI. El formato del archivo

es similar al que utiliza COPASI, es texto plano y en la primera columna está la línea celular, luego están las variables independientes y dependientes separadas por un TAB. Las columnas que representan un número de copias de una especie deben iniciar con *CNV\_* seguido por el nombre de la especie. El orden en que aparecen las columnas para cada línea celular no es relevante, se utiliza el nombre para su clasificación. Ejemplo de archivo:

LINE	CNV_MIR21	CNV_MIR20A	CNV_MIR17	MIR21	MIR20A	MIR17
BR:BT_549	1	1	2	7.1225	12.37	10.775
LINE	CNV_MIR21	CNV_MIR20A	CNV_MIR17	MIR21	MIR20A	MIR17
BR:HS578T	1	1	1	7.5625	10.94	9.595
LINE	CNV_MIR21	CNV_MIR20A	CNV_MIR17	MIR21	MIR20A	MIR17
BR:MCF7	1	2	1	7.265	11.28	9.45

- **LIST BY SOURCE:** lista los arcos de la red del usuario ordenados por nodo origen.
- **LIST BY TARGET:** lista los arcos de la red del usuario ordenados por nodo destino.
- **LOAD *network*:** carga una red que se haya almacenado previamente en el archivo *network*.
- **PARAMETERS FILE *file\_name*:** permite crear el modelo usando valores para los parámetros *K* especificados en un archivo. Cada línea del archivo de texto debe tener el nombre del parámetro seguido por el valor, separados por un TAB. Por ejemplo:
 

```
ks_ATF4    0.998007
kd_ATF4    0.106538
```
- **PATH *node\_1 node\_2*:** despliega todas las rutas entre los dos nodos en ambas direcciones, hasta una distancia máxima de tres arcos.
- **SAVE *network*:** guarda en el archivo *network* la red del usuario.
- **SOURCE *node*:** despliega en qué arcos *node* es el origen.
- **TAG *gen*:** marca un gen que es de interés al crear el modelo en COPASI, esto hace que se trate como gen en las fórmulas de las funciones, aunque sea un factor de transcripción.
- **TARGET *node*:** despliega en qué arcos *node* es el destino.
- **VIEW NETWORK:** muestra un gráfico de la red actual. Requiere que Cytoscape esté ejecutándose.

*Nota:* ni los comandos ni sus parámetros son sensitivos a mayúsculas.

## 11. Bibliografía

- Abaan, O. D., Polley, E. C., Davis, S. R., Zhu, Y. J., Bilke, S., Walker, R. L., Pineda, M., Gindin, Y., Jiang, Y., Reinhold, W. C., Holbeck, S. L., Simon, R. M., Doroshov, J. H., Pommier, Y., & Meltzer, P. S. (2013). The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology. *Cancer Research*, 73(14), 4372–4382.  
<https://doi.org/10.1158/0008-5472.CAN-12-3342>
- Acón, M. S. (2018). *Plataforma de biología de sistemas para la predicción de microARNs involucrados en redes de regulación de dosis génica en cáncer aneuploide*. Universidad de Costa Rica.
- Acón, M. S., Siles, F., & Mora, R. (2016). *A biocomputational platform for the automated construction of large-scale mathematical models of miRNA-transcription factor networks for studies on gene dosage compensation*. 1–7.  
<https://doi.org/10.1109/CONCAPAN.2016.7942348>
- Aliee, H., Massip, F., Qi, C., de Biase, M. S., van Nijnatten, J., Kersten, E. T. G., Kermani, N. Z., Khuder, B., Vonk, J. M., Vermeulen, R. C. H., Neighbors, M., Tew, G. W., Grimbaldeston, M., ten Hacken, N. H. T., Hu, S., Guo, Y., Zhang, X., Sun, K., Hiemstra, P. S., ... Faiz, A. (2020). Determinants of SARS-CoV-2 receptor gene expression in upper and lower airways. *MedRxiv*, 2020.08.31.20169946.  
<https://doi.org/10.1101/2020.08.31.20169946>
- Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F. A., Lenhof, H.-P., Keller, A., & Meese, E. (2019). An estimate of the total number of true human miRNAs. *Nucleic Acids Research*, 47(7), 3353–3364.  
<https://doi.org/10.1093/nar/gkz097>

- Alon, U. (2006). *An Introduction to Systems Biology. Design Principles of Biological Circuits*.  
(First).
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., ... Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–607. <https://doi.org/10.1038/nature11003>
- Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., & Benenson, Y. (2011). Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology*, *7*, 519. <https://doi.org/10.1038/msb.2011.49>
- Brennan, C. M., Vaites, L. P., Wells, J. N., Santaguida, S., Paulo, J. A., Storchova, Z., Harper, J. W., Marsh, J. A., & Amon, A. (2019). Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells. *Genes & Development*, *33*(15–16), 1031–1047. <https://doi.org/10.1101/gad.327494.119>
- Brown, J. S., & Jackson, S. P. (2015). Ubiquitylation, neddylation and the DNA damage response. *Open Biology*, *5*(4). <https://doi.org/10.1098/rsob.150018>
- Carbo, A., Bassaganya-Riera, J., Pedragosa, M., Viladomiu, M., Marathe, M., Eubank, S., Wendelsdorf, K., Bisset, K., Hoops, S., Deng, X., Alam, M., Kronsteiner, B., Mei, Y., & Hontecillas, R. (2013). Predictive Computational Modeling of the Mucosal Immune Responses during *Helicobacter pylori* Infection. *PLoS ONE*, *8*(9). <https://doi.org/10.1371/journal.pone.0073365>
- Dang, C. V. (2012). MYC on the Path to Cancer. *Cell*, *149*(1), 22–35. <https://doi.org/10.1016/j.cell.2012.03.003>

- Dastsooz, H., Cereda, M., Donna, D., & Oliviero, S. (2019). A Comprehensive Bioinformatics Analysis of UBE2C in Cancers. *International Journal of Molecular Sciences*, 20(9).  
<https://doi.org/10.3390/ijms20092228>
- de-Leon, S. B.-T., & Davidson, E. H. (2009). Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Developmental Biology*, 325(2), 317–328.  
<https://doi.org/10.1016/j.ydbio.2008.10.043>
- Eftimie, R., Gillard, J. J., & Cantrell, D. A. (2016). Mathematical Models for Immunology: Current State of the Art and Future Research Directions. *Bulletin of Mathematical Biology*, 78(10), 2091–2134. <https://doi.org/10.1007/s11538-016-0214-9>
- Emmert-Streib, F., & Dehmer, M. (2011). Emmert Streib F, Dehmer M (May 2011) Networks for systems biology: Conceptual connection of data and function. *IET Syst Biol. IET Systems Biology*, 5, 185–207. <https://doi.org/10.1049/iet-syb.2010.0025>
- Fischer, H. P. (2008). Mathematical Modeling of Complex Biological Systems. *Alcohol Research & Health*, 31(1), 49–59.
- Flynt, A. S., & Lai, E. C. (2008). Biological principles of microRNA-mediated regulation: Shared themes amid diversity. *Nature Reviews. Genetics*, 9(11), 831–842.  
<https://doi.org/10.1038/nrg2455>
- Fu, X., Qie, J., Fu, Q., Chen, J., Jin, Y., & Ding, Z. (2020). MiR-20a-5p/TGFBR2 Axis Affects Pro-inflammatory Macrophages and Aggravates Liver Fibrosis. *Frontiers in Oncology*, 10. <https://doi.org/10.3389/fonc.2020.00107>
- Gao, X., Qin, T., Mao, J., Zhang, J., Fan, S., Lu, Y., Sun, Z., Zhang, Q., Song, B., & Li, L. (2019). PTENP1/miR-20a/PTEN axis contributes to breast cancer progression by

- regulating PTEN via PI3K/AKT pathway. *Journal of Experimental & Clinical Cancer Research : CR*, 38. <https://doi.org/10.1186/s13046-019-1260-6>
- Guo, T., Luna, A., Rajapakse, V. N., Koh, C. C., Wu, Z., Liu, W., Sun, Y., Gao, H., Menden, M. P., Xu, C., Calzone, L., Martignetti, L., Auwerx, C., Buljan, M., Banaei-Esfahani, A., Ori, A., Iskar, M., Gillet, L., Bi, R., ... Aebersold, R. (2019). Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines. *IScience*, 21, 664–680. <https://doi.org/10.1016/j.isci.2019.10.059>
- Guo, Y., Alexander, K., Clark, A. G., Grimson, A., & Yu, H. (2016). Integrated network analysis reveals distinct regulatory roles of transcription factors and microRNAs. *RNA*, 22(11), 1663–1672. <https://doi.org/10.1261/rna.048025.114>
- Hao, Z., Zhang, H., & Cowell, J. (2012). Ubiquitin-conjugating enzyme UBE2C: Molecular biology, role in tumorigenesis, and potential as a biomarker. *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, 33(3), 723–730. <https://doi.org/10.1007/s13277-011-0291-1>
- Hershberg, R., & Margalit, H. (2006). Co-evolution of transcription factors and their targets depends on mode of regulation. *Genome Biology*, 7(7), R62. <https://doi.org/10.1186/gb-2006-7-7-r62>
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., & Kummer, U. (2006). COPASI—a COMplex PATHway SIMulator. *Bioinformatics*, 22(24), 3067–3074. <https://doi.org/10.1093/bioinformatics/bt1485>
- Hu, Z., Zhou, S., Luo, H., Ji, M., Zheng, J., Huang, F., & Wang, F. (2019). MiRNA-17 promotes nasopharyngeal carcinoma radioresistance by targeting PTEN/AKT. *International Journal of Clinical and Experimental Pathology*, 12(1), 229–240.



- Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., & Liu, Y. (2009). miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research*, *37*(Database issue), D98-104.  
<https://doi.org/10.1093/nar/gkn714>
- Karmiloff-Smith, A., Al-Janabi, T., D'Souza, H., Groet, J., Massand, E., Mok, K., Startin, C., Fisher, E., Hardy, J., Nizetic, D., Tybulewicz, V., & Strydom, A. (2016). The importance of understanding individual differences in Down syndrome. *F1000Research*, *5*.  
<https://doi.org/10.12688/f1000research.7506.1>
- Kaur, M., & Cole, M. D. (2013). MYC Acts via the PTEN Tumor Suppressor to Elicit Autoregulation and Genome-Wide Gene Repression by Activation of the Ezh2 Methyltransferase. *Cancer Research*, *73*(2), 695–705.
- Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, *68*(6), 540–546. <https://doi.org/10.4097/kjae.2015.68.6.540>
- Kojima, S., & Cimini, D. (2019). Aneuploidy and gene expression: Is there dosage compensation? *Epigenomics*, *11*(16), 1827–1837. <https://doi.org/10.2217/epi-2019-0135>
- Kourtis, N., & Tavernarakis, N. (2009). Cell-Specific Monitoring of Protein Synthesis In Vivo. *PLoS ONE*, *4*(2). <https://doi.org/10.1371/journal.pone.0004547>
- Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(Database issue), D155–D162.  
<https://doi.org/10.1093/nar/gky1141>
- Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(Database issue), D68–D73.  
<https://doi.org/10.1093/nar/gkt1181>

- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., & Tuschl, T. (2002). Identification of Tissue-Specific MicroRNAs from Mouse. *Current Biology*, *12*(9), 735–739. [https://doi.org/10.1016/S0960-9822\(02\)00809-6](https://doi.org/10.1016/S0960-9822(02)00809-6)
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>
- Lee, H., Jeong, A. J., & Ye, S.-K. (2019). Highlighted STAT3 as a potential drug target for cancer therapy. *BMB Reports*, *52*(7), 415–423. <https://doi.org/10.5483/BMBRep.2019.52.7.152>
- Li, H., & Johnson, A. D. (2010). Evolution of Transcription Networks—Lessons from Yeasts. *Current Biology : CB*, *20*(17), R746–R753. <https://doi.org/10.1016/j.cub.2010.06.056>
- Li, H., Liu, J., Sun, L., & Yuan, Y. (2017). A Meta-Analysis of the Association between DNMT1 Polymorphisms and Cancer Risk. *BioMed Research International*, *2017*. <https://doi.org/10.1155/2017/3971259>
- Liu, H.-Y., Zhang, Y.-Y., Zhu, B.-L., Feng, F.-Z., Yan, H., Zhang, H.-Y., & Zhou, B. (2019). MiR-21 regulates the proliferation and apoptosis of ovarian cancer cells through PTEN/PI3K/AKT. *European Review for Medical and Pharmacological Sciences*, *23*(10), 4149–4155. [https://doi.org/10.26355/eurev\\_201905\\_17917](https://doi.org/10.26355/eurev_201905_17917)
- Ma, T., Li, H., Sun, M., Yuan, Y., & Sun, L.-P. (2017). DNMT1 overexpression predicting gastric carcinogenesis, subsequent progression and prognosis: A meta and bioinformatic analysis. *Oncotarget*, *8*(56), 96396–96408. <https://doi.org/10.18632/oncotarget.21480>
- Medina, P. P., & Slack, F. J. (2008). microRNAs and cancer: An overview. *Cell Cycle (Georgetown, Tex.)*, *7*(16), 2485–2492. <https://doi.org/10.4161/cc.7.16.6453>

- Mishra, S., Deng, J. J., Gowda, P. S., Rao, M. K., Lin, C.-L., Chen, C. L., Huang, T., & Sun, L.-Z. (2014). Androgen Receptor and MicroRNA-21 axis down-regulates transforming growth factor beta receptor II (TGFB2) expression in Prostate Cancer. *Oncogene*, 33(31), 4097–4106. <https://doi.org/10.1038/onc.2013.374>
- Mott, J. L., & Mohr, A. M. (2015). Overview of MicroRNA Biology. *Seminars in Liver Disease*, 35(1), 3–11. <https://doi.org/10.1055/s-0034-1397344>
- Najafi, A., Bidkhorji, G., Bozorgmehr, J. H., Koch, I., & Masoudi-Nejad, A. (2014). Genome Scale Modeling in Systems Biology: Algorithms and Resources. *Current Genomics*, 15(2), 130–159. <https://doi.org/10.2174/1389202915666140319002221>
- Nakanishi, K. (2016). Anatomy of RISC: How do small RNAs and chaperones activate Argonaute proteins? *Wiley Interdisciplinary Reviews. RNA*, 7(5), 637–660. <https://doi.org/10.1002/wrna.1356>
- Nitti, M., Piras, S., Brondolo, L., Marinari, U. M., Pronzato, M. A., & Furfaro, A. L. (2018). Heme Oxygenase 1 in the Nervous System: Does It Favor Neuronal Cell Survival or Induce Neurodegeneration? *International Journal of Molecular Sciences*, 19(8). <https://doi.org/10.3390/ijms19082260>
- Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daniélsdóttir, A. D., Krecke, M., Merten, D., Haraldsdóttir, H. S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D. A., Sahoo, S., Gawron, P., Friscioni, L., Garcia, B., Prendergast, M., ... Thiele, I. (2019). The Virtual Metabolic Human database: Integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1), D614–D624. <https://doi.org/10.1093/nar/gky992>

- Oulas, A., Minadakis, G., Zachariou, M., Sokratous, K., Bourdakou, M. M., & Spyrou, G. M. (2017). Systems Bioinformatics: Increasing precision of computational diagnostics and therapeutics through network-based approaches. *Briefings in Bioinformatics*, *20*(3), 806–824. <https://doi.org/10.1093/bib/bbx151>
- Padi, M., & Quackenbush, J. (2015). Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Systems Biology*, *9*. <https://doi.org/10.1186/s12918-015-0228-1>
- Politano, G., Orso, F., Raimo, M., Benso, A., Savino, A., Taverna, D., & Di Carlo, S. (2016). CyTRANSFINDER: A Cytoscape 3.3 plugin for three-component (TF, gene, miRNA) signal transduction pathway construction. *BMC Bioinformatics*, *17*. <https://doi.org/10.1186/s12859-016-0964-2>
- Rahiminejad, S., Maurya, M. R., & Subramaniam, S. (2019). Topological and functional comparison of community detection algorithms in biological networks. *BMC Bioinformatics*, *20*. <https://doi.org/10.1186/s12859-019-2746-0>
- Ramírez-Clavijo, S., & Montoya-Ortíz, G. (2013). Gene expression and regulation. In *Autoimmunity: From Bench to Bedside [Internet]*. El Rosario University Press. <https://www.ncbi.nlm.nih.gov/books/NBK459456/>
- Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., Doroshov, J., & Pommier, Y. (2012). CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Research*, *72*(14), 3499–3511. <https://doi.org/10.1158/0008-5472.CAN-12-1370>
- Roy, S., Lagree, S., Hou, Z., Thomson, J. A., Stewart, R., & Gasch, A. P. (2013). Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling

- Networks. *PLoS Computational Biology*, 9(10).  
<https://doi.org/10.1371/journal.pcbi.1003252>
- Safran, M., Solomon, I., Shmueli, O., Lapidot, M., Shen-Orr, S., Adato, A., Ben-Dor, U., Esterman, N., Rosen, N., Peter, I., Olender, T., Chalifa-Caspi, V., & Lancet, D. (2002). GeneCards<sup>TM</sup> 2002: Towards a complete, object-oriented, human gene compendium. *Bioinformatics*, 18(11), 1542–1543. <https://doi.org/10.1093/bioinformatics/18.11.1542>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sheltzer, J. M., & Amon, A. (2011). The Aneuploidy Paradox: Costs and Benefits of an Incorrect Karyotype. *Trends in Genetics : TIG*, 27(11), 446–453.  
<https://doi.org/10.1016/j.tig.2011.07.003>
- Song, J., Zhang, X., Liao, Z., Liang, H., Chu, L., Dong, W., Zhang, X., Ge, Q., Liu, Q., Fan, P., Zhang, Z., & Zhang, B. (2019). 14–3–3 $\zeta$  inhibits heme oxygenase-1 (HO-1) degradation and promotes hepatocellular carcinoma proliferation: Involvement of STAT3 signaling. *Journal of Experimental & Clinical Cancer Research : CR*, 38.  
<https://doi.org/10.1186/s13046-018-1007-9>
- Song, Q., Grene, R., Heath, L. S., & Li, S. (2017). Identification of regulatory modules in genome scale transcription regulatory networks. *BMC Systems Biology*, 11.  
<https://doi.org/10.1186/s12918-017-0493-2>
- Tan, Q., Wang, H., Hu, Y., Hu, M., Li, X., Aodengqimuge, Ma, Y., Wei, C., & Song, L. (2015). Src/STAT3-dependent heme oxygenase-1 induction mediates chemoresistance of breast

- cancer cells to doxorubicin by promoting autophagy. *Cancer Science*, *106*(8), 1023–1032. <https://doi.org/10.1111/cas.12712>
- Tjian, R. (1978). The binding site on SV40 DNA for a T antigen-related protein. *Cell*, *13*(1), 165–179.
- Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemporary Oncology*, *19*(1A), A68–A77. <https://doi.org/10.5114/wo.2014.47136>
- Torres, E. M., Williams, B. R., & Amon, A. (2008). Aneuploidy: Cells Losing Their Balance. *Genetics*, *179*(2), 737–746. <https://doi.org/10.1534/genetics.108.090878>
- Tsang, J., Zhu, J., & van Oudenaarden, A. (2007). MicroRNA-mediated Feedback and Feedforward Loops are Recurrent Network Motifs in Mammals. *Molecular Cell*, *26*(5), 753–767. <https://doi.org/10.1016/j.molcel.2007.05.018>
- Velapasamy, S., Dawson, C. W., Young, L. S., Paterson, I. C., & Yap, L. F. (2018). The Dynamic Roles of TGF- $\beta$  Signalling in EBV-Associated Cancers. *Cancers*, *10*(8). <https://doi.org/10.3390/cancers10080247>
- Vidotto, T., Tiezzi, D. G., & Squire, J. A. (2018). Distinct subtypes of genomic PTEN deletion size influence the landscape of aneuploidy and outcome in prostate cancer. *Molecular Cytogenetics*, *11*. <https://doi.org/10.1186/s13039-017-0348-y>
- Wang, R., Song, Y., Liu, X., Wang, Q., Wang, Y., Li, L., Kang, C., & Zhang, Q. (2017). UBE2C induces EMT through Wnt/ $\beta$ -catenin and PI3K/Akt signaling pathways by regulating phosphorylation levels of Aurora-A. *International Journal of Oncology*, *50*(4), 1116–1126. <https://doi.org/10.3892/ijo.2017.3880>

- Wang, S., Ren, T., Jiao, G., Huang, Y., Bao, X., Zhang, F., Liu, K., Zheng, B., Sun, K., & Guo, W. (2017). BMPR2 promotes invasion and metastasis via the RhoA-ROCK-LIMK2 pathway in human osteosarcoma cells. *Oncotarget*, 8(35), 58625–58641.  
<https://doi.org/10.18632/oncotarget.17382>
- Wightman, B., Ha, I., & Ruvkun, G. (1994). Wightman, B. , Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855-862. *Cell*, 75, 855–862.  
[https://doi.org/10.1016/0092-8674\(93\)90530-4](https://doi.org/10.1016/0092-8674(93)90530-4)
- Wikman, H., Lamszus, K., Detels, N., UsLAR, L., Wrage, M., Benner, C., Hohensee, I., Ylstra, B., Eylmann, K., Zapatka, M., Sauter, G., Kemming, D., Glatzel, M., Müller, V., Westphal, M., & Pantel, K. (2012). Relevance of PTEN loss in brain metastasis formation in breast cancer patients. *Breast Cancer Research : BCR*, 14(2), R49.  
<https://doi.org/10.1186/bcr3150>
- Yang, H., Zhang, H., Zhong, Y., Wang, Q., Yang, L., Kang, H., Gao, X., Yu, H., Xie, C., Zhou, F., & Zhou, Y. (2017). Concomitant underexpression of TGFBR2 and overexpression of hTERT are associated with poor prognosis in cervical cancer. *Scientific Reports*, 7.  
<https://doi.org/10.1038/srep41670>
- Zhang, H., Kuang, S., Xiong, X., Gao, T., Liu, C., & Guo, A.-Y. (2015). Transcription factor and microRNA co-regulatory loops: Important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics*, 16(1), 45–58. <https://doi.org/10.1093/bib/bbt085>
- Zhang, H.-Q., Wang, J.-Y., Li, Z.-F., Cui, L., Huang, S.-S., Zhu, L.-B., Sun, Y., Yang, R., Fan, H.-H., Zhang, X., & Zhu, J.-H. (2021). DNA Methyltransferase 1 Is Dysregulated in

- Parkinson's Disease via Mediation of miR-17. *Molecular Neurobiology*, 58(6), 2620–2633. <https://doi.org/10.1007/s12035-021-02298-w>
- Zhang, Y., Han, T., Wei, G., & Wang, Y. (2015). Inhibition of microRNA-17/20a suppresses cell proliferation in gastric cancer by modulating UBE2C expression. *Oncology Reports*, 33(5), 2529–2536. <https://doi.org/10.3892/or.2015.3835>
- Zhang, Y., Liu, X., Zhang, J., Xu, Y., Shao, J., Hu, Y., Shu, P., & Cheng, H. (2020). Inhibition of miR-19a partially reversed the resistance of colorectal cancer to oxaliplatin via PTEN/PI3K/AKT pathway. *Aging (Albany NY)*, 12(7), 5640–5650. <https://doi.org/10.18632/aging.102929>
- Zhao, Q., Liu, H., Yao, C., Shuai, J., & Sun, X. (2016). Effect of Dynamic Interaction between microRNA and Transcription Factor on Gene Expression. *BioMed Research International*, 2016, 1–10. <https://doi.org/10.1155/2016/2676282>
- Zhou, D., Wan, Y., Xie, D., Wang, Y., Wei, J., Yan, Q., Lu, P., Mo, L., Xie, J., Yang, S., & Qi, X. (2015). DNMT1 mediates chemosensitivity by reducing methylation of miRNA-20a promoter in glioma cells. *Experimental & Molecular Medicine*, 47(9), e182. <https://doi.org/10.1038/emm.2015.57>