

DOI: 10.2436/20.1501.02.133
ISSN (paper): 0212-3037
<http://revistes.iec.cat/index.php/TSCB>

Les malalties del segle XXI
(Jordi Barquinero i Jaume Reventós, ed.)

Treballs de la SCB. Vol. 64 (2013), p. 29-47

SEQÜENCIACIÓ DE NOVA GENERACIÓ I SALUT EN EL SEGLE XXI

FRANCISCO VIDAL

*Unitat de Diagnòstic i Teràpia Molecular, Divisió de Coagulopaties Congènites,
Banc de Sang i Teixits*

Adreça per a la correspondència: Francisco Vidal. Edifici Dr. Frederic Duran i Jordà.
Pg. de Taulat, 116. 08005 Barcelona. Tel.: 935 573 500. Adreça electrònica: fvidal@bst.cat.

RESUM

En la primera dècada del segle XXI l'aparició de les plataformes de seqüenciació massiva de nova generació (NGS) ha produït una autèntica revolució en la seqüenciació del DNA. Aquesta tecnologia permet seqüenciar en paral·lel milions de molècules de DNA a un preu molt inferior al de la seqüenciació tradicional i s'està traslladant ràpidament des del terreny de la investigació bàsica a l'assistència de capçalera al pacient. El seu extraordinari potencial en diferents especialitats mèdiques afectarà pràcticament tots els aspectes de l'atenció clínica, ja que permetrà fer rutinàriament proves diagnòstiques fins ara inconcebibles. En els propers deu anys probablement apareixeran noves plataformes de NGS per a ús clínic, encara més versàtils, precises, assequibles i fins i tot portàtils, que permetran plantejar la seqüenciació del genoma complet com a mètode universal de diagnòstic genètic. Tot això, unit a la millora del coneixement per predir com una variació genètica afecta la susceptibilitat a la malaltia, l'evolució d'aquesta o la resposta al tractament mèdic, sens dubte tindrà un efecte transformador en l'ús i protagonisme de la informació genètica en la rutina clínica del segle XXI.

Paraules clau: diagnòstic, genètic, genoma, seqüenciació de nova generació, seqüenciació massiva paral·lela.

NEXT GENERATION SEQUENCING AND HEALTH OF THE 21ST CENTURY

SUMMARY

In the first decade of the 21st century the emergence of massive Next Generation Sequencing (NGS) platforms has produced a revolution in DNA sequencing. This technol-

ogy allows parallel sequencing of millions of DNA molecules at a much lower price than traditional sequencing and it is rapidly moving from the field of basic research to routinely use in clinical practice. Its tremendous potential in several medical specialties will affect virtually all aspects of clinical care, as it will routinely perform diagnostic tests so far inconceivable. In the next ten years probably new NGS platforms for clinical use will appear more versatile, accurate, affordable and even portable that will enable sequencing the whole genome as a universal method of genetic diagnosis. This issue, together with the improved knowledge to predict how genetic variation affects the disease susceptibility, the evolution of illness or the response to medical treatment, will certainly have a transformative effect on the usage and significance of genetic information in the clinical routine of the 21st century.

Key words: diagnostics, genetic, genome, next-generation sequencing, massively parallel sequencing.

INTRODUCCIÓ

Un dels descobriments essencials de la biologia moderna és la identificació, el 1953, de l'estructura en doble hèlix del DNA (Watson i Crick, 1953). Després d'exposar les característiques de l'estructura, els autors assenyalaven que l'aparellament específic de les cadenes suggeria un possible mecanisme de còpia del DNA i que la seqüència precisa de les bases seria el codi que transportava la informació genètica. En els anys següents diversos investigadors van estudiar la relació entre la seqüència dels àcids nucleics i de les proteïnes amb l'objectiu de desxifrar el codi genètic, mentre que altres van desenvolupar tècniques per seqüenciar el DNA. D'aquesta manera, el 1977, Maxam i Gilbert (Maxam i Gilbert, 1977) i Sanger i col·laboradors (Sanger *et al.*, 1977) van desenvolupar simultàniament mètodes per a la determinació de la seqüència de nucleòtids, cosa que va suposar un progrés extraordinari en els camps de la genètica humana i del diagnòstic molecular. El mètode de Sanger, que finalment es va imposar per la seva simplicitat, longitud de lectura i fidelitat de la seqüència proporcionada, s'ha considerat durant anys com el mètode de referència per al genotipatge i per a la de-

tecció de variants i mutacions en gens d'interès clínic com els causants de trastorns monogènics, poligènics i multifactorials. Dues importants fites per al desenvolupament d'aquesta tecnologia van ser l'aparició dels fluorocroms en substitució dels marcatges radioactius i l'automatització del procés amb l'aparició dels seqüenciadors automàtics, especialment aquells que permeten fer una electroforesi capil·lar. A mitjan anys vuitanta, de l'estudi dels gens es va passar a l'estudi del genoma gràcies a un ambiciós projecte internacional, conegut com a Projecte Genoma Humà (PGH), l'objectiu del qual consistia a seqüenciar i cartografiar el genoma de la nostra espècie. El PGH es va iniciar oficialment el 1990 i van ser necessaris tretze anys i més de 2.800 científics per dur-lo a terme. L'any 1998 una iniciativa privada va entrar en la carrera per a l'obtenció de la seqüència completa del genoma humà, va encoratjar la competència i va accelerar la consecució de l'objectiu final. Així, al febrer de 2001, les revistes *Nature* (McPherson *et al.*, 2001) i *Science* (Venter *et al.*, 2001) publicaven simultàniament dues edicions especials que van proporcionar el primer esborrany de la seqüència del genoma humà. S'ha complert, per tant, la primera dècada d'aquesta transcendental fita que

ha ocupat un paper decisiu en el desenvolupament de la biomedicina i ha donat lloc a productives sinergies entre els investigadors, els clínics i la indústria farmacèutica i biotecnològica. No obstant això, amb les eines disponibles en aquell moment resultava difícil progressar àgilment envers el coneixement detallat de la variabilitat genètica i les implicacions d'aquesta en la salut humana. Tot això està canviant i encara canviarà més durant les pròximes dècades: el desenvolupament de noves tècniques de seqüenciació del DNA aparegudes durant els últims anys propiciaran que genètica i genòmica caminin juntes i la integració de les dades d'ambdues disciplines ens durà a una interpretació millor de l'efecte de la variabilitat genètica sobre els diferents aspectes de la salut humana (Gibbs, 2011).

LA SEQÜENCIACIÓ DE NOVA GENERACIÓ

A partir de l'any 2005 es produeix una autèntica revolució en les tècniques de seqüenciació del DNA amb l'aparició de les plataformes de seqüenciació massiva de nova generació o NGS (del terme anglès *next generation sequencing*). Sense pretendre fer una relació exhaustiva de totes les plataformes en el mercat, podem destacar que a partir d'aquest any comencen a comercialitzar successivament una sèrie d'equips diferents empreses (FLX-GS Roche/454 Life Sciences, Illumina Genome Analyzer i SOLiD™ System de Life Technologies), basades en estratègies diverses amb el denominador comú que totes requereixen una amplificació clonal del DNA abans de procedir a la seqüenciació. Les característiques diferencials principals es posen de manifest en aspectes com la química utilitzada, la longitud de lectura i el nombre de molècules que són capaces de seqüenciar en pa-

ral·lel. Aquesta tecnologia ha provocat que la seqüenciació del DNA resulti més de dues-centes vegades més ràpida i econòmica que amb la seqüenciació tradicional de Sanger, i ha constituït indubtablement una segona generació tecnològica en aquest camp. Com a exemple de la seva capacitat, cal assenyalar que el 2008 una d'aquestes plataformes (FLX-GS Roche/454 Life Sciences) va permetre seqüenciar en tan sols quatre mesos el genoma complet de James D. Watson (Wheeler *et al.*, 2008), un dels descobridors de l'estructura de doble hèlix del DNA. És destacable que el PGH va invertir tretze anys a assolir aquest objectiu.

Tot això ha posat en relleu el seu extraordinari potencial com a alternativa o complement a la seqüenciació tradicional i, durant els últims anys, l'aplicació més estesa de la NGS ha consistit en la seqüenciació de genomes complets amb l'objectiu d'ampliar el coneixement sobre com les diferències genètiques trobades a escala poblacional afecten la salut o poden provocar una determinada malaltia. De fet, gràcies a aquestes i altres tecnologies d'anàlisi massiva, en els últims anys s'han pogut engregar grans projectes multinacionals impensables fa una dècada. Un és el Projecte 1000 Genomes (<http://www.1000genomes.org>), que té com a objectiu establir un catàleg exhaustiu de la variabilitat genètica humana. Igualment, altres projectes que ja estaven en marxa s'han beneficiat de manera significativa amb l'entrada en escena de la seqüenciació d'alt rendiment a baix cost. Aquest és el cas del projecte internacional ENCODE (*Encyclopedia of DNA elements*), l'objectiu del qual és delimitar tots els elements funcionals al llarg de la seqüència del genoma humà (Consortium *et al.*, 2012).

Tot i les seves virtuts, la NGS no està exempta d'inconvenients tècnics, alguns compartits per les diferents plataformes i altres de particulars de cadascuna. Entre

les principals limitacions hi ha, per exemple, la curta longitud de les seqüències que poden llegir les tecnologies Illumina GA i SOLiD™ System, que provoca dificultats en el procés d'acoblament *de novo* i en l'alineament enfront de seqüències de referència, en particular en les regions repetitives (Casals *et al.*, 2012). D'altra banda, no totes les seqüències són processades igualment, i les regions de DNA riques en GC són particularment propenses a una baixa cobertura. Finalment, però no menys important, la introducció d'errors en la seqüència és un tret comú a totes les plataformes: les seqüències més llargues són propenses a tenir errors de lectura, sobretot en els extrems, i les seqüències repetitives i els homopolímers són també motiu d'error en alguns d'aquests seqüenciadors (Hui, 2012). Ara bé, aquesta revolució tecnològica està lluny d'haver-se aturat i això es veu reflectit en una progressiva reducció dels costos, en la simplificació de la metodologia de treball, en l'augment en la capacitat i velocitat de seqüenciació i, sobretot, en el fet que molts dels problemes assenyalats es van corregint a mesura que evoluciona la tecnologia i apareixen noves versions dels equips i dels reactius.

Seqüenciació de tercera generació

Encara no hem assimilat l'impacte d'aquesta nova generació de seqüenciadors quan ja comença a entrar en escena una tercera generació. La principal novetat d'aquesta tercera generació (anomenada *seqüenciació a temps real d'una sola molècula*) és que són capaces de detectar la polimerització de molècules individuals de DNA sense necessitat d'una amplificació prèvia (Schadt *et al.*, 2010). Helicos Biosciences, la primera empresa que va anunciar una tecnologia de seqüenciació d'aquestes característiques, ha desenvolupat la *Helicos' true single mole-*

cule sequencing (tSMS): la seqüenciació detecta la incorporació dels quatre nucleòtids marcats fluorescentment sobre un suport sòlid, i dona lloc a lectures de 30-45 nucleòtids. Més recentment, Pacific Biosciences ha comercialitzat un sistema de tercera generació basat en la tecnologia de detectors *zero-mode waveguide* (ZMW): la paral·lelització d'aquestes estructures nanofòtiques permet la lectura en temps real de la fluorescència de nucleòtids marcats en volums molt reduïts juntament amb temps de carrera molt curts i seqüències de més de mil nucleòtids (Metzker, 2010).

Altres tecnologies de tercera generació, encara no disponibles comercialment, inclouen la detecció de bases de DNA individuals a mesura que passen a través d'un nanopor, que pot ser de naturalesa biològica (Oxford Nanopore) o de naturalesa sintètica (en desenvolupament a partir d'una aliança entre IBM i Roche). Finalment cal esmentar les tècniques de seqüenciació a partir de la imatge directa de les molècules individuals de DNA mitjançant microscòpia electrònica de transmissió desenvolupades, entre d'altres, per Halcyon Molecular (Casals *et al.*, 2012) i ZS Genetics (Schadt *et al.*, 2010). Així doncs, el panorama a pocs anys vista és considerablement variable; no obstant això, les tecnologies de seqüenciació de tercera generació a partir d'una sola molècula encara presenten importants limitacions, i actualment el principal problema és la seva elevada taxa d'error (Schadt *et al.*, 2010).

Noves plataformes NGS de sobretaula

A l'espera que aquesta tercera generació sigui una realitat a tots els nivells, l'evolució de la NGS sembla orientada cap a l'ús clínic rutinari. Per aconseguir aquest objectiu és necessari complir amb una sèrie de requi-

sits com són augmentar la precisió de les tècniques, simplificar els assajos, disposar d'instruments més petits i de baix preu, flexibilitzar-ne el rendiment, reduir els temps de treball i, el més transcendental, aconseguir una anàlisi de les dades simplificada. Des de l'aparició de les primeres plataformes de NGS, s'han produït una sèrie d'avanços tecnològics, com la química de seqüenciació millorada i noves metodologies de detecció de senyals, així com l'optimització dels protocols de preparació de les mostres. Això ha donat com a resultat la disponibilitat de noves plataformes de NGS més petites (també conegudes com a seqüenciadors «de sobretaula») com ara MiSeq (Illumina), 454 Junior GS (Roche) i Ion Torrent PGM (Life Technologies), que suposen una inversió econòmica inicial menor, tenen protocols per a la preparació de les mostres més fàcils, ràpids i automatitzats i, fins i tot, faciliten l'anàlisi de dades per a aplicacions concretes. D'altra banda, algunes de les noves tecnologies de seqüenciació no utilitzen fluorescència o quimioluminescència, cosa que representa un estalvi significatiu de costos per base seqüenciada. Per exemple, Ion Torrent PGM és capaç de detectar i quantificar un senyal elèctric generat pels protons alliberats després de l'addició seqüencial de nucleòtids no marcats en la síntesi de la molècula de DNA. Per tant, aquestes noves plataformes de sobretaula sembla que tenen algunes de les característiques esmentades abans que les fan més adients per a ús clínic. No obstant això, hi ha un consens a dir que encara s'enfronten reptes importants en aquesta transició i, sobretot, s'haurien de simplificar més les eines d'anàlisi de dades per minimitzar la necessitat d'un suport bioinformàtic massa sofisticat (Desai i Jere, 2012).

APLICACIONS CLÍNIQUES DE LA NGS

No és estrany que la revolució de les tecnologies de seqüenciació suposi un profund impacte en la nostra comprensió de la genètica i la biologia del genoma. En el context de la investigació bàsica, la NGS ha estat àmpliament aplicada per a la seqüenciació de genomes *de novo*, la reseqüenciació dirigida de gens, la seqüència de transcritomes i en epigenòmica. Però, indubtablement, una de les contribucions de la NGS que desperta més expectació és l'aplicació clínic. Sense pretendre fer una recopilació exhaustiva, aquestes aplicacions inclouen: la identificació de mutacions causants de malalties genètiques rares a través de la seqüenciació de tot el genoma o de regions discretes del genoma, el diagnòstic molecular de malalties monogèniques causades per múltiples gens, l'estudi de la base genètica de malalties poligèniques comunes, la identificació de variants genètiques rares associades amb trastorns monogènics mendelians, la detecció i caracterització ràpida de microorganismes patògens, la detecció eficient de mutacions hereditàries o somàtiques en els gens relacionats amb el desenvolupament de tumors, el pronòstic i personalització del tractament farmacològic, les proves prenatales invasives i no invasives, la tipificació HLA per a trasplantament d'òrgans i cèl·lules progenitores, etc. Per tant, podem assenyalar que, amb la introducció de les noves plataformes de seqüenciació, les possibilitats de fer estudis genètics per donar resposta a diferents qüestions mèdiques estan augmentant ràpidament. Encara que és impossible descriure totes les aplicacions clíniques potencials de la NGS, dedicarem els apartats següents a fer una breu presentació de les més destacades.

Diagnòstic de malalties monogèniques amb gens responsables coneguts

La seqüenciació sistemàtica dels gens implicats en malalties mendelianes és l'estratègia emprada habitualment per identificar les mutacions responsables en malalties monogèniques o factors de risc coneguts en malalties complexes. No obstant això, el cost de la seqüenciació tradicional, basada en el mètode de Sanger, segueix sent un important obstacle en moltes aplicacions diagnòstiques i el desenvolupament i optimització de solucions moleculars rendibles per donar suport al diagnòstic clínic són sempre ben rebuts. En aquest sentit, malgrat que la NGS és en molts aspectes superior a la seqüenciació tradicional de Sanger, encara hi ha pocs exemples del seu ús efectiu en el diagnòstic de rutina de malalties monogèniques causades per un gen conegut. Això és degut en part al fet que les plataformes de NGS van ser concebudes originalment per seqüenciar genomes complets i aquest no és habitualment l'objectiu en l'anàlisi de malalties monogèniques, sinó que és necessari, abans de la seqüenciació, la captura i enriquiment de la regió d'interès, que pot ser un gen, un grup de gens o una regió genòmica. En aquesta direcció, s'han desenvolupat estratègies imaginatives, incloent-hi la PCR de llarg abast (LR-PCR), la seqüenciació d'amplicons, la captura de seqüències mitjançant xips (per exemple, Nimblegen), la captura de seqüències en solució (per exemple, Agilent Sure Select), etc. Però aquí se'ns planteja un nou problema, ja que si la regió enriquida és molt petita no s'aprofita tota la capacitat del seqüenciador i el preu resultarà prohibitiu. Per tant, un dels reptes més importants en l'aplicació de la NGS a aquestes malalties és la cerca d'estratègies que permetin mesclar un gran nombre de mostres de diferents pacients i seqüenciar-les simultàniament. Ara

bé, després de la seqüenciació en paral·lel, és necessari poder diferenciar d'alguna manera a quina mostra pertany cada seqüència obtinguda. Per abordar aquesta qüestió, el mètode més estès consisteix a indexar les mostres amb un «codi de barres» molecular (seqüències curtes afegides als extrems dels fragments de DNA). Aquest procés es fa habitualment durant la construcció de les biblioteques, un pas fonamental en la preparació de la mostra abans de la seqüenciació. Una alternativa desenvolupada recentment, basada en la tecnologia de microfluids (Fluidigm Corporation), s'ha mostrat com una excel·lent eina per construir les biblioteques a la vegada que s'incorpora el «codi de barres» molecular a cada mostra a un preu raonable (Grossmann *et al.*, 2011). Aquest detall és particularment important, ja que la preparació de les biblioteques és un dels passos més costosos en el procediment complet de la NGS. En conclusió, i tenint en compte la progressió tecnològica dels últims anys, podem afirmar que es comencen a establir les bases perquè properament el diagnòstic molecular de moltes malalties monogèniques mitjançant la NGS s'incorpori a la rutina clínica i reemplaci, en bona part, la seqüenciació tradicional.

Diagnòstic de malalties monogèniques rares en les quals no es coneix el gen responsable

Des de l'aparició de la NGS, el ritme de la identificació de gens responsables de malalties genètiques rares ha augmentat considerablement, circumstància que previsiblement continuarà durant els pròxims anys. La utilitat de la NGS en la cerca del gen o gens responsables d'aquestes malalties és evident, ja que és una tecnologia que ens permet seqüenciar el genoma o l'exoma de l'individu afectat, encara que aquesta no és

una tasca senzilla. El genoma humà conté al voltant de 23.000 gens que s'estenen al llarg de 3.000 milions de parells de bases. Només l'1 % del nostre genoma es tradueix en cadenes d'aminoàcids que donen lloc a les proteïnes, i és aquesta part del genoma, anomenada *exoma*, on es produeixen la gran majoria de mutacions responsables de les malalties hereditàries que coneixem. Per tant, seqüenciar únicament els trenta milions de bases de l'exoma sembla una estratègia lògica en la recerca inicial dels defectes responsables d'aquest tipus de malalties. Per capturar o enriquir l'exoma s'han adaptat amb èxit sistemes com els esmentats abans que es basen en la hibridació del DNA genòmic fragmentat en oligonucleòtids immobilitzats sobre microxips (Nimblegen) o en solució (Agilent Sure Select). De fet, aquesta aproximació ha mostrat un considerable èxit en la identificació del gen responsable de malalties mendelianes quan es disposa de més d'un individu afectat no emparentat o l'evidència de lligament en almenys una família. Fins i tot, s'ha aconseguit identificar la causa de malalties molt rares a partir de la seqüenciació de l'exoma de només un pacient (Choi *et al.*, 2009; Vissers *et al.*, 2010). No obstant això i encara que els resultats són esperançadors, resulta difícil predir la taxa d'èxit que aquesta estratègia pot assolir en un entorn clínic en el qual s'incloguin pacients amb un ampli ventall de manifestacions fenotípiques (Need *et al.*, 2012).

Diagnòstic de malalties monogèniques causades per múltiples gens

Algunes malalties monogèniques com el retràs mental, la sordesa, la cardiomiopatia familiar i la retinitis pigmentosa, entre d'altres, poden ser causades per mutacions en desenes o centenars de gens diferents.

El diagnòstic molecular per a aquestes malalties es basa fonamentalment en l'anàlisi seqüencial dels gens potencials implicats mitjançant la tècnica de seqüenciació convencional de Sanger. Atès que això representa un treball feixuc i costós, aquesta estratègia habitualment no és capaç de satisfer les necessitats diagnòstiques d'aquest tipus de trastorns. L'aparició de la NGS ha permès el desenvolupament de mètodes per analitzar simultàniament múltiples gens, i en resulta un diagnòstic a un preu assequible. Malgrat que aquestes aproximacions no han estat incorporades íntegrament en la prestació de serveis clínics, previsiblement s'incorporaran en els propers anys (O'Sullivan *et al.*, 2012). Com a exemple del potencial i viabilitat de la NGS en el diagnòstic de trastorns genèticament heterogenis en podem destacar l'aplicació a la retinitis pigmentosa. En aquest cas, les proves convencionals són només aplicables a una petita part dels pacients amb la malaltia i permeten identificar les mutacions en menys del 25 % dels individus avaluats. En canvi, la seqüenciació simultània mitjançant la NGS de 105 gens, que se sap que han mutat en alguna de les formes de retinitis, augmenta la taxa d'identificació de mutació fins al 50-55 % (O'Sullivan *et al.*, 2012). D'igual manera, la seqüenciació de l'exoma s'ha plantejat com una alternativa al diagnòstic molecular de rutina en aquest tipus de malalties i podria augmentar encara més l'eficiència en la identificació de la mutació responsable en cada família. La detecció per NGS de variants patogèniques responsables d'un trastorn genèticament heterogeni com la retinitis pigmentosa no és més que el procediment paradigmàtic a seguir en altres malalties causades per múltiples gens. Una vegada posada a punt, l'estratègia és aplicable a la majoria de famílies que pateixen la mateixa malaltia.

Estudi de la base genètica de malalties poligèniques complexes i susceptibilitats

Alguns trastorns, com l'osteoporosi, l'artrosi, la diabetis o la hipertensió arterial, tendeixen a ser més freqüents en determinades famílies, cosa que en reflecteix el component hereditari, encara que el risc de patir-los no s'explica per l'alteració d'un sol gen. Per això la denominació de *malalties poligèniques* o *complexes*. Durant els darrers anys, l'aproximació al coneixement de les causes últimes d'aquestes malalties s'ha dut a terme fent ús dels estudis d'associació del genoma complet o GWAS (Genome Wide Association Study), que han permès identificar locus o regions del genoma associats amb trastorns freqüents com ara l'esquizofrènia, la depressió, la hipertensió, el trastorn bipolar, la malaltia de Crohn, l'artritis reumatoide i la diabetis, entre d'altres. En els estudis de GWA s'utilitzen generalment microxips de SNP (*single nucleotide polymorphisms*) i s'analitzen milers de mostres d'individus diferents. L'enfocament més comú dels estudis GWA és el que compara dos grans grups de persones, un grup control sa i un grup de casos afectats per alguna malaltia. Es fa el genotipatge d'un nombre de SNP variable segons l'estudi, habitualment al voltant del milió, i l'objectiu final és tractar de trobar diferències en les freqüències al·lèliques dels SNP d'ambdós grups i que això marqui una associació amb un gen o regió genòmica que després s'haurà d'estudiar.

Malgrat que els microxips són un potent instrument d'investigació, hi ha una certa resistència a acceptar-los com una eina de diagnòstic a causa de la complexitat en l'anàlisi de les dades, la manca de fiabilitat en la interpretació i la validació insuficient. Fins i tot, la utilització de microxips amb milions de sondes només permet identificar associacions amb variants molt comu-

nes, ja que els SNP utilitzats són els descrits en estudis previs com a més freqüents en les poblacions. A més, els algorismes estadístics per determinar les associacions solament són vàlids utilitzant polimorfismes amb freqüències elevades. A aquest motiu s'ha atribuït que els resultats dels grans estudis de GWA no hagin ofert una informació tan útil com s'esperava. En aquesta etapa pot tenir molt a dir la NGS, ja que la seva naturalesa no esbiaixada i l'alta cobertura que ofereix permeten analitzar no només les variants més comunes, sinó també les variants més rares i no únicament del tipus SNP. Aquests detalls són crucials, ja que avui coneixem que múltiples variacions rares amb un petit efecte actuen conjuntament per donar lloc al fenotip característic. D'altra banda, per als trastorns més comuns, hi ha un ampli ventall de símptomes, susceptibilitats i capacitats de resposta a la teràpia que probablement resulten de la conjunció única del genotip de l'individu i el medi ambient. Únicament tecnologies com la NGS tenen una resolució suficient per detectar les subtils variacions genètiques responsables d'aquests trets diferencials. És en aquest context on la medicina personalitzada propiciada pels sistemes de NGS representa una eina fonamental per poder portar a terme l'anàlisi d'alta resolució per a la identificació dels gens associats a malalties complexes (Desai i Jere, 2012).

Aplicacions de la NGS en el diagnòstic, pronòstic i teràpia del càncer

Atès que el càncer és una malaltia genètica hereditària o desencadenada per mutacions somàtiques, les noves tecnologies de seqüenciació del DNA tenen un impacte significatiu en la detecció, control i tractament de la malaltia. En aquest sentit, hi ha esforços generalitzats per al desenvolupament

pament de protocols basats en la NGS que permetran catalogar les mutacions responsables de múltiples tipus de càncer, fet que es traduirà en el descobriment de noves dianes diagnòstiques, pronòstiques i terapèutiques. Així mateix, és esperable que en els pròxims anys la informació procedent de la seqüenciació completa de genomes s'incorpori cada vegada més als assajos clínics en oncològica, i els resultats obtinguts mitjançant la investigació genòmica haurien de traduir-se en una millora en la pràctica clínica.

Actualment la NGS ja és una realitat pel que fa al diagnòstic clínic de rutina, com per exemple en la seqüenciació de gens d'alta penetrància en càncer familiar (*BRCA1*, *BRCA2*, *APC*, els gens de reparació del DNA, etc.), que estan indicades en persones considerades d'alt risc per la seva història familiar i clínica. La introducció de la NGS suposa un important estalvi dels costos de diagnòstic gràcies a la seqüenciació simultània de múltiples gens i mostres, que redueix els temps de resposta de mesos a setmanes. Un impacte positiu immediat de la reducció del cost de les proves genètiques per a malalties com el càncer de mama i d'ovari és que molts dels individus, que actualment no compleixen amb els estrictes criteris per a la inclusió en les proves genètiques, podran ser analitzats en un cribratge menys restrictiu. En aquest sentit, cal destacar que amb els criteris actuals al voltant del 30-50 % dels individus amb mutació no tenen una història familiar significativa per justificar la prova (Meldrum *et al.*, 2011). D'altra banda, els gens menys freqüentment implicats en síndromes de càncer familiar, i que no són testejats amb regularitat, també podrien incloure's en el cribratge genètic rutinari per conèixer si el risc pot atribuir-se a mutacions en aquests gens.

Un altre paper important de la NGS en

oncologia el trobem a escala terapèutica. Actualment hi ha una sèrie de teràpies dirigides per a diversos tipus de càncer com el melanoma, el càncer de pulmó o el càncer colorectal. L'èxit d'aquests tractaments depèn, en un alt grau, del perfil genètic del tumor tractat, i ja es coneixen diverses mutacions en determinats gens predictives de l'èxit o fracàs d'un fàrmac concret. A més, és probable que les mutacions somàtiques, ja siguin primàries o adquirides, també puguin contribuir com un mecanisme de recaiguda de la malaltia. Per exemple, s'ha descrit la mutació *p.Thr790Met* en el gen *EGFR* com un factor de resistència per a teràpies dirigides amb potencial per predir la probabilitat de recaiguda. Per tant, el seguiment mitjançant NGS de la resposta de la malaltia a la teràpia, per detectar la resistència emergent, i les proves genètiques concomitants per a la detecció de mutacions *de novo*, probablement formaran part de moltes estratègies terapèutiques futures (Meldrum *et al.*, 2011).

D'altra banda, es comencen a conèixer els primers resultats de la utilització clínica de la NGS de panells de múltiples gens en mostres tumorals de pacients. Per exemple, un estudi recent ha reportat l'ús de la seqüenciació massiva paral·lela de 138 gens específics en mostres de melanoma que ha permès descobrir un nou mecanisme de resistència adquirida en l'inhibidor de *RAF* PLX4032 (Wagle *et al.*, 2011). No obstant això, la seqüenciació de mostres tumorals planteja dificultats en la interpretació de les noves troballes en la pràctica clínica, ja que l'anàlisi bioinformàtica de les dades de seqüenciació d'espècimens tumorals és especialment complexa per les característiques inherents de les mostres: aneuploidia, heterogeneïtat i la contaminació potencial amb teixit normal. Per aquest motiu, l'elaboració de bones bases de coneixement dels gens i les mutacions que confereixin resistència

a un agent diana terapèutic serà transcendent per a la interpretació dels resultats de la seqüenciació i el desenvolupament de teràpies combinades eficaces (Meldrum *et al.*, 2011).

Finalment, encara que aquest capítol està dedicat a les aplicacions clíniques de la NGS, volem fer esment de la importància de l'epigenètica, que, malgrat no tenir una aplicació rutinària, molt probablement en els propers anys gaudirà d'aquest paper en el camp de l'oncologia. El terme *epigenètica* es refereix normalment a modificacions bioquímiques reversibles del DNA i proteïnes associades (com les histones) que afecten la modulació de l'expressió gènica sense alteració directa de la seqüència del DNA. Entre aquestes, la metilació del DNA és una de les més estudiades i amb més implicacions regulatòries. Les anomalies epigenètiques, com la metilació aberrant de les illes CpG, s'hereten a través de les divisions cel·lulars i tenen un paper important en la carcinogènesi. L'epigenètica va ser una de les primeres disciplines a treure profit del desenvolupament de la NGS, combinant-la amb mètodes ja establerts per capturar regions genòmiques modificades epigenèticament (Meaburn i Schulz, 2012), i no sols centrant-se en la metilació en CG, sinó analitzant els patrons globals de metilació que permet la NGS. A més dels estudis individuals, la NGS ha permès donar impuls a grans consorcis internacionals com l'International Human Epigenome Consortium (<http://www.epigenome.org>), que contribuirà al coneixement i la comprensió de les funcions de l'epigenètica subjacents a les malalties humanes. És indiscutible que els estudis epigenètics mitjançant la NGS ja tenen aplicacions múltiples en el terreny de la investigació (Miyamoto i Ushijima, 2005). Ara bé, probablement les primeres proves rutinàries es desenvoluparan en l'àmbit de l'oncologia per a la detecció de patrons

de metilació aberrants amb distintes aplicacions pràctiques: per detectar les cèl·lules tumorals en mostres de biòpsia o en el DNA lliure en el plasma; per fer un pronòstic de resposta a la quimioteràpia i a l'aparició d'efectes adversos; com a marcador associat amb un risc o susceptibilitat al desenvolupament d'un determinat càncer (Brosnan i Iacobuzio-Donahue, 2012).

Aplicacions de la NGS en microbiologia

Probablement la microbiologia clínica és una de les especialitats que més rendiment està obtenint de les noves tecnologies de seqüenciació. Comencem per dir que la seqüenciació massiva de microorganismes ha revolucionat la identificació bacteriana i fúngica, i ha permès millorar el diagnòstic de moltes malalties infeccioses, però també té altres aplicacions de diversa naturalesa, com ara la seqüenciació d'alt rendiment de genomes complets, el descobriment de nous bacteris i virus mitjançant aproximacions metagenòmiques, la investigació de les comunitats microbianes en el medi ambient i en el cos humà en contextos de salut i de malaltia, l'anàlisi de la variabilitat del genoma víric en l'hoste, el monitoratge ultrasensible de resistència a fàrmacs antivírics, la recerca de la diversitat vírica i l'avaluació del viroma humà (Barzon *et al.*, 2011). És impossible descriure totes aquestes aplicacions en poques línies, de manera que ens limitarem a comentar només alguns exemples il·lustratius.

En el camp de la virologia, l'esforç de seqüenciació s'ha centrat essencialment en organismes de gran rellevància clínica, i disset dels vint virus seqüenciats són causants de malalties en els éssers humans. La informació generada té repercussió en diferents àrees de la virologia, que van des del diagnòstic fins a la patogènesi, passant pel

disseny de vacunes (Radford *et al.*, 2012). Un dels millors exemples de l'aplicació clínica de la seqüenciació massiva en aquest camp és la caracterització de quasiespècies víriques en mostres de pacients infectats. El terme *quasiespècie* fa referència a la complexitat genotípica de les poblacions víriques deguda a les altes taxes de mutació inherents a la replicació dels virus RNA. Això dona lloc que diferents versions del genoma del virus, amb només lleugeres variacions en la seqüència, es trobin presents en un mateix individu en un moment determinat. Aquesta naturalesa plural de les poblacions víriques les permet adaptar-se ràpidament als canvis de l'entorn mitjançant la replicació selectiva de variants amb una millor condició per al nou escenari. Per tant, una caracterització precisa de les quasiespècies en mostres de pacients és crítica, ja que de la composició d'aquestes dependrà l'evolució vírica, la virulència, l'evasió enfront la resposta immunitària, el disseny de vacunes i la resistència als fàrmacs antivírics (Barzon *et al.*, 2011). Prenent com a exemple aquesta darrera característica, podem esmentar que el mètode de referència per detectar la resistència als fàrmacs del VIH deguda a mutacions concretes és la seqüenciació convencional directa de RT-PCR. Una limitació important d'aquest mètode és la seva ineficàcia per detectar variants presents en la quasiespècie en menys d'un 20-25 %. Això és especialment negatiu, ja que diversos estudis han demostrat que aquestes variants resistents no detectades són clínicament rellevants i sovint responsables del fracàs d'una nova pauta de tractament antiretrovíric. Per contra, la característica d'amplificació clonal de la NGS, unida a la profunditat de seqüència a la qual pot arribar, permet detectar variants extremament poc representades en la població, fet que la fa idònia per portar a terme la caracterització de les quasies-

pècies de manera senzilla, ràpida i econòmica.

Una altra aplicació pràctica de la NGS, amb un gran impacte en el camp de la microbiologia, és l'estudi de la diversitat microbiana de mostres ambientals i clíniques. El microbioma humà és el conjunt dels microorganismes, els seus genomes i les interaccions d'aquests amb l'ambient en què viuen. Una eina fonamental per estudiar-lo és la metagenòmica o l'anàlisi, mitjançant les noves tecnologies de seqüenciació, del genoma dels organismes d'una població sense necessitat d'aïllar i cultivar cadascun d'aquests organismes. L'objectiu en aquest cas no és el d'obtenir informació relativa a la bioquímica o al metabolisme de l'organisme, sinó més aviat l'obtenció d'empremtes específiques que permetin identificar les espècies presents en la mostra, establir-ne el nombre aproximat i deduir la distància genètica que les separa. La metagenòmica serveix també per estudiar la resposta d'una comunitat de microorganismes davant determinats factors i per comprovar com es modifica el conjunt de genomes d'aquesta comunitat en resposta a diferents estímuls o malalties. Per exemple, se sap que els canvis en les comunitats microbianes al cos estan estretament relacionats amb la funció del sistema immunitari, l'obesitat o el càncer. Sobre la base d'aquesta aproximació, el 2007 els instituts nacionals de la salut (NIH) dels Estats Units es van embarcar en el Projecte Microbioma Humà (Human Microbiome Project o HMP) i el 2008 la Comissió Europea i la Xina van crear el seu homòleg, MetaHIT (Metagenomics of the Human Intestinal Tract).

L'objectiu de l'HMP és obtenir un mapa complet del microbioma humà, i per això s'analitzen els gens de la microflora intestinal, la flora de la cavitat bucal i vaginal i els microorganismes que es troben a la pell. Els investigadors ja han pogut

estimar que més de deu mil espècies de microorganismes viuen en associació amb els éssers humans. Anteriorment tan sols uns pocs centenars d'espècies de bacteris havien estat identificats. Tenint en compte que actualment estan en marxa al voltant de mil projectes de seqüenciació de bacteris en diferents graus de desenvolupament, encara estem lluny de cobrir l'amplitud de la diversitat filogenètica del microbioma humà. Així mateix, han posat de manifest que cada indret del cos humà té la seva pròpia «signatura» de microorganismes i que la diversitat taxonòmica i genètica és més gran en les mostres de dents i femta, intermèdia en pell i en la superfície interna de la galta, i baixa en les mostres vaginals.

Per la seva banda, el projecte MetaHIT té com a objectiu tractar de trobar associacions entre gens del microbioma intestinal i la malaltia, centrant-se especialment en dues àrees amb un creixent interès clínic a Europa: l'obesitat i la malaltia inflammatòria intestinal. Els primers resultats publicats d'aquest projecte mostren troballes sorprenents. Després d'analitzar la flora intestinal de gairebé dues-centes persones de sis nacionalitats diferents, els investigadors del projecte esperaven trobar diferències en el microbioma segons races, nacionalitats, entorn o tipus de dieta. Per contra, els resultats han agrupat els humans (independentment de la seva procedència) en tres grans grups, segons el bacteri dominant. És aquest bacteri el que determina quines altres espècies hi conviuen i, per tant, quines altres espècies conformen el microbioma d'un individu (Arumugam *et al.*, 2011). Els resultats d'ambdós projectes representaran un primer pas per a l'enteniment de com el microbioma humà contribueix i es correlaciona amb diferents aspectes de la salut i la malaltia, de manera que, en el futur, l'anàlisi del microbioma de cada individu sigui un paràmetre biomètric d'utilitat clínica.

Diagnòstic prenatal no invasiu

El diagnòstic prenatal no invasiu (DPNI) consisteix a fer el diagnòstic prenatal a partir de mostres de DNA fetal lliure en plasma de dones gestants, i actualment és un dels camps en què la NGS té un potencial extraordinari. El descobriment el 1997 de la presència de DNA fetal lliure en el plasma matern va obrir noves possibilitats per a l'estudi molecular d'aquest DNA o de les cèl·lules fetals mateixes en sang materna (Lo *et al.*, 1997). Durant els primers anys, la investigació en aquest camp es va centrar en el desenvolupament de proves de determinació de l'Rh del fetus (detecció del gen *D*) per evitar la malaltia hemolítica del nou-nat o de determinació del sexe fetal (detecció del cromosoma Y) per a malalties lligades al sexe. El DPNI és un camp d'investigació de gran interès i llarg recorregut i, malgrat que encara està en procés de desenvolupament, en el futur serà una realitat que evitarà el diagnòstic prenatal mitjançant procediments de risc (Lo, 2012; Metzker, 2010).

S'ha iniciat el camí per poder portar a terme el DPNI d'aneuploidies cromosòmiques i malalties monogèniques mitjançant mètodes basats en la NGS. El 2007 es va demostrar que els mètodes que permeten analitzar i comptabilitzar molècules individuals de DNA són capaces de detectar la presència d'una trisomia fetal, sempre que el nombre de molècules analitzades sigui suficient. Encara que el DNA fetal representa una petita fracció del DNA present en el plasma matern, gràcies al rendiment de la NGS milions de molècules d'una regió de DNA es poden seqüenciar fàcilment de manera individual. D'aquesta manera, és possible detectar directament una aneuploidia fetal a partir d'una quantificació precisa de les seqüències totals (fetals + maternes) del cromosoma aneuploide i determinar si estan excessivament o insuficientment represen-

tades en comparació amb les seqüències d'altres cromosomes. Amb aquesta aproximació, la NGS ha permès fer DPNI de fetus amb trisomies 13, 18 i 21.

D'altra banda, tot apunta que la tecnologia de referència per al DPNI de malalties monogèniques es basarà també en mètodes d'anàlisi d'una sola molècula, com ara la PCR digital (Tsui *et al.*, 2011) o la NGS (Lo, 2012). Per exemple, la NGS es pot aplicar al diagnòstic de malalties monogèniques heretades per via paterna, ja que l'allel fetal procedent del pare portador de la mutació no està present en el genoma matern i es pot diferenciar fàcilment en el plasma de la gestant mitjançant seqüenciació massiva de la regió corresponent. Fins i tot, en els casos més complexos, com són les malalties lligades al sexe, en les quals és la mare la que transmet la mutació al fetus, hi ha aproximacions alternatives. L'estratègia utilitzada en aquest cas és la denominada *dosi relativa de mutació* (DRM), que es basa en una quantificació molt precisa dels allels mutant i salvatge en el plasma de les dones heterozigotes gestants de fetus masculins. L'allel trobat més freqüentment indicarà quin és el genotip del fetus.

Així doncs, quan la detecció rutinària de mutacions puntuals o aneuploidies a partir de plasma matern estigui suficientment validada, permetrà obviar l'amniocentesi o la biòpsia de vellositats corials, fet que suposarà una revolució sense precedents en el camp del diagnòstic prenatal.

Tipificació HLA per al trasplantament al·logènic d'òrgans i de cèl·lules progenitores

El sistema HLA (*human leukocyte antigen*) és el sistema genètic més polimòrfic en els éssers humans. Està implicat en el reconeixement immunitari i en la senyalitza-

ció entre cèl·lules, i la seva tipificació és fonamental per determinar la compatibilitat existent entre el donant i el receptor, condició imprescindible per a l'èxit del trasplantament al·logènic d'òrgans i de cèl·lules progenitores. La combinació dels nombrosos polimorfismes dels diferents gens que formen el sistema HLA dona lloc a milers d'allels, que es classifiquen en grups. La tipificació es pot fer a diferents resolucions en funció de si únicament s'identifica el grup (baixa resolució) o es determina l'allel individual (alta resolució). Inicialment la identificació dels polimorfismes de l'HLA es va fer mitjançant l'ús de tècniques serològiques, però a partir del desenvolupament de la PCR els mètodes basats en l'amplificació del DNA han predominat en els laboratoris d'histocompatibilitat. L'anàlisi del DNA genòmic és més exacta i reproducible, i ofereix més flexibilitat i resolució. Actualment s'empren diverses tècniques basades en el DNA (Dunn, 2011), incloent-hi la PCR de *primers* específics de seqüència (PCR-SSP), la PCR de sondes d'oligonucleòtids de seqüència específica (PCR-SSOP), la tipificació basada en seqüència (PCR-SBT) i la PCR a temps real (RT-PCR). Totes són extremament útils, però, alhora, totes tenen desavantatges. És un consens que la tipificació de baixa resolució no és suficient per al trasplantament i resulta en més mortalitat, així com en una malaltia de l'empelt contra l'hoste més aguda, i disminueix les possibilitats d'èxit del trasplantament. Per tant, actualment es persegueix la tipificació d'alta resolució segons dues vies alternatives: una és a partir d'una aproximació preliminar de baixa resolució que després es completa seleccionant sondes i *primers* específics per al genotipatge d'alta resolució. L'altra és la selecció de la seqüenciació com a mètode inicial, complementant-la a continuació amb PCR-SSP o PCR-SSOP per resoldre les ambigüitats. Per tant, la combi-

nació de tècniques actuals per resoldre les ambigüitats no només és laboriosa sinó també costosa i lenta.

Afortunadament, la recerca de noves maneres de treballar amb el creixent nombre d'allels descrits i la complexitat de la resolució d'aquests allels ha donat lloc a publicacions recents que demostren la utilitat de la NGS en la tipificació HLA d'alta resolució i d'alt rendiment sense els problemes experimentats amb altres mètodes (Gabriel *et al.*, 2011). Pràcticament tots aquests treballs es basen en l'ús de la plataforma GS-FLX de Roche, fonamentalment a causa de la longitud de lectura aconseguida (200-400 pb), considerablement més gran que l'obtinguda amb les altres plataformes. Aquesta longitud de lectura, juntament amb l'amplificació clonal, permet eliminar gran part de les ambigüitats de fase trobades en la PCR-SBT mitjançant el mètode de Sanger. Per tant, és imminent la tipificació de l'HLA per NGS, però encara és necessari introduir algunes millores en el processament, en el programari i en els reactius utilitzats abans que es pugui implementar en el diagnòstic de rutina. En aquest sentit, els nous seqüenciadors de sobretaula esmentats abans sembla que poden ser els més adequats des del punt de vista econòmic i per la possibilitat de dissenyar un procediment de treball totalment automatitzable. De la mateixa manera, la flexibilitat d'aquestes noves plataformes permetrà una transició senzilla des de les aplicacions de seqüenciació ja existents, basades en l'electroforesi capil·lar (Dunn, 2011).

ANÀLISI BIOINFORMÀTICA DE LES DADES

No és necessari dir que el desenvolupament que ha experimentat la genètica en les últimes dècades hauria estat impossible

sense el desenvolupament en paral·lel de la informàtica. Tampoc no cal esmentar que la ingent quantitat de dades derivada de la seqüenciació massiva d'individus i genomes suposa un problema a l'hora d'emmagatzemar-les, tractar-les i comparar-les. Encara que la discussió detallada de la bioinformàtica necessària associada a la NGS queda fora dels objectius d'aquest capítol, no podem deixar d'esmentar els problemes que suposen alguns dels processos més delicats del circuit d'anàlisi, com són la qualitat de seqüència, l'alineament i l'anotació, entre d'altres. Així mateix, la manipulació dels arxius de sortida característics, incloent tant les lectures individuals com els seus alineaments en seqüències contínues, pot ser un repte a causa de la seva grandària, sobretot quan parlem d'exomes o genomes complets (Radford *et al.*, 2012). Podria dir-se que ara és fàcil generar molts milions de bases de seqüència que superen la capacitat de còmput de la majoria dels sistemes i que el repte pot ser l'ús eficient. Aquest escenari és molt desitjable en el marc de la investigació bàsica, en què l'atenció se centra en l'obtenció de tanta informació com sigui possible sense la restricció del temps emprat a analitzar les dades per obtenir resultats significatius. Per contra, la gran quantitat de dades generades pot resultar un important obstacle en el trasllat a la rutina clínica.

Des de l'anàlisi de només un gen en el passat recent, passant per l'anàlisi de panells de múltiples gens, fins a la seqüenciació de l'exoma o del genoma complet, hem vist com la complexitat de la tecnologia i la bioinformàtica augmentaven espectacularment i les seves aplicacions clíniques han demostrat ser més complicades del que es plantejava inicialment. El cost de la infraestructura informàtica i de la mà d'obra altament qualificada és enorme i fins i tot inassequible per a petits laboratoris de diagnòstic. El preu d'administrar, emma-

gatzemar i analitzar dades de NGS es dispara fàcilment en centenars de milers d'euros i actualment sembla un obstacle greu per a la implementació de sistemes de NGS en la rutina clínica (Desai i Jere, 2012; Guigó, 2012). Per tant, en la transició de la NGS cap a la clínica, un dels reptes principals és la simplificació de l'anàlisi de la gran quantitat de dades generades per obtenir informació útil de manera eficient, la simplificació del procés d'identificació de les mutacions/variacions i també l'aplicació en el laboratori de rutina feta per persones sense una formació bioinformàtica excepcional. Si aconseguim caminar en aquesta direcció, començarem a extreure un benefici extraordinari de les múltiples aplicacions de la NGS en la clínica, incloent-hi la seqüenciació rutinària del genoma.

PERSPECTIVES DE FUTUR

Totes les aplicacions esmentades fins aquí són només el principi. L'impacte de la NGS en el diagnòstic molecular i tot el que comporta en l'àmbit de la salut humana evolucionarà durant les pròximes dècades a una velocitat difícilment predictable ara mateix. Com ja s'ha comentat anteriorment, encara estem assimilant l'impacte de la denominada *seqüenciació de nova generació*, mentre les plataformes de tercera generació comencen a mostrar les seves potencialitats. D'acord amb les estimacions de la revista *Nature*, per a finals de l'any 2012 s'espera que en el món s'hauran seqüenciat cent mil genomes humans complets («Human genome: genomes by the thousand», Schadt *et al.*, 2010). Cal recordar que només ha passat una dècada des de la finalització del Projecte Genoma Humà i abans d'aquesta data el nombre de genomes seqüenciats era zero. Unes poques dotzenes el 2009, 2.700 l'any 2010 i 30.000 a la fi de 2011. A aquest ritme expo-

nencial, per a l'any 2020 podria ser factible, almenys matemàticament, descodificar el DNA de cada membre de la humanitat en només 12 mesos. Així com qualsevol PC portàtil actual té la capacitat dels grans ordinadors centrals de fa poques dècades, és previsible que la seqüenciació del genoma complet estarà aviat accessible i assequible per a qualsevol persona en el nostre entorn. La capacitat d'emmagatzemar tota aquesta informació de manera eficient ha millorat ostensiblement durant els últims anys. De fet, avui es podria emmagatzemar la informació rellevant del genoma propi en solament unes poques megabases i tenir-la allotjada en el núvol, i és molt probable que en pocs anys la majoria de nosaltres tinguem la nostra pròpia informació genètica en lloc segur per un cost modest. Un problema diferent és la interpretació d'aquestes dades, que de moment resulta extremament complexa quan es tracta d'analitzar malalties i caràcters multifactorials o susceptibilitats i predisposicions genètiques enfront de determinats factors o variables mediambientals. Sens dubte, aquesta interpretació millorarà en els pròxims anys, una vegada que tinguem milions de genomes seqüenciats, a mesura que es vagin coneixent més dades sobre la variabilitat genètica humana i es desenvolupin algorismes apropiats per esbrinar què signifiquen en realitat cadascuna de les variacions poblacionals i individuals (actualment n'hi ha descrites més de vint-i-dos milions). El repte futur i l'oportunitat per a organismes públics i empreses privades es trobarà en el desenvolupament d'aquests algorismes a mida per a determinades malalties o condicions. Si l'obtenció de la seqüència del genoma és la primera etapa, desxifrar-la és la segona. La gran conseqüència següent serà sens dubte el veritable desenvolupament de la medicina predictiva i personalitzada. Aquests conceptes, dels quals s'ha abu-

sat considerablement en el passat, és molt probable que, ara sí, estiguin a les portes de convertir-se en una realitat; si més no s'estan establint les bases perquè sigui possible des del punt de vista tecnològic. De fet, tal com s'ha apuntat, en especialitats com l'oncologia o la virologia comença a ser una realitat quan el perfil genètic del pacient o del virus s'utilitza per a la prescripció del fàrmac més adequat en determinats processos tumorals o infecciosos. La seqüenciació de gens individuals i l'ús de xips de DNA en les proves genètiques serà reemplaçat en la pròxima dècada per la seqüenciació del genoma complet, i quan això succeeixi, la medicina predictiva començarà a arribar al seu ple potencial.

És evident que la NGS i la seqüenciació de l'exoma o del genoma complet és una eina prometedora amb un gran potencial per millorar l'assistència sanitària i la salut humana en general i, especialment, per prevenir malalties en les persones amb certes predisposicions de base genètica. No obstant això, també planteja importants consideracions ètiques que haurien de tenir-se en compte analitzant-les des de diferents punts de vista. L'estudi genètic d'humans sempre ha estat controvertit des del punt de vista ètic i legal: la preocupació pel consentiment informat, el perill d'estigmatització (és a dir, a ser jutjat o etiquetat com a conseqüència dels resultats de proves genètiques), la privacitat (tant individual com de grup), els interessos d'entitats patronals i asseguradores, la discriminació deguda a les característiques genètiques, etc. Totes les qüestions plantejades per les tecnologies anteriors s'exacerben amb l'aparició de la NGS, i a més planteja noves polèmiques. Encara més, des del moment que es pot arribar a considerar la seqüenciació d'exomes o genomes complets com a mètode de diagnòstic genètic d'aplicació universal, en lloc de la seqüenciació de regions

discretes del genoma. En terme mitjà, cada persona és portadora aproximadament de 250 a 300 variants amb pèrdua de funció en gens anotats i de 50 a 100 variants prèviament descrites com a causants de malalties hereditàries (Genomes Project, 2010). Per tant, la seqüenciació d'exomes o genomes no només revela la mutació causant de la malaltia, sinó que també pot posar de manifest mutacions causants d'altres malalties monogèniques o factors genètics de risc per a malalties complexes. Atès que és molt probable que en un futur no gaire llunyà aquest tipus d'aproximacions sigui d'aplicació generalitzada en la població, la polèmica clàssica sobre el dret a no saber i a no ser discriminat a causa de les característiques genètiques es torna a plantejar. Una solució a aquesta qüestió implicaria tractar les dades genètiques com qualsevol altre tipus d'informació personal, el propietari de la qual és la pròpia persona. Atès que les dades de seqüència sense una interpretació no tenen valor clínic rellevant, únicament s'haurien d'analitzar les dades corresponents als gens o regions necessaris per portar a terme el diagnòstic requerit en un moment determinat i descartar o emmagatzemar les dades de la seqüència irrellevants, fins a usos futurs amb el consentiment explícit de la persona. No obstant això, es pot presentar una segona polèmica, ja que el dret de la persona a no saber pot desencadenar que la seva descendència pateixi malalties genètiques que es podrien haver evitat en cas d'haver estat informat. Per tant, la nova qüestió és si la informació genètica és propietat de l'individu o de tota la família, incloent-hi els que encara no han nascut. No és l'objectiu d'aquest capítol debatre exhaustivament sobre les diferents qüestions ètiques i legals que es plantejaran, però és necessari posar en relleu que la transició a una nova etapa en la investigació centrada en el genoma en el seu conjunt

comporta qüestions ètiques cada vegada més abundants i més complexes, i els comitès ètics són i continuaran sent cada vegada més cridats a revisar l'ètica de la investigació que implica l'ús d'aquestes tecnologies emergents en la investigació amb éssers humans fins que els legisladors desenvolupin un consens i una orientació ètica consistent (Diamandis, 2009). Al fil d'aquestes consideracions és difícil no recordar una de les grans pel·lícules del cinema de ciència ficció: *Gattaca*, d'Andrew Niccol (1997). El protagonista, Vincent Anton Freeman, és un individu «no vàlid» nascut amb un 99 % de probabilitats de morir per fallada cardíaca i amb una esperança de vida en néixer de trenta anys, segons es desprèn de la interpretació de la seqüència del seu genoma. Per aquest únic motiu, se li nega la possibilitat de ni tan sols intentar convertir-se en pilot espacial. No és previsible que la nostra societat camini en aquesta direcció però cal tenir en compte que el que l'any 1997 semblava molt llunyà des del punt de vista tècnic, està convertint-se avui en una realitat. No obstant els perills potencials, els beneficis que en general suposarà tenir accés a la seqüència genòmica pròpia per a l'ús en la gestió de la salut individual tenen un potencial suficient per fer el treball que sigui necessari per a la resolució dels riscos ètics i les qüestions legals que es puguin plantejar. Superades aquestes consideracions, és inqüestionable que en els pròxims anys les noves tecnologies de seqüenciació tindran un efecte transformador en la utilització i protagonisme de la informació genètica en la rutina clínica que suposarà un impacte en la salut del segle XXI sense precedents.

BIBLIOGRAFIA

ARUMUGAM, M. [et al.] (2011). «Enterotypes of the human gut microbiome». *Nature*, 473: 174-180.

- BARZON, L.; LAVEZZO, E.; MILITELLO, V.; TOPPO, S.; PALU, G. (2011). «Applications of next-generation sequencing technologies to diagnostic virology». *Int. J. Mol. Sci.*, 12: 7861-7884.
- BROSNAN, J. A.; IACOBUZIO-DONAHUE, C. A. (2012). «A new branch on the tree: next-generation sequencing in the study of cancer evolution». *Semin. Cell Dev. Biol.*, 23: 237-242.
- CASALS, F.; IDAGHDOUR, Y.; HUSSIN, J.; AWADALLA, P. (2012). «Next-generation sequencing approaches for genetic mapping of complex diseases». *J. Neuroimmunol.*, 248: 10-22.
- CHOL, M.; SCHOLL, U. I.; JI, W.; LIU, T.; TIKHONOVA, I. R.; ZUMBO, P.; NAYIR, A.; BAKKALOGLU, A.; OZEN, S.; SANJAD, S.; NELSON-WILLIAMS, C.; FARHI, A.; MANE, S.; LIFTON, R. P. (2009). «Genetic diagnosis by whole exome capture and massively parallel DNA sequencing». *Proc. Natl. Acad. Sci. USA*, 106: 19096-19101.
- CONSORTIUM, E. P.; BERNSTEIN, B. E.; BIRNEY, E.; DUNHAM, I.; GREEN, E. D.; GUNTER, C.; SNYDER, M. (2012). «An integrated encyclopedia of DNA elements in the human genome». *Nature*, 489: 57-74.
- DESAL, A. N.; JERE, A. (2012). «Next-generation sequencing: ready for the clinics?». *Clin. Genet.*, 81: 503-510.
- DIAMANDIS, E. P. (2009). «Next-generation sequencing: a new revolution in molecular diagnostics?». *Clin. Chem.*, 55: 2088-2092.
- DUNN, P. P. (2011). «Human leucocyte antigen typing: techniques and technology, a critical appraisal». *Int. J. Immunogenet.*, 38: 463-473.
- GABRIEL, C.; STABENTHEINER, S.; DANZER, M.; PROLL, J. (2011). «What next? The next transit from biology to diagnostics: next generation sequencing for immunogenetics». *Transfus. Med. Hemother.*, 38: 308-317.
- GENOMES PROJECT, C. (2010). «A map of human genome variation from population-scale sequencing». *Nature*, 467: 1061-1073.
- GIBBS, R. A. (2011). «Genome-sequencing anniversary. Bringing genomics and genetics back together». *Science*, 331: 548.
- GROSSMANN, V.; KOHLMANN, A.; EDER, C.; HAFERLACH, C.; KERN, W.; CROSS, N. C.; HAFERLACH, T.; SCHNITTEGER, S. (2011). «Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in > 80% of patients with TET2 and EZH2 being of high prognostic relevance». *Leukemia*, 25: 877-879.
- GUIGÓ, R. (2012). «Biologia i computació». *Treb. Soc. Cat. Biol.*, 63: 183-198.
- HUI, P. (2012). «Next generation sequencing: chemistry, technology and applications». *Top. Curr. Chem.* [En premsa]

- «Human genome: genomes by the thousand». (2010). *Nature*, 467: 1026-1027.
- LO, Y. M. (2012). «Non-invasive prenatal diagnosis by massively parallel sequencing of maternal plasma DNA». *Open Biol.*, 2: 120086.
- LO, Y. M.; CORBETTA, N.; CHAMBERLAIN, P. F.; RAI, V.; SARGENT, I. L.; REDMAN, C. W.; WAINSCOT, J. S. (1997). «Presence of fetal DNA in maternal plasma and serum». *Lancet*, 350: 485-487.
- MAXAM, A. M.; GILBERT, W. (1977). «A new method for sequencing DNA». *Proc. Natl. Acad. Sci. USA*, 74: 560-564.
- McPHERSON, J. D. [et al.] (2001). «A physical map of the human genome». *Nature*, 409: 934-941.
- MEABURN, E.; SCHULZ, R. (2012). «Next generation sequencing in epigenetics: insights and challenges». *Semin. Cell Dev. Biol.*, 23: 192-199.
- MELDRUM, C.; DOYLE, M. A.; TOTHILL, R. W. (2011). «Next-generation sequencing for cancer diagnostics: a practical perspective». *Clin. Biochem. Rev.*, 32: 177-195.
- METZKER, M. L. (2010). «Sequencing technologies - the next generation». *Nat. Rev. Genet.*, 11: 31-46.
- MIYAMOTO, K.; USHIJIMA, T. (2005). «Diagnostic and therapeutic applications of epigenetics». *Jpn. J. Clin. Oncol.*, 35: 293-301.
- NEED, A. C.; SHASHI, V.; HITOMI, Y.; SCHOCH, K.; SHIANNA, K. V.; McDONALD, M. T.; MEISLER, M. H.; GOLDSTEIN, D. B. (2012). «Clinical application of exome sequencing in undiagnosed genetic conditions». *J. Med. Genet.*, 49: 353-361.
- O'SULLIVAN, J.; MULLANEY, B. G.; BHASKAR, S. S.; DICKERSON, J. E.; HALL, G.; O'GRADY, A.; WEBSTER, A.; RAMSDEN, S. C.; BLACK, G. C. (2012). «A paradigm shift in the delivery of services for diagnosis of inherited retinal disease». *J. Med. Genet.*, 49: 322-326.
- RADFORD, A. D.; CHAPMAN, D.; DIXON, L.; CHANTREY, J.; DARBY, A. C.; HALL, N. (2012). «Application of next-generation sequencing technologies in virology». *J. Gen. Virol.*, 93: 1853-1868.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. (1977). «DNA sequencing with chain-terminating inhibitors». *Proc. Natl. Acad. Sci. USA*, 74: 5463-5467.
- SCHADT, E. E.; TURNER, S.; KASARSKIS, A. (2010). «A window into third-generation sequencing». *Hum. Mol. Genet.*, 19: R227-240.
- TSUI, N. B.; KADIR, R. A.; CHAN, K. C.; CHI, C.; MELLARS, G.; TUDDENHAM, E. G.; LEUNG, T. Y.; LAU, T. K.; CHIU, R. W.; LO, Y. M. (2011). «Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA». *Blood*, 117: 3684-3691.
- VENTER, J. C. [et al.] «The sequence of the human genome». *Science*, 291: 1304-1351.
- VISSERS, L. E.; LIGT, J. de; GILISSEN, C.; JANSSEN, I.; STEEHOUWER, M.; VRIES, P. de; LIER, B. van; ARTS, P.; WIESKAMP, N.; ROSARIO, M. del; BON, B. W. van; HOISCHEN, A.; VRIES, B. B. de; BRUNNER, H. G.; VELTMAN, J. A. (2010). «A de novo paradigm for mental retardation». *Nat. Genet.*, 42: 1109-1112.
- WAGLE, N.; EMERY, C.; BERGER, M. F.; DAVIS, M. J.; SAWYER, A.; POCHANARD, P.; KEHOE, S. M.; JOHANNESSEN, C. M.; MACCONAILL, L. E.; HAHN, W. C.; MEYERSON, M.; GARRAWAY, L. A. (2011). «Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling». *J. Clin. Oncol.*, 29: 3085-3096.
- WATSON, J. D.; CRICK, F. H. (1953). «Molecular structure of nucleic acids, a structure for deoxyribose nucleic acid». *Nature*, 171: 737-738.
- WHEELER, D. A.; SRINIVASAN, M.; EGHOLM, M.; SHEN, Y.; CHEN, L.; MCGUIRE, A.; HE, W.; CHEN, Y. J.; MAKHJANI, V.; ROTH, G. T.; GOMES, X.; TARTARO, K.; NIAZI, F.; TURCOTTE, C. L.; IRZYK, G. P.; LUPSKI, J. R.; CHINAULT, C.; SONG, X. Z.; LIU, Y.; YUAN, Y.; NAZARETH, L.; QIN, X.; MUZYNY, D. M.; MARGULIES, M.; WEINSTOCK, G. M.; GIBBS, R. A.; ROTHBERG, J. M. (2008). «The complete genome of an individual by massively parallel DNA sequencing». *Nature*, 452: 872-876.

SOBRE L'AUTOR

Francisco Vidal (Villablino, Lleó, 1966) és facultatiu adjunt de Banc de Sang i Teixits. Va estudiar biologia a la Universitat de Barcelona i es va doctorar al Departament de Bioquímica i Biologia Molecular d'aquesta universitat l'any 1996. Ha desenvolupat la seva trajectòria investigadora al llarg dels últims vint anys en diferents entitats com ara els laboratoris Ferrer Internacional, la Universitat de Barcelona, la Universitat Autònoma de Barcelona i el Centre d'Investigació Cardiovascular (HSCSP-CSIC-ICCC). L'any 1998 s'incorporà com a investigador al Banc de Sang i Teixits, on, des de 2006, és el responsable del Laboratori de Coagulopaties Congènites. També, és investigador principal del Grup de Diagnòstic i Teràpia Molecular de l'Institut de Recerca Vall d'Hebron. La seva activitat professional

actual té un caràcter dual: per una banda, l'activitat assistencial donant suport al diagnòstic molecular, consell genètic i diagnòstic prenatal dels trastorns congènits de la coagulació i altres malalties hereditàries; i per l'altra, la investigació i el desenvolupament de noves aproximacions en el camp del diagnòstic i la terapèutica. Les seves línies principals d'investigació se centren en l'estudi de diversos aspectes bàsics i aplicats de malalties hereditàries de la sang amb elevada rellevància clínica, econòmica i social com ara l'hemofília, la mal-

tia de Von Willebrand i altres coagulopaties minoritàries. Les seves contribucions en l'àmbit de les coagulopaties congènites han merescut el reconeixement del Premi Internacional d'Investigació Duquessa de Sòria i el Premi de la Reial Fundació Victòria Eugènia, entre d'altres. Ha participat en nombrosos projectes nacionals i europeus com a col·laborador i com a investigador principal, la seva recerca ha estat finançada de manera continuada durant els darrers anys per entitats públiques i privades i és coautor en més de quaranta publicacions científiques.