

Tendencias

Hacia una economía del conocimiento

JOSEP CORBELLA
Barcelona

Barcelona, tenemos un problema: el volumen de datos biológicos que se está obteniendo en centros de investigación y hospitales está creciendo más rápido que la capacidad de almacenarlos y digerirlos. Llegará un punto, a medio plazo, en que la capacidad de los ordenadores no dará abasto para gestionar la avalancha de datos que generan médicos y biólogos.

Para una ciudad que se mira en el espejo de Boston para construir una economía del conocimiento, y que cuenta ya con algunos de los equipos de bioinformáticos más importantes de Europa, “resolver este problema es una cuestión estratégica”, afirma Modesto Orozco, bioquímico de la Universitat de Barcelona (UB) e investigador del Institut de Recerca Biomèdica (IRB) y del Barcelona Supercomputing Center (BSC).

El problema, con el que no sólo se encuentra Barcelona sino todas las capitales de investigación biomédica, se explica por el drástico abaratamiento de secuenciar genomas que se ha registrado en los últimos cinco años. Si en el

Barcelona, capital GENÓMICA

La caída del coste de investigar el ADN provoca una avalancha de datos biológicos

2007 el coste de secuenciar cada genoma humano era de diez millones de dólares, la llegada de una nueva tecnología de secuenciación ha hecho que el precio haya caído ahora por debajo de los 10.000 dólares. Este abaratamiento ha llevado a que un número creciente de investigadores y de médicos hayan empezado a secuenciar genomas en cuanto han podido permitírselo.

Hoy día se secuencian genomas de células tumorales, de variedades de melón, de fósiles de neandertales... En el caso del cán-

cer, se quieren comparar los genomas de células tumorales con los de células sanas del mismo paciente para ver cuáles son los genes que están alterados en el tumor. En un futuro próximo, se prevé poder comparar el genoma de células tumorales antes y después del tratamiento para evaluar la eficacia de los distintos fármacos... Y no sólo se almacenan datos de genomas. También se han empezado a estudiar epigenomas (que indican qué genes están activos y cuáles están silenciados en un momento dado en una

célula concreta); proteomas (que indican qué proteínas hay en una célula); transcriptomas (que indican qué ARN hay)... En fin, una cantidad enorme de datos que van camino de desbordar la capacidad de los superordenadores.

Según la famosa ley de Moore, la capacidad de almacenamiento de datos de los chips se duplica cada 18 meses. Pero según una estimación del Institut Europeu de Bioinformàtica, con sede en el Reino Unido, la cantidad de datos biológicos que se almacenan se duplica cada nueve meses.

“Si el volumen de datos aumenta más rápido que la capacidad de almacenarlos, nos encontramos con un cuello de botella que limitará la investigación y dificultará que estos datos se apliquen en beneficio de los pacientes”, advierte Ivo Gut, director del Centro Nacional de Análisis Genómico (CNAG) en Barcelona.

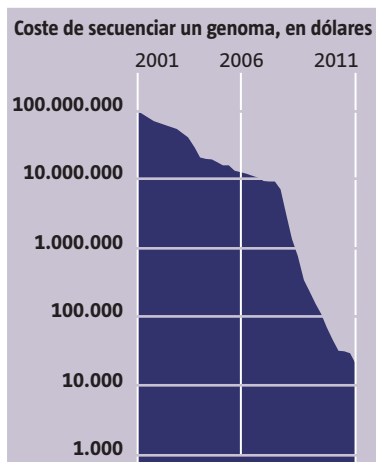
La solución, según Modesto Orozco, requiere gestionar los datos de la manera más eficiente posible. Si hasta ahora no se ha hecho, es porque la secuenciación de genomas es una tecnología joven con la que cada centro de investigación ha empezado a experimentar por su cuenta. Pero, llegados a la actual crisis de crecimiento de esta tecnología, “es mejor compartir recursos que duplicarlos”, sostiene Orozco.

Compartir recursos significaría establecer un banco central de datos biológicos que guardara la información de los distintos hospitales y centros de investigación. “No tiene sentido –según Orozco– que cada institución gestione sus propios datos de manera aislada; sería mucho más eficiente que estuviéramos todos conectados”.

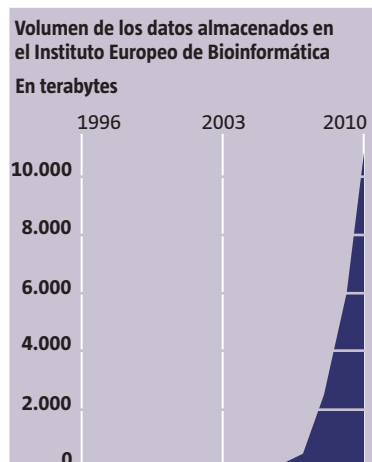
El Biodata Center –como lo llama Orozco de manera provisional– estará vinculado al superordenador Mare Nostrum en el

Los ordenadores no dan abasto para gestionar la avalancha de datos que aportan los biólogos

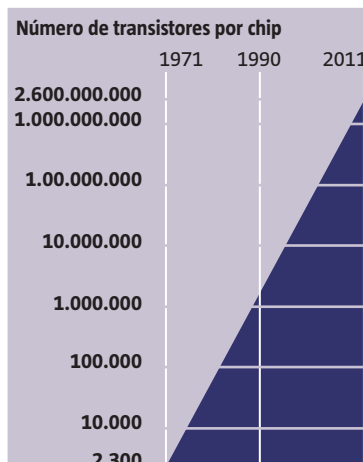
Como el precio de secuenciar genomas cae en picado desde el 2007...



... el volumen de datos biológicos se multiplica por dos cada nueve meses



Sin embargo, la capacidad de almacenar datos en un chip se duplica cada 18 meses



Por lo tanto, el aumento de la capacidad de almacenamiento no será suficiente para gestionar todo el volumen de datos biológicos

FUENTE: Elixir / BSC

LA VANGUARDIA

Trece países crean un consorcio bioinformático europeo

■ Trece países europeos –España entre ellos– han creado el consorcio Elixir para gestionar la avalancha de datos biológicos derivados de la revolución genómica. El plan de trabajo prevé que Elixir se construirá como una infraestructura con un *hub* principal situado en el Reino Unido –el Estado que más invierte en el proyecto y el que lidera la investigación genómica en Europa– y varios

nodos distribuidos por toda Europa. En España, están involucrados en el proyecto Elixir el Centro Nacional de Investigaciones Oncológicas (CNIO) y Barcelona Supercomputing Center.

“El reto que plantea la avalancha de datos biológicos es demasiado grande para que ningún país lo pueda afrontar por sí solo”, declaró el 3 de septiembre Soren Brunak, presidente de

la junta directiva de Elixir.

La fase preparatoria del proyecto se inició en el 2007 con financiación de la Comisión Europea y concluirá en diciembre de este año. El calendario de trabajo prevé que la fase de construcción de la infraestructura dure cuatro años y termine en diciembre del 2016. A partir del año 2017 se iniciará la fase permanente de explotación de la red Elixir.



Modesto Orozco, en el Barcelona Supercomputing Center

LOS ESTADOS QUE MÁS APORTAN AL PROYECTO ELIXIR (en euros)

REINO UNIDO
100 millones

FINLANDIA
6,85 millones

NORUEGA
6,5 millones

ESPAÑA
5 millones

DINAMARCA
5 millones

SUECIA
1,7 millones

BSC, en el Campus Nord de la Universitat Politècnica. Pero para que el proyecto sea viable, será necesario que la conexión entre el BSC y los centros de investigación y hospitales se haga por una red de fibra óptica con un gran ancho de banda que actualmente no existe.

“Sería interesante tener un repositorio centralizado de datos, pero para hacer buena investiga-

ción es imprescindible poder tener un acceso muy rápido a los datos”, explica Joan Seoane, investigador del Instituto de Oncología de Vall d’Hebron (VHIO).

“Tiene sentido compartir recursos”, coincide Ivo Gut. “Pero incluso así, y con una transmisión de datos rápida, seguiremos necesitando nuestros propios ordenadores para tener una cierta autonomía respecto al BSC y poder priorizar las investigaciones que consideremos prioritarias”.

Actualmente el BSC ya está conectado por fibra óptica al CNAG, lo que ha facilitado la secuenciación del genoma de la leucemia entre ambas instituciones. Pero aún no lo está a otros centros de investigación biomédica ni a los grandes hospitales. Establecer esta red requeriría una inversión de varios millones de euros.

Llegados a este punto, la pregunta clave es si resultará más costoso construir la red de fibra óptica o no construirla. La respuesta depende, según Orozco, de lo importante que se considere la investigación biomédica para el país. “A corto plazo –sostiene–, podemos seguir como hasta ahora. Pero en tres o cuatro años, si no hacemos algo para evitarlo, llegaremos a una situación de colapso de los datos biológicos”.

PROBLEMA EMERGENTE

El volumen de datos biológicos aumenta más rápido que la capacidad informática

POSIBLES CONSECUENCIAS

El problema puede dificultar que los datos genómicos beneficien a los pacientes

SOLUCIÓN PROPUESTA

El BSC propone compartir recursos de computación entre centros científicos



LAURA GUERRERO

“Necesitamos bioinformáticos y no los estamos formando”

Roderic Guigó, coinvestigador principal del proyecto Encode

J. CORBELLA Barcelona

Roderic Guigó tuvo la gran suerte de que no le salieron bien dos de los primeros experimentos que hizo cuando empezaba su carrera investigadora. Decidió que el laboratorio no era lo suyo, que le iba más el trabajo teórico, y apostó por la bioinformática cuando aún nadie hablaba de bioinformática. Supo intuir en qué dirección iba a cambiar la biología y la apuesta le salió bien. Lideró el único equipo de investigación español que participó en la histórica secuenciación del genoma humano. Después tuvo una participación destacada en la secuenciación de otros genomas. Y, cuando se lanzó el proyecto Encode como continuación del proyecto Genoma humano en el 2003, fue seleccionado –y financiado por el Gobierno de Estados Unidos– como uno de los investigadores principales del proyecto. Investigador del Centre de Regulació Genòmica (CRG) y profesor de la Universitat Pompeu Fabra (UPF), está considerado uno de los mejores bioinformáticos de Europa.

¿Cómo se hace uno bioinformático?

En España es difícil. En mi caso, la trayectoria de investigación de mi equipo se ha basado siempre en la genómica computacional. Aquí desarrollamos un software que se utilizó para identificar genes en la secuenciación del genoma humano. Después tuvimos una participación importante en el genoma del ratón. Y cuando se inició el proyecto Encode, que era un proyecto competitivo abierto a investigadores de todo el mundo, era una continuación lógica del trabajo que habíamos hecho.

¿Por qué dice que en España es difícil?

Porque aquí, a diferencia de lo que ocurre en otros países, no se ofrece la posibilidad de estudiar al mismo tiempo biología e informática, lo cual es un error estratégico. España necesita bioinformáticos y no los está formando.

¿Para qué los necesita?

En el mundo de la biología, hasta hace poco la mayor parte del trabajo consistía en obtener da-

tos. Ahora se producen de forma automática cantidades ingentes de datos y el gran problema consiste en interpretarlos. Y lo mismo va a ocurrir pronto en el mundo de la medicina. La importancia de la computación en biología y en medicina es cada vez mayor. Por eso hay una alta demanda de personal cualificado en bioinformática y biología computacional.

¿Y si no se forma este personal?

Tendremos las herramientas para obtener los datos, como secuenciar el genoma de cada uno de nosotros, que cada día es más barato, pero no podremos utilizar estos datos porque no sabremos qué hacer con ellos. Es decir, tendremos los datos pero no sabremos interpretarlos.

precedente de la doble titulación en física y matemáticas, en la que se matriculan estudiantes altamente motivados. Habría que ofrecer también una doble titulación de biología e informática.

¿No sería mejor crear una carrera de bioinformática?

Es otra posibilidad, pero en mi opinión una superespecialización de los estudiantes desde los primeros cursos de carrera, como se ofrece ahora, es la dirección contraria de hacia dónde deberíamos ir. En un mundo rápidamente cambiante, deberíamos ofrecer carreras flexibles y formar a los estudiantes para que puedan adaptarse a los cambios.

¿Qué aptitudes hay que tener para ser buen bioinformático?



JORDI PLAY

Roderic Guigó, en el Centre de Regulació Genòmica

PROFESIÓN CON FUTURO

“Hay una alta demanda de personal cualificado en bioinformática”

FALLA LA UNIVERSIDAD

“Tendría que haber la opción de estudiar al mismo tiempo informática y biología”

¿Solución?

Habría que dinamitar el sistema universitario que tenemos, esto sería lo ideal. Hacerlo menos rígido, y mucho más libre y flexible.

¿Y si lo ideal no es posible?

Por lo menos tendría que haber la posibilidad de estudiar al mismo tiempo informática y biología, una opción que hoy día no existe en España. Hay el

Hay dos perfiles principales. Uno es el de personas fascinadas por la biología y a las que al mismo tiempo se les dan bien las matemáticas y la informática. Encode, por ejemplo, es sobre todo un proyecto para comprender cómo funcionan los seres vivos, pero hemos tenido que resolver problemas importantes de computación para sacarlo adelante.

¿Y el segundo perfil?

Es el de personas que tienen facilidad por la informática y las matemáticas, pero que las encuentran áridas y preferirían aplicarlas a problemas reales como los de la biología.

¿Recomendaría a estudiantes de bachillerato que se planteen dedicarse a la bioinformática?

Desde luego. El problema es que en España lo tienen difícil para formarse. Pero si les gustan la biología y la computación, es una profesión con futuro.●