

Tendències

Cap a una economia del coneixement

JOSEP CORBELLA
Barcelona

Barcelona, tenim un problema: el volum de dades biològiques que s'està obtenint en centres d'investigació i hospitals està creixent més ràpid que la capacitat d'emmagatzemar-los i digerir-los. Arribarà un punt, a mitjà termini, en què la capacitat dels ordinadors no donarà l'abast per gestionar l'allau de dades que generen metges i biòlegs.

Per a una ciutat que es mira al mirall de Boston per construir una economia del coneixement, i que disposa ja d'alguns dels equips de bioinformàtics més importants d'Europa, "resoldre aquest problema és una qüestió estratègica", afirma **Modesto Orozco**, bioquímic de la Universitat de Barcelona (UB) i investigador de l'Institut de Recerca Biomèdica (IRB) i del Barcelona Supercomputing Center (BSC).

El problema, amb el qual no només es troba Barcelona sinó totes les capitals d'investigació biomèdica, s'explica pel dràstic abaratiment de seqüenciar genomes que s'ha registrat en els últims cinc anys. Si el 2007 el cost de seqüenciar cada genoma

Barcelona, capital GENÒMICA

La caiguda del cost d'investigar l'ADN provoca una allau de dades biològiques

humà era de deu milions de dòlars, l'arribada d'una nova tecnologia de seqüenciació ha fet que el preu hagi caigut ara per sota dels 10.000 dòlars. Aquest abaratiment ha provocat que un nombre creixent d'investigadors i de metges hagin començat a seqüenciar genomes així que s'ho han pogut permetre.

Avui dia se seqüencien genomes de cèl·lules tumorals, de varietats de meló, de fòssils de neandertals... En el cas del càncer, es volen comparar els genomes de cèl·lules tumorals amb els de

cèl·lules sanes del mateix pacient per veure quins són els gens que estan alterats al tumor. En un futur pròxim, es preveu poder comparar el genoma de cèl·lules tumorals abans i després del tractament per avaluar l'eficàcia dels diferents fàrmacs... I no només s'emmagatzemen dades de genomes. També s'han començat a estudiar epigenomes (que indiquen quins gens estan actius i quins estan silenciats en un determinat moment en una cèl·lula concreta); proteomes (que indiquen quines

proteïnes hi ha en una cèl·lula); transcriptomes (que indiquen quin ARN hi ha)... En fi, una quantitat enorme de dades que van camí de desbordar la capacitat dels superordinadors.

Segons la famosa llei de Moore, la capacitat d'arxivament de dades dels xips es duplica cada 18 mesos. Però segons una estimació de l'Institut Europeu de Bioinformàtica, amb seu al Regne Unit, la quantitat de dades biològiques que s'emmagatzemen es duplica cada nou mesos.

"Si el volum de dades augmen-

ta més ràpid que la capacitat d'emmagatzemar-los, ens trobem amb un coll d'ampolla que limitarà la investigació i dificultarà que aquestes dades s'apliquin en benefici dels pacients", adverteix **Ivo Gut**, director del Centre Nacional d'Anàlisi Genòmica (CNAG) a Barcelona.

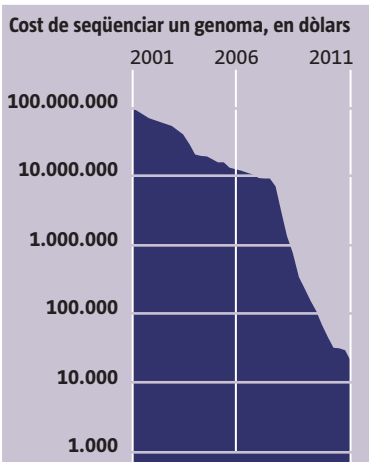
La solució, segons Modesto Orozco, suposa gestionar les dades de la manera més eficient possible. Si fins ara no s'ha fet, és perquè la seqüenciació de genomes és una tecnologia jove amb què cada centre d'investigació ha començat a experimentar pel seu compte. Però amb l'actual crisi de creixement d'aquesta tecnologia, "és millor compartir recursos que duplicar-los", afirma Orozco.

Compartir recursos significaria establir un banc central de dades biològiques que guardés la informació dels diferents hospitals i centres d'investigació. "No té sentit -segons Modesto Orozco- que cada institució gestioni les seves dades de manera aïllada; seria molt més eficient que estiguéssim tots connectats".

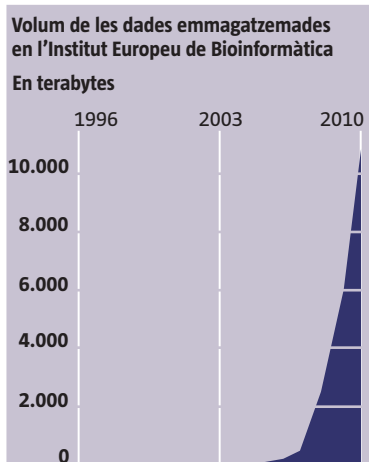
El Biodata Center -com l'anomena Orozco de manera provisional- estaria vinculat al superordinador Mare Nostrum al BSC, al Campus Nord de la Uni-

Els ordinadors no donen l'abast per gestionar l'allau de dades que aporten els biòlegs

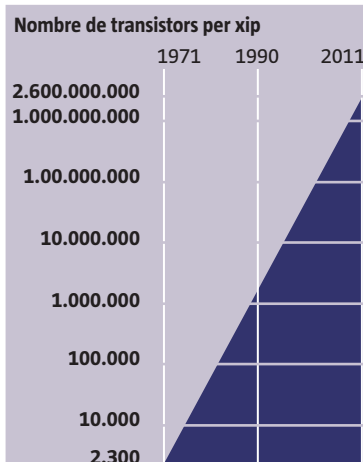
Com el preu de seqüenciar genomes cau en picat des del 2007...



... el volum de dades biològiques es multiplica per dos cada nou mesos



Tanmateix, la capacitat d'emmagatzemar dades en un xip es duplica cada 18 mesos



Per tant, l'augment de la capacitat d'emmagatzemament no serà suficient per gestionar tot el volum de dades biològiques

FONT: Elixir / BSC

LA VANGUARDIA

Tretze països creen un consorci bioinformàtic europeu

■ Tretze països europeus -entre els quals Espanya- han creat el consorci Elixir per gestionar l'allau de dades biològiques derivades de la revolució genòmica. El pla de treball preveu que Elixir es construirà com una infraestructura amb un *hub* principal situat al Regne Unit -l'Estat que més inverteix en el projecte i el que lidera la investigació genòmica a Europa- i diversos

nodes distribuïts per tot Europa. A Espanya, estan involucrats en el projecte Elixir els Centres Nacionals d'Investigacions Oncològiques (CNIO) i Barcelona Supercomputing Center.

"El repte que planteja l'allau de dades biològiques és massa gran perquè cap país no el pugui afrontar per si sol", va declarar el 3 de setembre Soren Brunak, president de

la junta directiva d'Elixir.

La fase preparatòria del projecte es va iniciar el 2007 amb finançament de la Comissió Europea i conclourà el desembre d'aquest any. El calendari de treball preveu que la fase de construcció de la infraestructura duri quatre anys i acabi el desembre del 2016. A partir de l'any 2017 s'iniciarà la fase permanent d'explotació de la xarxa Elixir.



Modesto Orozco, al Barcelona Supercomputing Center

ELS ESTATS QUE MÉS APORTEN AL PROJECTE ELIXIR (en euros)

REGNE UNIT
100 milions

FINLÀNDIA
6,85 milions

NORUEGA
6,5 milions

ESPANYA
5 milions

DINAMARCA
5 milions

SUÈCIA
1,7 milions

versitat Politècnica. Però perquè el projecte sigui viable, serà necessari que la connexió entre el BSC i els centres d'investigació i hospitals es faci per una xarxa de fibra òptica amb un gran ample de banda que actualment no existeix.

“Seria interessant tenir un reposador centralitzat de dades, però per fer bona investigació és imprescindible poder tenir un

PROBLEMA EMERGENT

El volum de dades biològiques augmenta més ràpid que la capacitat informàtica

POSSIBLES CONSEQÜÈNCIES

El problema pot dificultar que les dades genòmiques beneficiïn els pacients

SOLUCIÓ PROPOSADA

El BSC proposa compartir recursos de computació entre centres científics

accés molt ràpid a les dades”, explica Joan Seoane, investigador de l'Institut d'Oncologia de Vall d'Hebron (VHIO).

“Té sentit compartir recursos”, coincideix Ivo Gut. “Però fins i tot així, i amb una transmissió de dades ràpida, continuarem necessitant els nostres ordinadors per tenir una certa autonomia respecte al BSC i poder prioritzar les investigacions que considerem prioritàries”.

Actualment el BSC ja està connectat per fibra òptica al CNAG, la qual cosa ha facilitat la seqüenciació del genoma de la leucèmia entre les dues institucions. Però encara no ho està a altres centres d'investigació biomèdica ni als grans hospitals. Establir aquesta xarxa requeriria una inversió de diversos milions d'euros.

Arribats a aquest punt, la pregunta clau és si resultarà més costós construir la xarxa de fibra òptica o no construir-la. La resposta depèn, segons Modesto Orozco, de la importància que tingui la investigació biomèdica per al país. “A curt termini –sosté–, podem continuar com fins ara. Però en tres o quatre anys, si no fem alguna cosa per evitar-ho, arribarem a una situació de col·lapse de les dades biològiques”.

“Necessitem bioinformàtics i no els estem formant”

Roderic Guigó, coinvestigador principal del projecte Encode

J. CORBELLA Barcelona

Roderic Guigó va tenir la gran sort que no li van sortir bé dos dels primers experiments que va fer quan començava la seva carrera investigadora. Va decidir que el laboratori no era el seu fort, que li agradava més la feina teòrica, i va apostar per la bioinformàtica quan encara ningú no parlava de bioinformàtica. Va saber intuir en quina direcció canviaria la biologia i l'aposta li va sortir bé. Va liderar l'únic equip d'investigació espanyol que va participar en la històrica seqüenciació del genoma humà. Després va tenir una participació destacada en la seqüenciació d'altres genomes. I, quan va sortir el projecte Encode com a continuació del projecte Genoma humà el 2003, va ser seleccionat –i finançat pel Govern dels Estats Units– com un dels investigadors principals del projecte. Investigador del Centre de Regulació Genòmica (CRG) i professor de la Universitat Pompeu Fabra (UPF), està considerat com un dels millors bioinformàtics d'Europa.

Com es pot fer algú bioinformàtic?

A Espanya és difícil.

En el meu cas, la trajectòria d'investigació del meu equip s'ha basat sempre en la genòmica computacional. Aquí desenvolupem un software que es va fer servir per identificar gens en la seqüenciació del genoma humà. Després vam tenir una participació important en el genoma del ratolí. I quan es va iniciar el projecte Encode, que era un projecte competitiu obert a investigadors de tot el món, era una continuació lògica de la feina que havíem fet.

Per què diu que a Espanya és difícil?

Perquè aquí, a diferència del que passa en altres països, no s'ofereix la possibilitat d'estudiar alhora biologia i informàtica, la qual cosa és un error estratègic. Espanya necessita bioinformàtics i no els està formant.

Perquè els necessita?

Al món de la biologia, fins fa poc la major part de la feina

consistia a obtenir dades. Ara es generen de manera automàtica quantitats ingents de dades i el gran problema consisteix a interpretar-les. I el mateix passarà aviat al món de la medicina. La importància de la computació en biologia i en medicina és cada vegada més gran. Per això hi ha una alta demanda de personal qualificat en bioinformàtica i biologia computacional.

I si no es forma aquest personal?

Tindrem les eines per obtenir les dades, com seqüenciar el genoma de cadascun de nosaltres, que cada dia és més barat, però, tanmateix, no podem fer servir aquestes dades perquè no sabem què fer-ne. És a dir, tindrem les

opció que avui dia no existeix a Espanya. Hi ha el precedent de la doble titulació en física i matemàtiques, en la qual es matriculen estudiants altament motivats. Caldria oferir també una doble titulació de biologia i informàtica.

No seria millor crear una carrera de bioinformàtica?

És una altra possibilitat, però segons la meua opinió una superespecialització dels estudiants des dels primers cursos de carrera, com s'ofereix ara, és la direcció contrària de cap a on hauríem d'anar. En un món ràpidament canviant, hauríem d'oferir carreres flexibles i formar els estudiants perquè es puguin adaptar als canvis.

Quines aptituds cal tenir per ser bon bioinformàtic?



JORDI PLAY

Roderic Guigó, al Centre de Regulació Genòmica

PROFESSIÓ AMB FUTUR

“Hi ha una alta demanda de personal qualificat en bioinformàtica”

FALLA LA UNIVERSITAT

“Hi hauria d'haver l'opció d'estudiar alhora informàtica i biologia”

dades, però no les sabem interpretar.

Solució?

El que caldria seria dinamitar el sistema universitari que tenim, això seria l'ideal. Fer-lo menys rígid, i molt més lliure i flexible.

I si l'ideal no és possible?

Almenys hi hauria d'haver la possibilitat d'estudiar alhora informàtica i biologia, una

Hi ha dos perfils principals. Un és el de persones fascinades per la biologia i que, al mateix temps, tenen facilitat per les matemàtiques i la informàtica. Encode, per exemple, és sobretot un projecte per comprendre com funcionen els éssers vius, però hem hagut de resoldre problemes importants de computació per tirar-lo endavant.

I el segon perfil?

És el de persones que tenen facilitat per la informàtica i les matemàtiques, però que les troben àrides i preferirien aplicar-les a problemes reals com els de la biologia.

Recomanaria a estudiants de batxillerat que es plantegin dedicar-se a la bioinformàtica?

Sens dubte. El problema és que a Espanya ho tenen difícil per formar-se. Però, si els agraden la biologia i la computació, és una professió amb futur.



LAURA GUERRERO