

La nova generació de tecnologies de seqüenciació obre una nova era en la genòmica

Mònica Bayés

Centre Nacional d'Anàlisi Genòmica (CNAG-CRG)

Correspondència: Mònica Bayés. Centre Nacional d'Anàlisi Genòmica. Carrer de Baldiri Reixac, 4. 08028 Barcelona. Adreça electrònica: mbayes@pcb.ub.cat.

DOI: 10.2436/20.1501.02.151

ISSN (ed. impresa): 0212-3037

ISSN (ed. digital): 2013-9802

<http://revistes.iec.cat/index.php/TSCB>

Rebut: 20/10/2014

Acceptat: 04/12/2014

Resum

Les tecnologies i aplicacions de la genòmica han experimentat ràpids avanços en els darrers anys. La nova generació de tecnologies de seqüenciació (*next generation sequencing*, NGS) permeten recopilar informació a escala genòmica d'una mostra per identificar la seqüència dels fragments de DNA, la variació dels nivells d'expressió dels gens o les modificacions de les bases, en un temps relativament curt i a un cost relativament baix. Generen una quantitat immensa de dades que s'interpreten en el marc d'una disciplina, la bioinformàtica, i mitjançant l'ús de potents recursos informàtics. Les tecnologies NGS possibiliten avanços en camps molt diversos com ara la millora genètica dels conreus i del bestiar, la recerca relacionada amb la base molecular de les malalties humanes, el diagnòstic molecular, el desenvolupament de nous agents terapèutics i el descobriment de nous organismes infecciosos.

Paraules clau: genòmica, seqüenciació massiva, genoma, exoma, medicina personalitzada.

Introducció a la genòmica

La genòmica és la disciplina que té per objectiu caracteritzar el genoma sencer dels organismes, utilitzant tècniques de biologia molecular a gran escala i l'anàlisi bioinformàtica de les dades (<http://www.genome.gov/18016863>).

El genoma és el conjunt del material genètic de les cèl·lules d'un organisme que s'emmagatzema en forma de DNA (àcid desoxiribonucleic) i que conté tota la informació per al desenvolupament i funcionament correctes de l'organisme. La molècula de DNA està formada per dues cadenes d'unes unitats químiques que anomenem *nucleòtids* o bases, enrotllades al voltant d'un eix comú formant una doble hèlix. Hi ha quatre tipus de bases, que identifiquem amb les lletres A, T, G i C (adenina, timina, guanina i citosina). Les dues cadenes queden unides per ponts d'hidrogen formats entre la base d'una cadena i la de l'altra cadena amb la qual queda enfrontada. Els aparellaments sempre són entre A-T i G-C.

Anomenem *gens* els segments de DNA que confereixen instruccions específiques a la cèl·lula, sovint mitjançant la síntesi de proteïnes a partir de molècules intermèdies anomenades *RNA missatgers*. Les proteïnes són les que formen el òrgans i teixits del cos, i controlen les reaccions químiques i la comunicació entre les cèl·lules. Una mutació en el DNA d'una cèl·lula pot produir una proteïna aberrant que pot alterar els processos normals de l'organisme i donar lloc a una malaltia com ara el càncer.

El genoma humà té unes tres mil milions de bases i entorn de vint mil gens, cada un dels quals dóna lloc a tres proteïnes diferents de

Next-generation sequencing technologies open a new era in genomics

Summary

DNA sequencing technologies and applications have undergone a remarkable evolution during the last few years. Next-generation sequencing (NGS) technologies enable the collection of genome-wide information from a sample and the identification of the DNA nucleotide sequence, the levels of gene expression and/or the nucleotide modification profiles, very rapidly and at relatively low costs. They produce an overwhelming amount of data that is interpreted within a bioinformatics framework that requires significant computational infrastructure. NGS advances open up new opportunities in genomics-assisted breeding of crops and livestock, in research about the molecular base of human diseases and clinical molecular diagnostics, and in the development of new therapies and the discovery of new infectious agents.

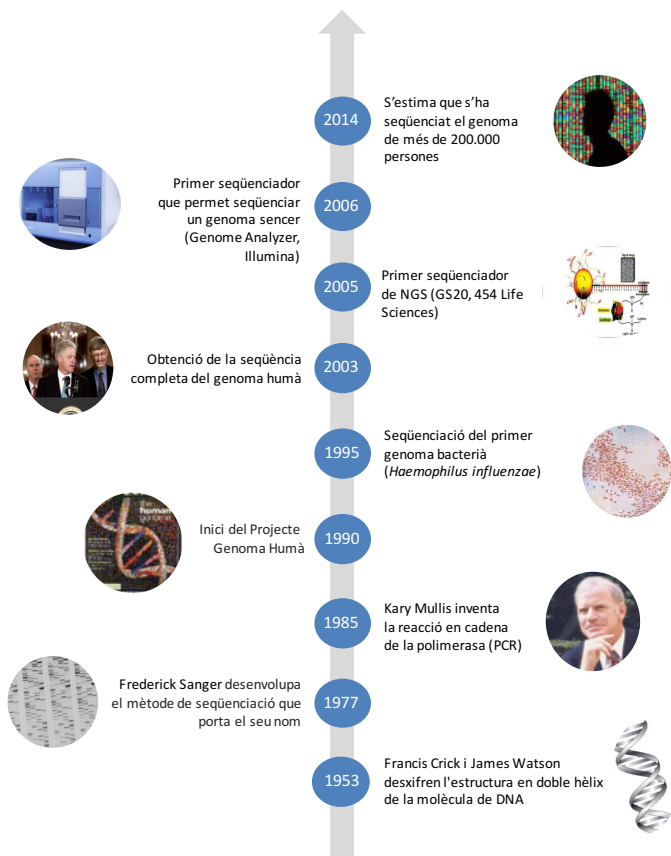
Keywords: genomics, massive sequencing, genome, exome, personalized medicine.

mitjana. La primera seqüència del genoma humà, és a dir, l'ordre de la gran majoria de les bases en la cadena del DNA, es va completar l'any 2003 en el marc d'una col·laboració internacional, el Projecte Genoma Humà (<http://www.genome.gov/10001772>). És tracta d'un dels assoliments més importants de la biologia; el projecte es va dur a terme en tretze anys i es calcula que va tenir un cost d'uns tres mil milions de dòlars (vegeu la figura 1). Totes les dades del projecte estan dipositades en bases de dades públiques i de lliure accés per tal de facilitar la recerca i promoure l'obtenció de beneficis i recursos per a la societat.

Els resultats del projecte van posar de manifest que dues persones qualssevol comparteixen el 99,9 % del seu genoma. Les posicions en què dos individus difereixen en una única base s'anomenen *SNP* (*single nucleotide polymorphisms*) (Brookes, 1999). En el genoma també podem trobar reordenaments genòmics o variants estructurals que afecten centenars o milers de bases, com ara inversions, delecions o duplicacions.

Les tècniques d'enginyeria genètica

El Projecte Genoma Humà va ser possible gràcies a les tecnologies d'enginyeria genètica desenvolupades a partir dels anys setanta (vegeu la figura 1). Aquestes tecnologies utilitzen enzims de restricció, polimerases i ligases per tallar, amplificar i unir fragments de DNA i generar molècules de DNA recombinant, formades per fragments que originalment no estaven junts. D'aquesta manera es van aïllar els primers gens i es van amplificar a l'interior de bacteris o virus en un procés que anomenem *clonatge*. Es marca un fragment de DNA homòleg



↑ Figura 1. Principals esdeveniments de la història de la genòmica.

al gen d'interès amb una molècula radioactiva o fluorescent, i s'utilitza com a sonda per identificar el clon que conté el gen, mitjançant un procés d'hibridació entre la sonda i el DNA del clon.

Des de l'any 1977 s'utilitza el mètode de Sanger per determinar l'ordre de les bases de fragments de DNA fins a una quilobase (1.000 pb) en un procés que anomenem *seqüenciació* (Sanger *et al.*, 1977). Es basa en l'ús d'un enzim, la DNA-polimerasa, per generar noves cadenes a partir de la cadena que es vol seqüenciar. En aquesta síntesi es generen fragments que acaben en les quatre possibles bases del DNA marcades cadascuna amb una molècula fluorescent diferent. Aquests fragments se separen després segons la mida en una matriu porosa i es detecten mitjançant el senyal fluorescent que emeten. Amb aquest mètode es va obtenir la primera seqüència del genoma humà i també de molts altres genomes. Actualment hi ha instruments amb noranta-sis capil·lars que permeten generar al voltant de dues megabases (2.000.000 pb) de seqüència en vint-i-quatre hores.

La tècnica de la reacció en cadena de la polimerasa (PCR) va ser inventada l'any 1985 per Kary B. Mullis (Saiki *et al.*, 1985). Permet obtenir moltes còpies de fragments específics de DNA a partir d'una quantitat original mínima. Aquesta tècnica utilitza DNA-polimerases per replicar fragments de DNA utilitzant cicles que alternen temperatures altes i baixes per separar i ajuntar les cadenes de DNA després de cada fase de replicació. La PCR va revolucionar molts aspectes de la biologia molecular i el diagnòstic clínic. S'utilitza també en investigacions de criminologia per identificar persones a partir de mostres de cabell.

Aquestes tècniques han proporcionat eines genètiques d'ús habitu-

al en els laboratoris de recerca, hospitals i empreses de biotecnologia. El seu impacte ha estat tan important que primer Sanger l'any 1980 i després Mullis l'any 1993 van ser guardonats amb el Premi Nobel de Química.

Els microxips de DNA

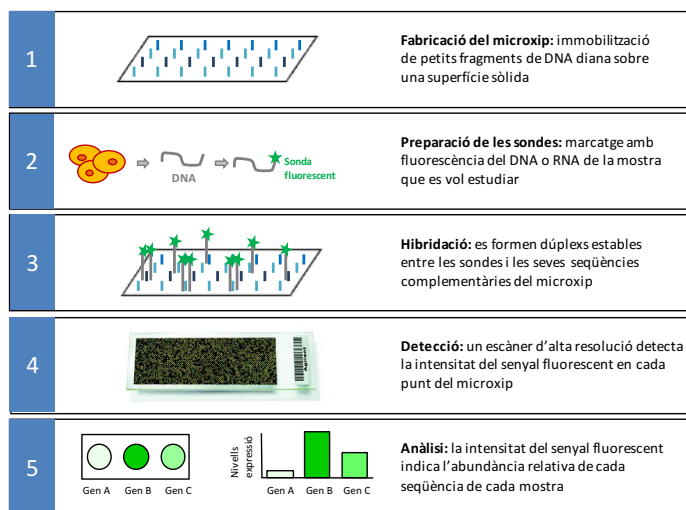
Un microxip de DNA és una col·lecció de milers de fragments petits de DNA coneguts i ancorats sobre una superfície sòlida, sovint una placa de vidre de la mida d'un portaobjectes per a observacions al microscopi (per a una revisió, vegeu Bumgarner, 2013). Hi ha diversos mètodes que permeten obtenir aquests fragments a gran escala i immobilitzar-los de manera ordenada en uns pocs centímetres o mil·límetres quadrats. En els experiments de microxips s'hibriden una o més mostres complexes de DNA o RNA marcades amb molècules fluorescentes, que anomenem *sondes*, amb els fragments de DNA immobilitzats, anomenats *dianes* (vegeu la figura 2). Quan les sondes troben les seves seqüències complementàries entre les dianes immobilitzades es formen dúplexs més o menys estables que es poden detectar amb un escàner d'alta resolució que registra la llum emesa en cada punt del microxip. La intensitat del senyal fluorescent que es detecta en un punt és proporcional al nombre de dúplexs que s'han format amb aquells fragments diana i és, per tant, un indicador de l'abundància relativa de cada seqüència de la mostra. Tot i la mida reduïda dels microxips, en cada experiment es genera una gran quantitat de dades que es processen amb programes informàtics especialitzats que permeten extreure les dades primàries d'intensitat de les imatges escanejades, normalitzar aquestes intensitats, eliminar el soroll de fons i finalment manipular les dades per treure'n les conclusions biològiques escaients.

Els microxips van sorgir als anys noranta com a mètode per determinar a gran escala l'expressió relativa dels gens en un teixit o en diverses situacions experimentals, com ara la determinació dels efectes d'un fàrmac sobre el conjunt de RNA de les cèl·lules (el transcriptoma) (Schena *et al.*, 1995). També es poden utilitzar per analitzar genomes. Hi ha plataformes de microxips comercials per analitzar milions de SNP coneguts al llarg del genoma (per a una revisió, vegeu Ragoussis, 2009). S'anomenen xips de genotipatge i contenen centenars de milers de petits fragments de DNA entre vint i seixanta nucleòtids que permeten determinar si el DNA de la mostra conté un nucleòtid determinat en una posició concreta. En un únic experiment es poden determinar fins a cinc milions de posicions del genoma humà. Finalment, també es poden utilitzar els microxips per identificar aquells elements reguladors del DNA als quals s'uneix una determinada proteïna o un factor de transcripció, o per detectar l'estructura de les diverses formes dels RNA missatgers.

Gràcies als microxips es poden interrogar milers o fins i tot milions de seqüències en un únic experiment, sense cap hipòtesi prèvia, de manera ràpida i relativament barata. A diferència de la seqüenciació, però, en aquests experiments només podem detectar aquelles seqüències que coneixem i que estan presents en el microxip.

La nova generació de tecnologies de seqüenciació

L'any 2004 hi va haver un canvi de paradigma en el camp de la seqüenciació del DNA i de la genòmica, amb l'aparició d'una nova generació de tecnologies de seqüenciació (NGS) o seqüenciació massiva (per a revisions, vegeu Buermans i Dunnen, 2014, i Dijk *et al.*, 2014). Les tecnologies de NGS combinen l'ús de tècniques d'enginyeria genètica, la nanotecnologia i la generació de milions de dades basades en la imatge.



↑ Figura 2. Esquema d'un experiment de microxips per detectar l'expressió de tots els gens del genoma.

A diferència de la seqüenciació pel mètode de Sanger, que es basa en l'anàlisi individual de fragments de DNA, els seqüenciadors de NGS són capaços d'analitzar milions de fragments de DNA en paral·lel i, en conseqüència, de seqüenciar un genoma humà sencer en pocs dies. Generen una quantitat de dades genòmiques impensable fa deu anys, de qualsevol organisme, en coneguem o no el genoma prèviament, i de manera robusta.

En els últims anys han aparegut diverses plataformes en el mercat però el seu flux de treball és conceptualment similar (vegeu la figura 3). El primer pas consisteix a fragmentar el DNA de manera aleatòria i afegir en els extrems dels milions de fragments resultants unes seqüències comunes. Després s'aïllen les molècules individuals d'aquesta biblioteca i s'obtenen milions de còpies de cadascuna mitjançant la tècnica de la PCR. Un cop amplificades les molècules originals, se seqüencien utilitzant equips que combinen sistemes de microfluids i la detecció d'imatges d'elevada resolució a temps real. La majoria de les plataformes actuals utilitza la seqüenciació per síntesi, és a dir, la fabricació d'una cadena a partir d'una cadena motlle gràcies a una DNA-polimerasa. Les lectures són una mica més curtes (generalment entre 75 i 400 pb) que les que s'obtenen pel mètode de Sanger (800-1.000 pb), però permeten seqüenciar els dos extrems de cada fragment i això facilita l'alineament amb el genoma de referència i la detecció de reordenaments genòmics.

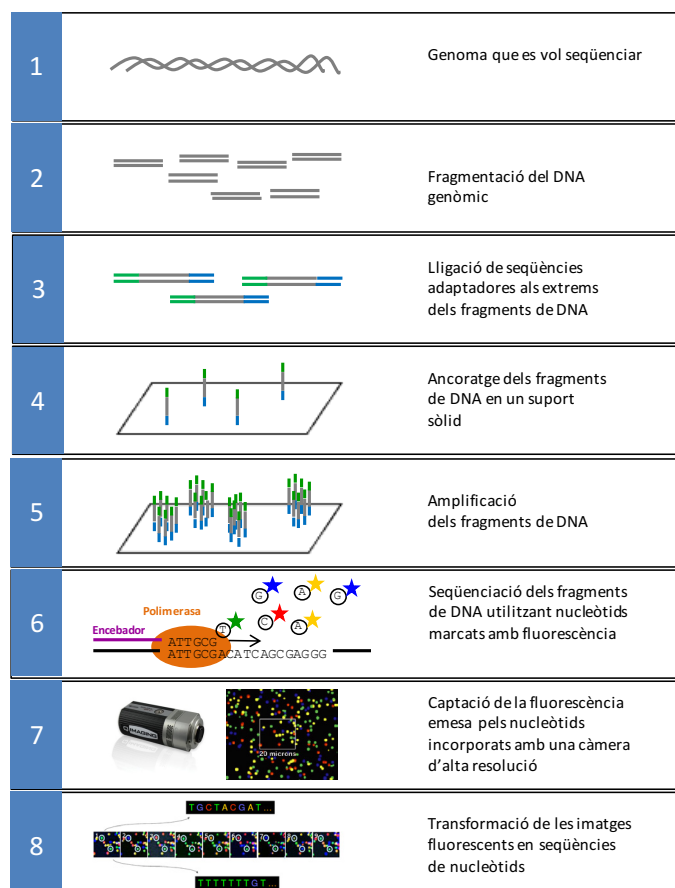
Diverses cases comercials han desenvolupat equips de NGS, com ara el HiSeq4000 d'Illumina, el Genome Sequencer FLX de Roche o el Proton de Life Technologies. Difereixen en els mètodes per preparar, aïllar i amplificar les biblioteques de DNA, en l'estratègia per detectar el senyal resultant de la seqüència, que pot ser òptic, químic o elèctric, i en la quantitat de gigabases (Gb, 1 Gb = 1.000.000.000 pb) que es poden generar en cada carrera o experiment. Avui dia el seqüenciadors més populars són els HiSeq4000 d'Illumina, que permeten generar 1.500 Gb de seqüència (l'equivalent a seqüenciar dotze genomes humans amb una cobertura alta) en una sola carrera que dura aproximadament quatre dies. Alguns d'aquests són instruments més petits, com el MiSeq (Illumina) o l'Ion Torrent (Life Technologies), pensats per a investigadors individuals o per a l'àmbit clínic. En l'extrem oposat trobem els nous equips X Ten d'Illumina, que permeten seqüenciar 18.000 geno-

mes humans a l'any per menys de mil dòlars cadascun. Recordem que es van necessitar més de deu anys i una inversió de tres mil milions de dòlars per obtenir la primera seqüència completa del genoma humà.

De manera paral·lela als avanços de les tecnologies de NGS, les eines bioinformàtiques per a l'anàlisi de les dades resultants han experimentat una ràpida evolució. Es tracta d'algoritmes i programari que ens permeten en primer lloc avaluar la qualitat dels resultats de la seqüenciació i alinear les lectures amb una seqüència ja coneguda del genoma, de manera que cadascuna dels milions de lectures de 75-400 pb que es generen en un experiment quedi aparellada amb les posicions equivalents del genoma de referència (vegeu la figura 4). El segon pas és identificar quines d'aquestes posicions o nucleòtids difereixen del genoma de referència. Finalment, cal anotar totes aquestes variacions, és a dir, identificar en quin gen es troben, si han estat descrites amb anterioritat i predir-ne les conseqüències. Tots aquests passos són crucials per a la interpretació correcta de les dades (per a una revisió, vegeu Dolled-Filhart *et al.*, 2013).

En aquells casos en què no es disposa d'un genoma de referència amb el qual es poden comparar els resultats, s'han desenvolupat algoritmes que reconstrueixen la seqüència de tot el genoma a partir de les lectures, acoblant unes amb altres, en un procés que s'anomena acoblament *de novo*. Es pot trobar una llista actualitzada d'aplicacions bioinformàtiques per a l'anàlisi de dades genòmiques en diversos fòrums, com per exemple SEQanswers (<http://www.seqanswers.com>).

↓ Figura 3. Esquema d'un experiment de seqüenciació d'un genoma amb les noves tecnologies de seqüenciació, utilitzant la plataforma desenvolupada per la casa comercial Illumina (<http://www.illumina.com>).



Les tecnologies de seqüenciació del futur

Actualment diverses companyies desenvolupen el que s'anomena la tercera generació de tecnologies de seqüenciació (Buermans i Dunnen, 2014). Aquestes tecnologies presenten sistemes de detecció extremadament sensibles que permeten la seqüenciació de molècules individuals de DNA, sense necessitat d'amplificació mitjançant PCR. Alguns d'aquests equips, com els de Complete Genomics i Pacific Bioscience, ja estan disponibles. Una altra casa comercial, Oxford Nanopores, està seguint una estratègia radicalment diferent, basada en la detecció dels canvis en el potencial de membrana a mesura que els nucleòtids passen a través d'uns petits porus. Es tracta d'un camp en constant evolució, però sembla clar que la tecnologia del futur podrà seqüenciar molècules individuals tant de DNA com de RNA, generar lectures de diverses megabases de longitud, de manera ràpida i amb una elevada precisió, i a un cost molt assequible.

Aplicacions clíniques de la nova generació de tecnologies de seqüenciació

En molts hospitals i centres de recerca s'utilitzen les tecnologies de seqüenciació massiva per desxifrar la complexitat dels genomes normals i dels genomes associats a malaltia.

L'objectiu del Consorci Internacional del Genoma del Càncer (ICGC, <http://www.icgc.org>) és obtenir una descripció completa dels canvis del genoma i del transcriptoma en cinquanta tipus de tumors diferents (The International Cancer Genome Consortium, 2010). Per identificar tots els canvis genòmics que contribueixen al procés tumoral es comparen les seqüències completes de DNA de mostres tumorals i de teixits normals del mateix individu. D'aquesta manera s'han obtingut els perfils moleculars del genoma tumoral d'individus amb leucèmia aguda mieloide, càncer de pulmó, glioblastoma, melanoma maligne i càncer de mama, entre d'altres. Aquests estudis han revelat que els tumors hemàtics (leucèmies i limfomes), els sarcomes i els càncers pediàtrics presenten taxes de mutació més baixes. El nombre més gran de mutacions el trobem en aquells tumors sotmesos a exposicions a agents mutagènics exògens, com en el càncer de pulmó i el melanoma (per a una revisió, vegeu Chmielecki i Meyerson, 2014). En aquests casos és molt difícil distingir les mutacions que realment tenen una influència en la patologia del càncer (mutacions de tipus *driver*) d'aquelles que apareixen a l'atzar (mutacions de tipus *passenger*). S'han trobat mutacions *driver* en gens de proliferació cel·lular, com ara gens de transducció de senyal i de regulació del cicle cel·lular, i també en gens del metabolisme, la transcripció i el processament de l'RNA, entre d'altres vies. En alguns casos aquestes mutacions permeten establir el pronòstic del tumor o ens donen pistes sobre el tipus de teràpia més adequada per a aquell individu, cosa que fa possible la medicina «personalitzada». El projecte PanCancer fa un pas més enllà i coordina l'anàlisi conjunta de les dades de diversos tipus de tumors per identificar patrons genòmics comuns en el càncer (Hoadley *et al.*, 2014).

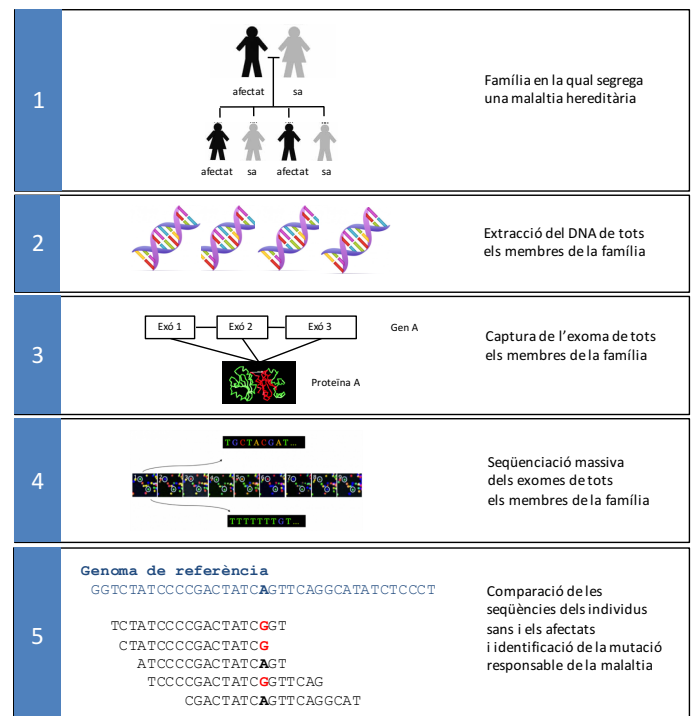
La majoria de les malalties hereditàries o familiars estan causades per mutacions molt poc freqüents situades en les regions dels gens que codifiquen proteïnes, els exons, i que representen una petita part (un 2 %) del genoma humà. S'anomena *exoma* el conjunt d'aquestes regions. S'han desenvolupat mètodes basats en la hibridació de fragments de DNA per capturar regions genòmiques específiques, com ara l'exoma, usant un conjunt de sondes de DNA que són complementàries a les regions d'interès. Aquests mètodes de captura de l'exoma, juntament amb les tecnologies de NGS, són extremadament valuosos per identi-

ficar les mutacions que causen malalties hereditàries. En aquests estudis se seqüencia l'exoma d'uns pocs individus afectats per la malaltia, s'identifiquen totes les variacions o mutacions fent comparacions amb una seqüència de referència, i se seleccionen aquelles que no s'han descrit en individus sans i que provoquen alteracions en la proteïna (vegeu la figura 4) (per a una revisió, vegeu Rabbani *et al.*, 2012). D'aquesta manera s'ha pogut identificar el gen responsable de la síndrome de Miller (Ng *et al.*, 2010) i d'unes cent cinquanta malalties hereditàries de les més de sis mil descrites.

Finalment, les tècniques de NGS també s'utilitzen per a la identificació ràpida de patògens en malalties cròniques o agudes (per a una revisió, vegeu Padmanabhan *et al.*, 2013). Permeten detectar quantitats molt baixes del patògen sense necessitat de cultivar-lo i detectar la presència, per exemple, de gens que confereixen resistència a antibiòtics. Per il·lustrar els avanços en aquest camp, en les bases de dades públiques trobem actualment la seqüència de més de deu mil mostres del virus de la grip.

Les tècniques d'enginyeria genètica i els microxips s'utilitzen des de fa temps per a la identificació d'individus amb risc elevat de desenvolupar càncer de mama, per a l'aplicació de teràpies personalitzades en càncer en funció del perfil de mutacions del tumor o per al consell genètic en algunes malalties hereditàries, entre d'altres. La seqüenciació mèdica tindrà un paper molt important en la pràctica clínica del segle XXI, també amb finalitats diagnòstiques, preventives o terapèutiques. Algunes empreses i institucions, com ara Illumina (<http://clinical.illumina.com/clinical.ilmn>) i el Baylor College of Medicine (<https://www.bcm.edu/research/medical-genetics-labs>), ofereixen serveis de seqüenciació de panells de gens, exomes o genomes amb finalitats mèdiques, en alguns casos directament als consumidors. És urgent establir un marc que reguli temes importants com ara la qualitat dels experi-

Figura 4. Ús de les noves tecnologies de seqüenciació per a la identificació de mutacions responsables de malalties hereditàries.



ments, la confidencialitat de les dades, el consentiment informat dels pacients i l'ús de la informació obtinguda no directament relacionada amb l'objecte de l'anàlisi (Abul-Husn *et al.*, 2014).

Altres aplicacions de la nova generació de tecnologies de seqüenciació

Les tecnologies de NGS possibiliten l'estudi dels organismes a una resolució sense precedents. Utilitzant aquestes plataformes podem cercar variacions en la seqüència del DNA però també determinar el nombre de còpies d'una seqüència, l'estructura, l'estat de metilació dels nucleòtids i els perfils d'expressió gènica. Processos tan diferents com la replicació, la transcripció, la traducció, la metilació i l'estructura del DNA es poden estudiar utilitzant la seqüenciació.

Actualment, en la base de dades del National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/assembly>) hi ha informació sobre la seqüència genòmica de més de dues mil espècies diferents. Hi ha iniciatives internacionals per seqüenciar mil genomes de fongs (<http://1000.fungalgenomes.org>), mil genomes d'insectes (<http://www.arthropodgenomes.org/wiki/i5K>) i deu mil genomes de vertebrats

(<https://genome10k.soe.ucsc.edu>), entre d'altres. En un futur no gaire llunyà es disposarà del genoma de com a mínim una espècie de cada grup taxonòmic (Ellegren, 2013).

Finalment, pel que fa a la indústria ramadera i de l'agricultura, la genòmica permet accelerar els processos de millora genètica tant en plantes com en animals. Es disposa ja del genoma de la majoria d'espècies vegetals d'interès agrícola i s'han utilitzat marcadors genètics per millorar la qualitat dels grans d'arròs, per aconseguir que els tomàquets madurin de manera uniforme o per controlar els processos de floració en patates (Bolger *et al.*, 2014). En el camp de la ramaderia les tècniques genòmiques són una eina fonamental per aconseguir bestiar més productiu i aliments més saludables (Ludu i Plastaw, 2013).

Agraïments

Vull expressar el meu agraïment a la doctora Berta Fuster i a l'Anna Borrell per la col·laboració en la confecció de les figures d'aquest article. El Centre Nacional d'Anàlisi Genòmica està finançat pel Ministeri d'Economia i Competitivitat i per la Generalitat de Catalunya.

Bibliografia

- ABUL-HUSN, N. S. [et al.] (2014). «Implementation and utilization of genetic testing in personalized medicine». *Pharmgenomics. Pers. Med.*, 7: 227-240.
- BOLGER, M. E. [et al.] (2014). «Plant genome sequencing - applications for crop improvement». *Curr. Opin. Biotechnol.*, 26: 31-37.
- BROOKES, A. J. (1999). «The essence of SNPs». *Gene*, 234: 177-186.
- BUERMANS, H. P.; DUNNEN, J. T. den (2014). «Next generation sequencing technology: Advances and applications». *Biochim. Biophys. Acta.*, 1842: 1932-1941.
- BUMGARNER, R. (2013). «Overview of DNA microarrays: types, applications, and their future». *Curr. Protoc. Mol. Biol.*, cap. 22, un. 22.1.
- CHMIELECKI, J.; MEYERSON, M. (2014). «DNA sequencing of cancer: what have we learned?». *Annu. Rev. Med.*, 65: 63-79.
- DIJK, E. L. van [et al.] (2014). «Ten years of next-generation sequencing technology». *Trends Genet.*, 30: 418-426.
- DOLLED-FILHART, M. P. [et al.] (2013). «Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing». *Scientific World Journal*, 2013: 730210.
- ELLEGREN, H. (2014). «Genome sequencing and population genomics in non-model organisms». *Trends Ecol. Evol.*, 29: 51-63.
- HOADLEY, K. A. [et al.] (2014). «Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin». *Cell*, 158: 929-944.
- LUDU, J. S.; PLASTOW, G. S. (2013). «Livestock and the promise of genomics». *Genome*, 56: 556-566.
- NG, S. B. [et al.] (2010). «Exome sequencing identifies the cause of a mendelian disorder». *Nat. Genet.*, 42: 30-35.
- PADMANABHAN, R. [et al.] (2013). «Genomics and metagenomics in medical microbiology». *J. Microbiol. Methods*, 95: 415-424.
- RABBANI, B. [et al.] (2012). «Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders». *J. Hum. Genet.*, 57: 621-632.
- RAGOSSIS, J. (2009). «Genotyping technologies for genetic research». *Annu. Rev. Genomics Hum. Genet.*, 10: 117-133.
- SAIKI, R. K. [et al.] (1985). «Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia». *Science*, 230: 1350-1354.
- SANGER, F. [et al.] (1977). «DNA sequencing with chain-terminating inhibitors». *Proc. Natl. Acad. Sci. USA*, 74: 5463-5467.
- SCHENA, M. [et al.] (1995). «Quantitative monitoring of gene expression patterns with a complementary DNA microarray». *Science*, 270: 467-470.
- THE INTERNATIONAL CANCER GENOME CONSORTIUM (2010). «International network of cancer genome projects». *Nature*, 464: 993-998.