



DOSSIER DE PREMSA DE PRESENTACIÓ DEL "gff2ps",

EL SOFTWARE QUE PERMET VISUALITZAR EL GENOMA HUMÀ

L'edició electrònica de la revista científica nord-americana Science publica avui un dossier especial, "The Sequence of the Human Genome", en el qual s'explica tot el procés que l'empresa Celera Genomics ha seguit per aconseguir la seqüència completa i contínua del genoma humà.

En aquest número especial de Science també es publica un article signat, entre altres autors, pels biòlegs Roderic Guigó i Josep Francesc Abril, investigadors de l'Institut Municipal d'Investigació Mèdica (IMIM) i de la Universitat Pompeu Fabra (UPF). En aquest article de la prestigiosa revista científica es parla del programa informàtic desenvolupat per aquests bioinformàtics que, per primera vegada, ha permès la visualització de la seqüència completa del genoma humà.

La seqüenciació del genoma humà: un pas més per conèixer els processos vitals més bàsics de l'ésser humà

La tecnologia que hi havia fins fa pocs anys només permetia seqüenciar fragments molt petits de l'ADN, de tan sols 700 o 800 lletres dels 3.000 milions de nucleòtids -o lletres- que constitueixen el nostre genoma. A finals del juny del 2000 es va anunciar la seqüència que cobreix, essencialment, el 95 % del genoma humà, en forma d'un conjunt de milions de fragments. Des d'aleshores s'ha intentat reconstruir la seqüència lineal dels 23 parells de cromosomes que conformen el genoma, a partir d'aquests fragments. El que els investigadors genètics disposen des d'ara és l'acoblament del genoma: una ordenació lineal d'aquests fragments sobre els cromosomes bastant bona, que cobreix gairebé la totalitat del genoma humà i amb força fiabilitat. Possiblement, la seqüència completa del genoma humà no es coneixerà fins l'any 2002, ja que encara queden petits forats o fragments que són difícils d'obtenir i de seqüenciar, és a dir, de posar en l'ordre correcte.

Això és el que coneixem, però el que ens interessa conèixer és, sobretot, els gens que estan codificats en aquest genoma, i com aquest gens i aquestes proteïnes interactuen en la cèl·lula, ja que són les responsables de les formes de la vida, el que obre noves portes a la recerca per a les properes dècades. És en el genoma on es produeix la interfase entre la vida i la no vida.

La bioinformàtica: un eina imprescindible per seqüenciar i visualitzar el genoma

La recerca de Roderic Guigó i Josep F. Abril, que es desenvolupa en el laboratori d'informàtica del genoma, dins del Grup de Recerca en Informàtica Biomèdica de l'IMIM-UPF, es centra en l'anàlisi de les seqüències genòmiques per intentar extreure informació rellevant de les pròpies seqüències. Els investigadors van desenvolupar un programa informàtic que permetia visualitzar aquesta informació de forma que fos útil. Com que la seva recerca és pública, van posar el seu programa a Internet, sota la llicència de software públic GNU General Public Licence. Diversos grups es van interessar per fer-lo servir, entre ells l'empresa Celera Genomics. La col·laboració de Roderic Guigó i Josep F. Abril amb Celera va començar el febrer del 2000 quan l'empresa va contactar amb ells per fer el

dibuix del cromosoma del genoma de la mosca del vinagre, la *Drosophila melanogaster*, i, des d'aleshores, l'empresa va quedar oberta a noves col·laboracions, ja que Celera estava treballant per obtenir, el més aviat possible, la seqüència del genoma humà, fet que es va produir el juny del 2000.

El programa "gff2ps", l'eina que permet visualitzar, per primera vegada, el genoma humà

Roderic Guigó i Josep Francesc Abril són els únics investigadors espanyols que han col·laborat amb Celera en l'anotació del genoma humà. La seva aportació ha estat decisiva, ja que Celera ha adoptat el software que els investigadors de l'IMIM-UPF han desenvolupat per representar dades genòmiques i així permetre la visualització completa del genoma humà. Aquest programa, anomenat "gff2ps", ha fet possible obtenir una figura en el qual s'han representat els aproximadament 3.000 milions de nucleòtids -o lletres- que componen el genoma humà. Aquest gràfic del genoma humà pot ser de gran utilitat per als investigadors genètics, ja que poden interpretar-lo a primera vista, com si es tractés d'un mapa de carreteres. Amb la representació de la seqüència completa del genoma, els genetistes poden localitzar de manera més senzilla les regions de l'ADN que contenen els gens i veure la seva distribució al llarg dels cromosomes.

L'avantatge que té aquest programa respecte d'altres és, sobretot, que permet tractar grans volums d'informació, i que permet imprimir-los en qualsevol format. Quan parlem de genoma, parlem d'una quantitat d'informació molt gran i, per tant, per tenir una visualització que sigui útil cal ser capaços d'imprimir (si es vol) en formats molt grans. Per exemple, l'edició en paper de Science, que es publicarà el proper dia 16 de febrer, inclourà, com a suplement, un pòster plegat, amb unes mides aproximades d'1,40 metres d'alçada i 1 metre d'amplada, en el qual es podrà gaudir del mapa del genoma humà que els dos biòlegs catalans han aconseguit representar.

L'aportació del programa "gff2ps" a la investigació genòmica

Bàsicament, la visualització de les dades de les seqüències genòmiques és una de les passes finals en l'anàlisi d'aquestes seqüències. Si s'agafa un dels 23 parells de cromosomes que hi ha en el genoma, s'obtenen uns fragments molt petits, els quals passen a les màquines de seqüenciació, d'on s'obtenen les seqüències de nucleòtids. Un cop acoblats aquests fragments, es disposa d'una seqüència completa del cromosoma.

Una seqüència del genoma acoblada és un llistat enorme de milers de lletres -A, C, G i T- que constitueixen la seqüència bàsica d'una regió genòmica. D'aquesta seqüència ens interessen regions concretes, en les que es troben els gens.

El problema és que en tot el procés es genera una quantitat increïble de dades i és quan es necessita un programa de visualització de les seqüències. El programa "gff2ps" permet visualitzar tot el genoma humà d'una vegada i possibilita veure en una figura única els gens, les seves extensions i les seves estructures. La visió del genoma en aquesta escala ens permet trobar regions en les quals els gens s'acumulen com si fossin una illa, i també podem veure quin tipus de gens s'hi acumulen. El codi de colors dels gens en aquest nivell ens permet definir com s'han trobat els mateixos, a partir d'evidències experimentals o si s'han obtingut per semblança amb altres seqüències conegudes o a partir de la predicció computacional de nous gens.

Perfils biogràfics de Roderic Guigó i Josep F. Abril

Roderic Guigó i Serra (Barcelona 1959) és doctor en Biologia per la Universitat de Barcelona. Actualment, és investigador de l'Institut Municipal d'Investigació Mèdica (IMIM), on dirigeix el laboratori d'informàtica del genoma, dins del Grup de Recerca en Informàtica Biomèdica, i és professor associat del Departament de Ciències Experimentals i de la Salut de la Universitat Pompeu Fabra.

Com a investigador, va treballar, en una primera etapa, a la Universitat de Harvard, a la Unitat de Recerca de Biologia Molecular Computacional del Dana Farber Cancer Institute. Més tard es va traslladar al Centre de Recerca d'Enginyeria Biomolecular de la Universitat de Boston, i l'any 1992 la seva recerca sobre anàlisi del genoma es va desenvolupar al Laboratori Nacional de Los Álamos. L'any 1994 retorna a Barcelona, a l'IMIM, on centra la seves investigacions en la seqüència genòmica i en la seva aplicació informàtica. És autor de nombrosos articles en revistes científiques, com ara Science, Genome Research i Bionformatics. Ha col·laborat també en la publicació de diversos llibres i en la realització de seminaris.

Jospe Francesc Abril Ferrando (Tortosa, 1970) és llicenciat en Biologia per la Universitat de Barcelona. Actualment, és alumne del Programa de Doctorat en Ciències de la Salut i de la Vida de la UPF i centra la seva recerca en l'evolució de l'estructura dels gens. La seva afició pels ordinadors el va portar cap al camp de la bioinformàtica. Des de fa uns anys, és becari de l'equip que dirigeix Roderic Guigó. Participant de diversos seminaris i conferències, entre les que es troba la ISMB'99 (Conferència Internacional de sistemes intel·ligents de biologia molecular), també ha publicat articles en revistes científiques.

La resta de membres del laboratori d'informàtica del genoma són: Enrique Blanco, Moisés Buset, Sergi Castellano, Xavi Fustero i Genís Parra.

Barcelona, 12 de febrer del 2001

Nota: més informació (àudio, vídeo i fotografies) a l'adreça d'Internet
<http://www.upf.es/grec/noticies/premsa/genoma.htm>

Marc Permanyer. Àrea de Comunicació de la UPF
Tels. 93 542 21 00 - 629 76 06 28 ? Fax. 93 542 20 81
A/e: marc.permanyer@grec.upf.es
Les notícies de la UPF a <http://www.upf.es/grec/premsa>

Institut Municipal d'Investigació Mèdica (IMIM)
Tel. 93 221 10 09 ? Fax. 93 221 32 27
A/e: elopez@imim.es i arevuelta@imim.es
La web de l'IMIM: <http://www.imim.es>