

Instituto Nacional de Bioinformática (INB) y Centro Nacional de Investigaciones Oncológicas (CNIO)

"Si lográramos hacer bioinformática teórica probablemente estaríamos muy cerca de curar enfermedades hasta ahora incurables"

José María Fernández González forma parte del Nodo 2 del INB. Antes de comenzar a trabajar para este centro, apenas había oído hablar de la bioinformática, quizás por ello su afán en darla a conocer a través de su blog en [madri+d](http://www.madridmasd.org/blogs/bioinformatica) (www.madridmasd.org/blogs/bioinformatica) abierto desde 2005 y en el que participa activamente.



José María Fernández González

José María Fernández llegó al mundo de la bioinformática sin proponérselo, pero son ya 12 años de experiencia los que acumula. Tras estudiar ingeniería informática en Málaga, y gracias a la insistencia de uno de sus profesores, se presentó y consiguió la plaza en la que hoy desarrolla su labor investigadora. Consciente de que, como él hace una década, somos muchos los que nada sabemos de esta rama de la ciencia, ejerce una importante labor divulgativa desde el blog de bioinformática de esta misma Web. Hoy, a través de esta entrevista, intentaremos descifrar en qué consiste y qué trata esta ciencia.

Montserrat Alvarez

"A grandes rasgos podemos definir la bioinformática como la informática aplicada a la resolución de problemas biológicos"- nos cuenta-.

"Por ponerte un ejemplo, ahora está muy de moda hablar de la *ultrasecuenciación*. En esencia esto significa poder secuenciar las secuencias genómicas de una muestra de un paciente (normalmente de cáncer), y diagnosticarlo de forma especializada de manera que no se le dé un medicamento 'general' sino uno 'particular' adaptado a su caso; un medicamento *a la carta* diseñado especialmente para él.

Pues bien, para poder llegar a esto, obviamente hay que hacer experimentos, pero también es necesario analizar todos los datos que se producen durante esa investigación y ahí es donde los bioinformáticos tenemos un papel relevante porque somos quienes, de toda esa marea de información que surge, extraemos lo que realmente es importante.

Por ejemplo, cuando se secuenció el genoma humano se consideró que había entre 40.000 y 50.000 genes (a día de hoy todavía no sabe el número exacto). ¿Imaginas la cantidad de información que nació de esta investigación? Pero durante la misma también fue necesario consultar y acudir a fuentes como bases de datos biológicos donde poder contrastar informaciones, compararlas, buscar pistas.

El problema es que hay tal cantidad de información que a simple vista sería imposible sacarle partido. Es como si estuvieras en un inmenso campo de amapolas y a necesitaras detectar 3 o 4 margaritas. Nosotros ayudamos a localizarlas. Somos la base estadística; nos basamos en información ya existente pero contrastada. Por supuesto, al final es un experto en la enfermedad el que toma la decisión de aplicar un tratamiento u otro, pero nosotros somos un apoyo sustancial en la toma de esa decisión."

Intuyo entonces que una de sus labores fundamentales es el tratamiento y la discriminación de información y así se lo sugiero.

"Por supuesto, pero no sólo eso - me corrige-. Hay que saber buscar, pero normalmente con eso sólo no es suficiente. También es necesario haber almacenado la información, y haber corregido errores para que las deducciones que se vayan a realizar a partir de esa información no sean equivocadas.

Te explicaré uno de los casos más comunes a lo que nos enfrentamos: secuenciamos un nuevo organismo y obtenemos nuevas proteínas, pero necesitamos saber cómo se comportan los genes de ese organismo. ¿Cómo podemos hacerlo? Pues por similitud con lo que ya se conoce, por comparación: ya que todavía no somos capaces de saber cómo se comporta un gen a partir de la secuencia del mismo, tenemos que hacerlo a través de la secuencia de otro gen parecido que sí sabemos cómo se comporta. Y cuando no podemos hacer esta comparación, recurrimos a otras fuentes como puede ser la literatura científica".

Pero entonces, os encontráis con el problema que me comentabas anteriormente de la sobreinformación- le apunto.

"Exacto, hay muchísima información y además no se puede extraer de forma sistematizada porque está preparada para que sea leída por humanos y no por máquinas. Verás, **PubMed** es una base de datos de artículos científicos que tiene más de 20 millones de registros, ¡y es de las pequeñas! Pues bien, aunque tienes la introducción a cada uno de los artículos, te puedes encontrar con problemas tan clásicos como que se escribió mal el nombre de un gen, o que el mismo gen se nombra de formas distintas según el país, o que la nomenclatura de los reactivos es diferente, etc. Todo esto hace que al buscar la información sobre ese gen, no se encuentre toda, que se produzcan errores de identificación, etc."

Proyectos e investigaciones

Una vez acotado y definido el campo de trabajo de un bioinformático, me intereso ahora por las investigaciones que está llevando a cabo nuestro experto.

"Yo trabajo para el Instituto Nacional de Bioinformática, que podemos definirlo como virtual porque está repartido por varios grupos de investigación por toda España; en total somos unas 40 personas.

Personalmente estoy inmerso en varios proyectos. El primero está relacionado con la búsqueda y el análisis de información sobre genes de humanos, de rata y de ratón.

Hace años diseñé unos servicios web que están siendo muy utilizados, los **IHOP's**; que podían ser utilizados por cualquier persona pero no se podían automatizar.

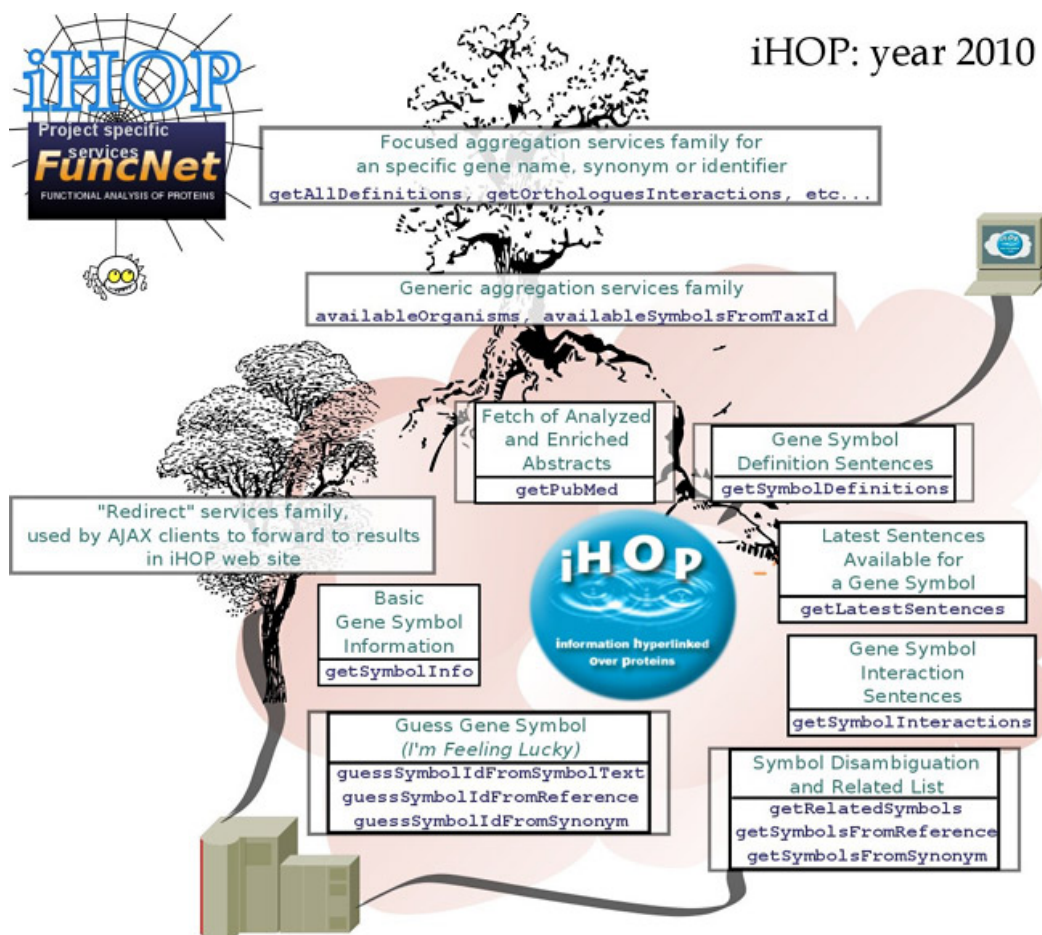
Para subsanar este inconveniente he diseñado **Web Services** que sí permiten utilizarlo de forma sistematizada: básicamente lo que he hecho ha sido mandar una lista de nombres de genes, tanto de humanos como de ratas, buscar en la literatura científica y en ortólogos y ver toda la información relacionada. ¿Con qué fin? Si los genes

Los genes ortólogos tienen un origen común: Hace millones de años un organismo tenía un gen. Debido a la evolución de ese organismo se originaron otros 2, 3 o 4; sus genes, aunque básicamente son iguales al primero, tienen modificaciones.

tienen un origen común, es de suponer que tienen una secuencia conocida, y si tienen una secuencia conocida seguramente funcionarán de forma similar.

Otro de mis proyectos tiene que ver con la **Evaluación de los distintos sistemas de anotación de funciones (Sistema GOPHER)**. Verás, un sistema de anotación de función toma una secuencia de proteínas o toma una estructura de proteínas de la cual no se conoce su función y se le asigna una función de forma sistematizada. Cuando hay una proteína parecida darle una asignación es fácil (por similitud), pero no lo es tanto cuando no hay ninguna similar. El sistema que estoy desarrollando es para poder evaluar 'cómo de bien' lo hace cada uno de los programas existentes para realizar esas anotaciones y hacerlo de forma continuada.

Además de todo esto, dentro de la **Red de Excelencia ENFIN**, en la que también colaboro, he trabajado en dos subproyectos, uno relacionado con el **Estudio de red de interacciones de las proteínas relacionadas con angiogénesis** (generación de tumores) y otro sobre la **Fase final de la condensación cromosómica celular**; cuando una célula va a realizar la mitosis para crear otra célula idéntica, en la fase final intervienen unas proteínas que ayudan a separar las copias de los cromosomas para que se pueda hacer la mitosis; pues bien estas proteínas son muy poco conocidas porque son muy difíciles de estudiar. La gran suerte es que uno de los participantes, con una nueva técnica experimental, va a poder validar nuestros resultados".



Aportaciones de la informática a la biología

A modo de resumen, le pido a José María que me de su opinión sobre las interacciones de su especialidad en el avance y desarrollo de la biología.

Nuestro experto reconoce tres claves.

"Lo primero que ha hecho es acortar los tiempos de investigación. Esto se ha notado sobre todo a nivel de genoma humano; la primera vez que se hizo una secuenciación del genoma humano costó muchísimos años y muchísimo dinero, ahora gracias al trabajo en colaboración de las técnicas experimentales y de las técnicas informáticas se puede realizar en sólo 2 días.

Pero también ha ayudado a resolver un problema muy acuciante en biología: la sobreinformación. Antiguamente había muchas preguntas y pocos datos, ahora hay demasiados datos. En esa marea la bioinformática ayuda organizar, a buscar y a deducir.

Por último, puedo decirte que tampoco la estructura tridimensional que conocemos de las proteínas podrían calcularse sin la ayuda de los ordenadores".

Futuro de la bioinformática

No quiero finalizar esta entrevista sin preguntar a nuestro experto por el futuro de la bioinformática. Esto es lo que nos cuenta.

"Lo ideal sería hacer bioinformática teórica pero es muy difícil porque hacer una simulación completa del entorno es complicado; si lo consiguiéramos ahorraríamos muchos experimentos y con ello tiempo y dinero.

Y la razón por la que no se puede hacer es porque no es fácil explicar un mundo en el que los conceptos van cambiando y evolucionando.

Hace 10 años se empezaba a estudiar la posibilidad de que se podría producir, a partir de un mismo gen, varias proteínas, pero no se esperaba que hubiera tantos genes en los que esto fuera posible (algunos estudios recientes indican que más del 90%). En biología molecular los conceptos van cambiando y de forma muy rápida, hay conceptos que son esperables pero otros no, y esto te puede modificar toda la línea de trabajo.

Por otro lado, si bajamos al nivel más básico, al atómico, es muy difícil simularlo, porque, entre otras cosas, no se pueden simular todos los efectos y hay que recurrir a campos de fuerza simulados y para ello también hace falta tener procesadores muy potentes.

Si consiguiéramos hacer bioinformática teórica, probablemente, estaríamos cerca de curar enfermedades hasta ahora incurables porque podríamos empezar a simular organismos completos (no sólo células) y ver que comportamiento tendrían de forma global ante determinada enfermedad y ante determinados fármacos.

Pero mientras tanto, debemos interesarnos en resolver problemas clásicos no resueltos como el ***Plegamiento de las proteínas*** (conseguir que de la secuenciación de la proteína se obtenga su estructura) un objetivo viable pero que, hoy por hoy, requiere de mucha potencia computacional, o la ***Anotación de función de proteína*** (resolver para qué sirve una proteína dentro de la tareas de un organismo).

Instituto Nacional de Bioinformática (INB) y Centro Nacional de Investigaciones Oncológicas (CNIO)

CENTRO

NODO 2 del Instituto Nacional de Bioinformática (INB) y del Centro Nacional de Investigaciones Científicas Oncológicas (CNIO)

Líneas de Investigación

Bioinformática, biología computacional e integración de datos.

Personal

Investigador: José María Fernández González

Datos de Contacto:

Dirección: c/ Melchor Fernández Almagro, 3
28029 Madrid

Teléfono: 91 732 80 00

e-mail: jmfernandez@cnio.es

Web: www.madrimasd.org/blogs/bioinformatica