

Aplicación web para convertir archivos a un formato compatible entre diferentes softwares de análisis de redes biológicas

Lic. María D. García¹, Dr. Daniel V. Guebel²

¹Investigador de [Sidar](#) y responsable de [Atedis](#). Buenos Aires. Argentina (e-mail: garcia.mariad@gmail.com)

²Consultor. Biotechnology Counseling Services. Buenos Aires. Argentina (e-mail: dvguebel@hotmail.com)

Aplicación web para convertir archivos a un formato compatible entre diferentes software de análisis de redes biológicas

Resumen: Se presenta una aplicación Web de acceso libre destinado a la comunidad científica. Su objetivo es compatibilizar formatos de diversos software utilizados en el análisis de redes biológicas (redes génicas, regulatorias, de señales, y/o metabólicas). La aplicación web transforma los archivos con sintaxis tipo CellNetAnalyzer a otros formatos aptos para los programas Cytoscape, Binom, Diva, y MFinder. Ello permite así potenciar las prestaciones obtenibles de cada software.

Además, la aplicación web también computa la matriz de incidencia, siendo ésta un insumo para la posterior determinación de la estructura modular del sistema. La aplicación web desarrollada tiene múltiples ventajas: (a) cubre una necesidad previamente insatisfecha; (b) permite el ingreso y procesamiento de datos de redes de gran tamaño evitando el ingreso individual de datos a través de interfaces gráficas; (c) no se requiere de ningún conocimiento de programación para su utilización; (d) el debugging incorporado localiza y facilita la corrección de errores en el archivo fuente. En conclusión, más investigadores podrán aplicar las herramientas del análisis de redes en los campos de biomedicina y biotecnología, así como su enseñanza en el ámbito universitario. La aplicación web se encuentra accesible para ser utilizada online en la dirección www.desarrollosdg.com.ar/Guebel/archivo1.php

Summary: A web application devoted to the scientific community with interest in the analysis of biological networks is presented. By this open service, different software packages frequently used for the analysis of complex biological networks (genes, regulatory, signals, and/or metabolic) can be reconciled. In this way, the practical scope of software packages is synergized. Hence, files with CellNetAnalyzer syntax can be transformed into other formats by the application developed, thus enabling the usage of the software Cytoscape, Binom, Diva, and MFinder. In addition, the web application also computes the incidence matrix, this being an input for the subsequent determination of the modular structure of the system. The application developed has several advantages: (a) it covers a need previously unsatisfied; (b) allows the entry and processing of data from large networks avoiding individual income data through graphical interfaces; (c) does not require any programming knowledge to use; (d) the built-in debugging locates and facilitates correction of errors in the source file. In conclusion, more investigators can apply this type of analysis in biomedicine and biotechnological research, as well as for teaching these disciplines at the university. The application is available to be used online in the site **¡Error! Referencia de hipervínculo no válida.**

Keywords: Systems biology, Boolean Logic, Multivalued logic, network analysis, software compatibility, Incidence matrix

Introducción

El actual paradigma de investigación biológica ha desplazado su foco de interés desde un enfoque de tipo “reduccionista” (estudio de mecanismos y sub-procesos individuales) hacia otros enfoques de carácter más integrativo y global (estudio de mecanismos y sub-procesos bajo condiciones de interacción potencial y/o efectiva). A su vez, este cambio de paradigma posibilitó la irrupción de nuevos campos disciplinares tales como la Biología de Sistemas, la Biología Integrativa, la Ingeniería Metabólica, y la Biología Sintética (Mardinoglu et al., 2013).

Siendo el concepto de *sistema global* la noción común a todos los nuevos campos arriba mencionados, ello trajo aparejado la necesidad de aplicar métodos formal-matemáticos para su estudio. Los modelos matemáticos pueden apelar eventualmente a diferentes niveles de abstracción. Sin embargo, en cualquier caso, los modelos deben permitir la representación, visualización, simulación y análisis de estructuras complejas. Tal el caso de las denominadas redes génicas, de señales regulatorias, de interacción de proteínas, y/o metabólicas a escala genómica entre otros varios tipos de redes posibles de interés (Berger and Iyengar, 2009).

Emulando los modelos provenientes del campo de la física y de la química, es deseable que los modelos de sistemas biológicos puedan formalizarse a través de sistemas de ecuaciones diferenciales ordinarias (ODE) y/o en derivadas parciales (PDE), o aún como modelos probabilísticos (Markov) según corresponda a la naturaleza del fenómeno analizado. Sin embargo, al expandirse la frontera de lo que se considera el sistema bajo estudio, por su considerable tamaño, hace que dichos métodos “deseables” se revelen en definitiva como prácticamente no-factibles.

Contribuye a ello, el alto grado de multidimensionalidad presente y las muchas dificultades presentes a la hora de parametrizar dichos modelos (Ryll et al., 2011). Esta situación ha llevado al desarrollo de sistemas formales alternativos, tal el caso de los métodos de análisis cualitativos (Boolean Logic) y/o semi-cuantitativos (Multi-valued Logic). Aunque menos potentes, los métodos lógicos tratan de explotar los bien establecidos formalismos matemáticos derivados de la teoría de los grafos y de los hypergrafos respectivamente. Desafortunadamente, aún cuando diversos paquetes de software específicos han sido desarrollados para aplicar los métodos cualitativos y semi-cuantitativos de análisis, una característica importante de los mismos es que dichos programas han sido desarrollados utilizando muy diferentes lenguajes de programación. En algunos casos, además el input a dichos programas está concebido a partir del uso de interfaces gráficas (reacción por reacción), que si bien son muy atractivas cuando se trata de sistemas pequeños, resultan imprácticas cuando se trata de sistemas de alta escala. En los hechos, esta realidad torna dificultosa la posibilidad que un investigador no-experto en informática pueda transformar o adaptar dichos programas a sus propias necesidades si carece de un conocimiento exhaustivo de los diversos lenguajes de programación. En dicho contexto, el servicio Web aquí presentado intenta poner al alcance de los investigadores las múltiples prestaciones que ofrecen los principales software disponibles, utilizando para ello un marco general común, tal que la transición de uno a otro software no revista mayores dificultades operativas.

Fundamentación

Existe una oferta limitada de programas para simular modelos en forma cualitativa. De entre ellos, CellNetAnalyzer (Klamt et al., 2007) es uno de lo más versátiles. En el sitio MUVAl (Guebel et al., 2012a) se encuentra disponible un servicio web que permite transformar un archivo de CellNetAnalyzer en otro que utilizando el mismo software,

permite aplicar lógica multi-valuada extendida con inhibiciones escalonadas (Guebel et al., 2012b). Los análisis de lógica Booleana y/o de lógica multi-valuada pueden también beneficiarse de ciertos estudios complementarios que ofrecen otros programas, tal el caso de Cytoscape (Smoot et al., 2011), BINOM (Bonnet et al., 2013), Diva (Potapov et al., 2008), MFINDER (Alon, 2007). Sin embargo, un inconveniente práctico es que estos programas utilizan un formato de archivo cuya sintaxis difiere de la utilizada por Cell Net Analyzer.

Objetivos

Basados en nuestra experiencia, aquí ofrecemos un sencillo servicio web que permite cumplir tres objetivos:

- (a) Transformar archivos de Cell Net Analyzer hacia el formato que requieren los otros programas.
- (b) Construir los archivos de Cell Net Analyzer por una vía alternativa, ya que, la carga manual de datos mediante la interface normal del programa, no resulta práctica en caso de redes extensas.
- (c) Obtener como resultado de las manipulaciones anteriores la Matriz de Incidencia del sistema, importante insumo para la ulterior realización de otros estudios, tales como el análisis de la estructura modular del sistema.

En definitiva, a través servicio web aquí ofrecido se podrá obtener una plena potenciación entre herramientas de análisis que provienen de diferentes orígenes.

Características de la Aplicación

La aplicación es de código abierto bajo licencia Creative Commons (CC BY-NCND), es decir que es públicamente visible y utilizable pero no puede ser modificada ni utilizada con fines comerciales. La única obligación de los usuarios es citar la herramienta si se ha hecho uso de la misma. La aplicación fue desarrollada en lenguaje PHP v.5, utilizándose como base de datos el sistema MySQL.

Input para la Aplicación

En el sitio web (<http://www.desarrollosdg.com.ar/Guebel/archivo1.php>) se encuentra disponible un manual para el usuario conteniendo documentación bilingüe (español/inglés) con instrucciones precisas acerca de cómo se debe operar el programa. Por dicho motivo esta información no será repetida aquí. Baste decir que como input, el programa sólo requiere la disponibilidad del modelo escrito respetando la sintaxis del CellNetAnalyzer pero bajo la forma de un archivo **.xls** (Excel formato) (ver Tabla 1).

Tabla 1: Ejemplo de pequeño archivo de prueba en formato **.xls** utilizado como input para el programa de conversión.

| | |
|-----|---|
| R80 | FOXO(c) + JNK1,2*(c) + MdmXP(Ser367)(c) + !NLK*(n) + !RTK*(pm) + !HAUSP(n) + !FOXO_p14-3-3(c) = PPFOXOUb(n) |
| R81 | c-Met*(pm) + NOD*(pm) + DUOX2(pm) = ROS(c) |
| R82 | Betacatenin(n) + PPFOXOUb(n) + p300P(Ser1834)(n) + !HAUSP(n) + !Sirt1*(n) = FOXO_Betacatenin(n) |
| R83 | PPFOXOUb(n) + p53Tetra(n) + p300P(Ser1834)(n) + !HAUSP(n) + !Sirt1*(n) = FOXO_p53(n) |
| R84 | TCF4E(n) + SOX17(n) + Proteasome(26S) = TCF4E(catabolized) |
| R85 | PIN1(n) = HAUSP(n) |

Observar que la primer columna de la Tabla 1 corresponde la identificación (ID) de cada una de las seis reacciones consideradas (R80-R85). La segunda columna de la Tabla 1 corresponde a las reacciones que conforman la red. Las especies químicas con rol inhibitorio (o eventualmente represor), –siguiendo la sintaxis del programa CellNetAnalyzer–, son señalados precediendo su nombre mediante un signo (!).

Así, la reacción R80 de la Tabla 1 indica que el factor de transcripción FOXO es inactivo mientras se localice en el citoplasma (FOXO(c)), pero se activa cuando es translocado al núcleo. La translocación sólo ocurre si FOXO inactivo es fosforilado por JNK activado (JNK1,2*(c)) y ubiquitinado por MdmXP(Ser367)(c) generándose entonces la especie activa PPFOXOUb(n). El proceso de activación de FOXO es modulado negativamente por cuatro especies químicas o eventos (NLK*, RTK*, HAUSP, y FOXO_p14-3-3). Fisiológicamente, ello se corresponde con la activación de proteína Nemo-like kinasa (NLK*), la activación de los receptores tirosina kinasa (RTK*), la presencia de la enzima deubiquitinasa HAUSP en el núcleo, y el secuestro del factor FOXO por formación de un complejo con la proteína 14-3-3 en el citoplasma (FOXO_p14-3-3(c)).

Se debe tener en cuenta que las situaciones de múltiple inhibición son muy frecuentes en la redes biológicas, tanto más cuánto más global sea el enfoque con que se analiza la red. A pesar de la importancia del fenómeno, dicha situación no puede ser tratada adecuadamente mediante lógica Booleana, como tampoco por lógica multi-valued convencional. De hecho, según se asuma un mecanismo de inhibición concertada o independiente, podría darse el caso que la reacción nunca resulte inhibida a pesar de la presencia de inhibidores (falsos negativos), o por el contrario siempre resulte inhibida aún cuando los niveles de los inhibidores fueran muy bajos (falsos positivos). Para evitar dicha situación puede aplicarse la denominada lógica multivaluada “extendida” con inhibiciones graduales aún utilizando el mismo software CellNetAnalyzer (Guebel et al., 2012b; web service MUVAl <http://calsys.informatik.uni-rostock.de/muval11/index.html>).

Observar en la Tabla 1 que todas aquellas especies químicas situadas a la izquierda del signo (=) indican los substratos y/o moduladores de cada reacción, en tanto que la única especie situado a la derecha del signo (=) indica el producto de la reacción. Si en una reacción se formaran más de un producto, la reacción deberá desdoblarse repitiendo la parte izquierda de la ecuación tantas veces como se forme un producto diferente, y colocando en la parte derecha de la ecuación los diferentes productos formados para una misma ecuación.

En la Tabla 1 el símbolo (l) denota el final de cada reacción. Las letras entre paréntesis adyacente a los nombres de las especies químicas denotan compartimento celular donde se localiza la especie química considerada. Así por ejemplo, (n): núcleo; (c): citoplasma; (pm): membrana plasmática; (m): mitocondria, etc. Para más detalles ver manual de CellNetAnalyzer (<http://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html>).

Outputs de la Aplicación

Como output del programa se genera un archivo con formato **CSV** (comma separated values). Desde allí el usuario puede fácilmente seleccionar la sección de resultados de su interés y exportarla a archivos tipo Excel. Estos archivos serán luego utilizados para correr otros programas tales como Cytoscape (ver Tabla 2) o MFinder (ver Tabla 4). Observar que output del programa M-finder (Tabla 4) no puede ser interpretada sin las equivalencias establecidas en la Tabla 3. Otro importante output que genera la aplicación es la matriz de incidencia del sistema (ver Tabla 5).

Tabla 2: Output compatible con programas Cytoscape-Binom-Diva que se obtiene a partir de correr el archivo de prueba.xls.

| Sustrate | Interaction | Product |
|-------------------|-------------|---------------------|
| FOXO(c) | positive | PPFOXOUb(n) |
| JNK1,2*(c) | positive | PPFOXOUb(n) |
| MdmXP(Ser367)(c) | positive | PPFOXOUb(n) |
| NLK*(n) | negative | PPFOXOUb(n) |
| RTK*(pm) | negative | PPFOXOUb(n) |
| HAUSP(n) | negative | PPFOXOUb(n) |
| FOXO_p14-3-3(c) | negative | PPFOXOUb(n) |
| c-Met*(pm) | positive | ROS(c) |
| NOD*(pm) | positive | ROS(c) |
| DUOX2(pm) | positive | ROS(c) |
| Betacatenin(n) | positive | FOXO_Betacatenin(n) |
| PPFOXOUb(n) | positive | FOXO_Betacatenin(n) |
| p300P(Ser1834)(n) | positive | FOXO_Betacatenin(n) |
| HAUSP(n) | negative | FOXO_Betacatenin(n) |
| Sirt1*(n) | negative | FOXO_Betacatenin(n) |
| PPFOXOUb(n) | positive | FOXO_p53(n) |
| p53Tetra(n) | positive | FOXO_p53(n) |
| p300P(Ser1834)(n) | positive | FOXO_p53(n) |
| HAUSP(n) | negative | FOXO_p53(n) |
| Sirt1*(n) | negative | FOXO_p53(n) |
| TCF4E(n) | positive | TCF4E(catabolized) |
| SOX17(n) | positive | TCF4E(catabolized) |
| Proteasome(26S) | positive | TCF4E(catabolized) |
| PIN1(n) | positive | HAUSP(n) |

Cabe señalar que la aplicación está programada para excluir automáticamente las interacciones redundantes que surgen del proceso de desagregación que opera en las Tablas 2 y Tabla 4. Dicho filtrado es importante, ya que las interacciones redundantes constituyen una insidiosa interferencia que luego no es eliminado por los programas Cytoscape, Binom, DIVA, o M-finder.

Otro uso práctico importante que surge del servicio Web implementado, es que la lista filtrada de metabolitos en la Tabla 3 puede utilizarse además para construir el denominado “archivo de metabolitos” que CellNetaAnalyzer requiere para poder correr.

La obtención de dichos listado es sumamente engorrosa cuando debe hacerse en forma manual teniendo en cuenta que la red puede estar constituida por varios cientos o miles de metabolitos.

Tabla 3: Output con equivalencias entres códigos alfanuméricos (requeridos por CellNetAnalyzer-Cytoscape-Binom-Diva) y códigos numéricos (requeridos por el programa M-Finder).

| | |
|---------------------|----|
| FOXO(c) | 1 |
| JNK1,2*(c) | 2 |
| MdmXP(Ser367)(c) | 3 |
| NLK*(n) | 4 |
| RTK*(pm) | 5 |
| HAUSP(n) | 6 |
| FOXO_p14-3-3(c) | 7 |
| c-Met*(pm) | 8 |
| NOD*(pm) | 9 |
| DUOX2(pm) | 10 |
| Betacatenin(n) | 11 |
| PPFOXOUb(n) | 12 |
| p300P(Ser1834)(n) | 13 |
| Sirt1*(n) | 14 |
| p53Tetra(n) | 15 |
| TCF4E(n) | 16 |
| SOX17(n) | 17 |
| Proteasome(26S) | 18 |
| PIN1(n) | 19 |
| ROS(c) | 20 |
| FOXO_Betacatenin(n) | 21 |
| FOXO_p53(n) | 22 |
| TCF4E(catabolized) | 23 |

Tabla 4. Desagregación del archivo de prueba.xls utilizado como input, siguiendo la nomenclatura establecida en Tabla 3 para así cumplir los requerimientos del programa M-finder.

| Species | Interaction | Species |
|---------|-------------|---------|
| 1 | positive | 12 |
| 2 | positive | 12 |
| 3 | positive | 12 |
| 4 | negative | 12 |
| 5 | negative | 12 |
| 6 | negative | 12 |
| 7 | negative | 12 |
| 8 | positive | 20 |
| 9 | positive | 20 |
| 10 | positive | 20 |
| 11 | positive | 21 |
| 12 | positive | 21 |
| 13 | positive | 21 |
| 6 | negative | 21 |
| 14 | negative | 21 |
| 12 | positive | 22 |
| 15 | positive | 22 |
| 13 | positive | 22 |
| 6 | negative | 22 |
| 14 | negative | 22 |
| 16 | positive | 23 |
| 17 | positive | 23 |
| 18 | positive | 23 |
| 19 | positive | 6 |

Tabla 5: Matriz de incidencia computada a partir del archivo de prueba.xls utilizado como input.

| Nombre/ID | R80 | R81 | R82 | R83 | R84 | R85 |
|---------------------|-----|-----|-----|-----|-----|-----|
| FOXO(c) | 1 | 0 | 0 | 0 | 0 | 0 |
| JNK1,2*(c) | 1 | 0 | 0 | 0 | 0 | 0 |
| MdmXP(Ser367)(c) | 1 | 0 | 0 | 0 | 0 | 0 |
| NLK*(n) | 1 | 0 | 0 | 0 | 0 | 0 |
| RTK*(pm) | 1 | 0 | 0 | 0 | 0 | 0 |
| HAUSP(n) | 1 | 0 | 1 | 1 | 0 | 1 |
| FOXO_p14-3-3(c) | 1 | 0 | 0 | 0 | 0 | 0 |
| c-Met*(pm) | 0 | 1 | 0 | 0 | 0 | 0 |
| NOD*(pm) | 0 | 1 | 0 | 0 | 0 | 0 |
| DUOX2(pm) | 0 | 1 | 0 | 0 | 0 | 0 |
| Betacatenin(n) | 0 | 0 | 1 | 0 | 0 | 0 |
| PPFOXOUb(n) | 1 | 0 | 1 | 1 | 0 | 0 |
| p300P(Ser1834)(n) | 0 | 0 | 1 | 1 | 0 | 0 |
| Sirt1*(n) | 0 | 0 | 1 | 1 | 0 | 0 |
| p53Tetra(n) | 0 | 0 | 0 | 1 | 0 | 0 |
| TCF4E(n) | 0 | 0 | 0 | 0 | 1 | 0 |
| SOX17(n) | 0 | 0 | 0 | 0 | 1 | 0 |
| Proteasome(26S) | 0 | 0 | 0 | 0 | 1 | 0 |
| PIN1(n) | 0 | 0 | 0 | 0 | 0 | 1 |
| ROS(c) | 0 | 1 | 0 | 0 | 0 | 0 |
| FOXO_Betacatenin(n) | 0 | 0 | 1 | 0 | 0 | 0 |
| FOXO_p53(n) | 0 | 0 | 0 | 1 | 0 | 0 |
| TCF4E(catabolized) | 0 | 0 | 0 | 0 | 1 | 0 |

Como se ve en Tabla 5, la matriz de incidencia codifica la participación de las diversas especies química en cada una de los procesos o reacciones del sistema. Como tal, a diferencia de lo que sucede con los softwares que sólo trabajan sobre la base de lógica Booleana (Cytoscape, Diva, M-Finder), la matriz de incidencia es apta para el análisis de los hipergrafos ya que las reacciones individuales no son desagregadas en sub-reacciones elementales (ver [Guebel et al, 2012b](#)).

Experiencia con el sitio web de la aplicación

De acuerdo a los registros obtenidos mediante Google Analytics, en el mes inmediato anterior a esta presentación, el sitio fue visitado por 11 usuarios internacionales: 5 españoles, 5 norteamericanos, 1 Gran Bretaña. Mediante la misma herramienta se pudo además verificar que en el período 2013-2014 la mayoría de las visitas culminaron en la utilización del programa (ver Fig. 1).



Fig. 1. Mapeo de las actividades desarrolladas por las visitas en el sitio web de la aplicación durante el último mes (20 de Marzo-19 de Abril 2014), atendiendo al país de procedencia de la PC utilizada (Fuente: Google Analytics).

Así pudo establecerse que el programa fue visitado/utilizado por gente de diversas universidades alemanas (Berlin, Rostock, Freiburg, Friederich Alexander), estadounidenses (Palo Alto, Maryland, Baltimore), británicas (Manchester), españolas (CSIC, Madrid, Andalucía, Barcelona), belgas (Leuven), colombianas (Universidad de Los Andes), brasileñas y chilenas entre otras identificables.

Contribución de los autores: Maria Dolores Garcia es la responsable de la programación y mantenimiento del sitio web. Daniel Victor Guebel concibió y supervisó el desarrollo de la aplicación. Ambos autores contribuyeron a la redacción de esta presentación.

Bibliografía

- Alon U, Network motifs: theory and experimental approaches. Nature Reviews Genetics (2007), 8, 450-461
- Berger S.I., Iyengar R. Network analyses in systems pharmacology. Bioinformatics (2009), 25: 2466-2472
- Bonnet E., Calzone L., Rovera D., Stoll G., Barillot E, Zinovyev A. BiNoM 2.0, a Cytoscape plugin for accessing and analyzing pathways using standard systems biology formats. BMC Systems Biology (2013), 7: 18
- Guebel DV, Schmitz U, Wolkenhauer O, Vera J. Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach. Mol.BioSyst (2012a), 8: 1230-1242
- Guebel DV, Schmitz U, Wolkenhauer O, Vera J. Supplementary Materials of Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach. Mol. BioSyst (2012b), 8: 1230-1242.
- Klamt S, Saez-Rodriguez J, Gilles E D. Structural and functional analysis of cellular networks with CellNetAnalyzer.(2007), BMC Systems Biology, 1:2
- Mardinoglu A., Gatto F., Nielsen J. Genomic scale modeling of human metabolism –a system biology approach. Biotechnology Journal (2013), 8: 985-996
- Potapov AP, Goemann B, Wingender E. The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks. BMC Bioinformatics (2008), 9:227
- Ryll A, Samaga R, Schaper F, Alexopoulos LG, Klamt S. Large-scale network models of IL-1 and IL-6 signalling and their hepatocellular specification. Mol.Biosyst. (2011), 7: 3253-70.
- Smoot M, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics (2011), 27: 431–432.

La aplicación Web desarrollada se encuentra referenciado desde los siguientes sitios:

- Plugins for CellNetAnalyzer
(<http://www2.mpi-magdeburg.mpg.de/projects/cna/plugins.html>)
- ResearchGate
(https://www.researchgate.net/profile/Daniel_Guebel?ev=hdr_xprf)