

## TÈCNiques PER ESTUDIAR ELS GENS I LA SEVA EXPRESSIÓ

GLÒRIA RIBAS<sup>1</sup> I JOSEP CLOTET<sup>2</sup>

<sup>1</sup> *Grup de Genètica Humana, CNIO.*

<sup>2</sup> *Àrea de Biologia Molecular i Cel·lular, Facultat de Ciències de la Salut, Universitat Internacional de Catalunya.*

Adreça per a la correspondència: Josep Clotet. Àrea de Biologia Molecular i Cel·lular, Facultat de Ciències de la Salut, Universitat Internacional de Catalunya. Josep Trueta, s/n. 08195 Sant Cugat del Vallès. Adreça electrònica: [jclotet@csc.uic.es](mailto:jclotet@csc.uic.es).

### RESUM

La seqüenciació del DNA ens permet conèixer la composició química d'un determinat gen. Existeixen tècniques per comprovar si un gen s'expressa en una cèl·lula o teixit: són els protocols coneguts amb el nom de transferència *northern* (detecta la presència de RNA missatger) o la transferència *western* (que detecta la presència de proteïna). En els darrers deu anys han sorgit tot un seguit de tècniques que han permès passar de l'estudi individual dels gens a l'estudi global. Així, la genòmica estudia la seqüència de milers de gens alhora, la transcriptòmica comprova la presència de tots els missatgers presents en una cèl·lula en un moment determinat i la proteòmica fa el mateix, però analitzant el contingut global de proteïnes. En aquest capítol es presenten de manera resumida què són i què aporten cadascuna d'aquestes tècniques.

**Paraules clau:** genòmica, transcriptòmica, proteòmica.

### LABORATORY TECHNIQUES FOR THE STUDY OF GENES AND THEIR EXPRESSION

### SUMMARY

The DNA sequencing allows us to know the chemistry composition of a gene. Otherwise, it can be evaluated if a gene is expressed in a cell by the northern blot protocol (which checks for the presence of a messenger RNA) or by the western blot (which evaluate the levels of a protein). In the past 10 years, the development of the new molecular biology techniques allowed us to study the genes of a genome in a more global view. Then, the genomics study the sequence of a thousands of genes at the same time, the transcriptomics check for the presence of all the messengers in a cell, and the proteomics do the same but

analyzing the protein global content. In this chapter are presented the main concepts and the principal achievements of all these techniques.

**Key words:** genomics, transcriptomics, proteomics.

## INTRODUCCIÓ

Per a un biòleg que fa vint anys va finalitzar els seus estudis a la universitat, la *genòmica*, la *proteòmica*, els *microxips de DNA*, o els *xips de proteïnes* poden semblar paraules d'altres branques del coneixement, potser una fusió de la informàtica i la biologia. I el problema no és que els professors de la facultat deixessin d'explicar una determinada part del temari: és que en els darrers deu anys ha aparegut tota una metodologia d'anàlisi dels gens i de la seva expressió que ha revolucionat sens cap mena de dubte la biologia.

Per entrar pausadament en tots aquests conceptes nous, val a dir que no són més que aproximacions globals a l'estudi molecular de les cèl·lules. En el capítol anterior es feia referència a les tècniques que permeten l'estudi individual de gens. Altres eines que permeten conèixer gens de manera individualitzada són la seqüenciació del DNA i l'anàlisi de la seva expressió mitjançant l'estudi dels RNA missatgers i les proteïnes. En canvi, la genòmica, la transcriptòmica o la proteòmica permeten l'anàlisi de milers de gens, milers de missatgers o milers de proteïnes alhora. Per tant, més enllà de les paraules, les «òmiques» no són metodologies estranyes, sinó anàlisis massives dels components moleculars de la cèl·lula.

Per facilitar la comprensió d'aquestes tecnologies noves, en aquest capítol presentarem la tècnica que permet l'estudi individual d'un gen (per exemple la seqüenciació del DNA) i tot seguit la seva «òmica» corresponent (per exemple, els estudis genòmics). Ho farem seguint el fil argumental proposat en la presentació de diapositives que podeu trobar en el CD. En aquest capítol és

clau la visualització d'esquemes i dibuixos; per tant, és millor llegir-lo tenint al davant les imatges de la presentació.

## ES POT LLEGIR LA SEQÜÈNCIA D'UN GEN

L'estudi de les diferències entre seqüències dels àcids nucleics es pot realitzar mitjançant les tècniques de seqüenciació. Hi ha diferents estratègies per seqüenciar els gens, però la més implementada avui dia és la inventada el 1975 por Frederick Sanger, de la Universitat de Cambridge (Sanger *et al.*, 1977)

La tècnica de Sanger consisteix a sintetitzar fragments de DNA utilitzant com a motlle el DNA que es vol seqüenciar, mitjançant la reacció de PCR (vegeu el capítol anterior). La clau del mètode està en el fet que la síntesi de DNA es pot parar si en lloc d'un desoxiribonucleòtid (dNTP) s'afegeix un anàleg artificial del tipus didesoxiribonucleòtid (ddNTP) que, en no tenir el grup -OH en el carboni 3', no permet l'addició posterior d'un dNTP nou, i d'aquesta manera s'interromp l'elongació de la cadena de DNA. Aquests ddNTP són, per tant, nucleòtids terminadors de la reacció de síntesi del DNA.

Per dur a terme el procés de seqüenciació, es prepara un tub i s'afegeix una barreja dels quatre dNTP (dATP, dTTP, dGTP, dCTP) i una barreja dels quatre ddNTP. Aquests darrers s'afegeixen a una concentració molt menor per permetre que, durant la síntesi de les cadenes de DNA, s'incorpori la majoria de les vegades el dNTP i no el nucleòtid terminador. De totes maneres, de tant en tant s'incorporarà un terminador i

la síntesi s'aturarà, i es produiran fragments de DNA de grandàries diferents en funció del lloc on s'hagin incorporat. Un cop finalitzada la reacció, els productes obtinguts se separen per electroforesi en gels de poliacrilamida. Aquests gels tenen el mateix principi de separació que els gels d'agarosa (vegeu el capítol anterior), però a causa de la seva estructura altament ordenada tenen una grandària de porus més uniforme, fet que permet una resolució molt alta a l'hora de separar fragments de pesos similars; de fet, poden separar fragments de DNA de fins a un parell de bases de diferència. Tot i que els seqüenciadors en gel encara són molt utilitzats, els equips de seqüenciació més moderns utilitzen sistemes d'electroforesi capillar, que presenten molts avantatges, comparats amb una electroforesi tradicional en gel. Els capillars presenten una dissipació de calor molt més eficient, cosa que permet la utilització de voltatges més alts i, per tant, temps de carrera molt més curts; també, com que el sistema de detecció és més sensible, es requereix menys DNA per mostra que amb els gels tradicionals (Righetti *et al.*, 2002).

Els quatre nucleòtids terminadors estan marcats cadascun amb un fluorocrom diferent. Així, al final de l'electroforesi, no solament tenim separats els fragments de DNA en funció de la seva grandària, sinó que també es poden visualitzar les bandes de color diferent en funció del nucleòtid terminador incorporat. Per tant, al final del procés només cal llegir l'ordre d'aparició de bandes de colors diferents per saber l'ordre de la seqüència del DNA (vegeu-ne l'esquema a la presentació).

Potser l'aplicació més coneguda de les tècniques de seqüenciació és el Projecte Genoma Humà. La idea del projecte va sorgir el 1985, però el treball no es va iniciar formalment fins l'any 1990, i s'hi varen implicar centenars de laboratoris de tot el món. El febrer de 2001 es varen publicar conjunta-

ment dos estudis sobre la seqüenciació del genoma humà, l'un liderat per una empresa privada (Venter *et al.*, 2001) i l'altre per un grup de científics de la comunitat internacional (Lander *et al.*, 2001), aquest últim amb l'objectiu de fer públics i accessibles els resultats obtinguts. L'objectiu d'aquest projecte era catalogar i analitzar tots els gens que es troben en els cromosomes i que constitueixen el genoma de l'espècie *Homo sapiens*. En definitiva, es pretenia descriure la seqüència de tot el DNA present en el nucli d'una cèl·lula humana, és a dir, esbrinar l'ordre de les «lletres químiques» o *nuclèotids*. De totes maneres, la seqüenciació del DNA és la tasca més senzilla de tot el projecte. La interpretació de les dades és el pas més complicat: és resoldre un trencaclosques, i és la veritable missió del projecte.

Per desgràcia, la tècnica de seqüenciació de Sanger solament és capaç de generar seqüències d'uns sis-cents nucleòtids de mitjana d'un sol cop. Per això, per obtenir la seqüència del genoma humà cal descompondre el DNA en fragments de cinc-centes lletres, seqüenciar-los, i enganxar-los en l'ordre correcte. Per tant, cal imaginar el Projecte Genoma Humà com un trencaclosques gegant, format per sis milions de peces! (tres mil milions de nucleòtids dividits per cinc-cents nucleòtids obtinguts en cada seqüència), que cal anar empalmant laboriosament fins poder llegir la seqüència de cada cromosoma de manera contínua. El gran problema del projecte ha estat aconseguir ordinadors suficientment potents per poder processar tal quantitat d'informació i aconseguir empalmar tots els fragments.

Una altra de les dificultats que han de resoldre els ordinadors és que en aquesta enorme sopa de lletres cal cercar els elements que constitueixen els gens. El genoma humà solament conté entre trenta mil i quaranta mil gens, i la resta són fragments de DNA no codificants, en molts casos sense funció coneguda: introns, regions entre

gens, etc...). Del total de tres mil milions de nucleòtids, solament un 2 % porta el missatge dels gens. De nou, fan falta ordinadors molt potents capaços de processar una quantitat tan gran d'informació.

## ES POT LLEGIR LA SEQÜÈNCIA DE MILERS DE GENS D'UNA MATEIXA PERSONA ALHORA

Les tècniques de genòmica fan referència a un conjunt de protocols heterogenis que permeten analitzar milers de gens alhora sense haver d'utilitzar el laboriós mètode de seqüenciació. Aquestes tècniques no donen resultats detallats sobre la seqüència dels gens, però donen una visió global del genoma. D'aquesta manera es poden comparar genomes de la mateixa espècie, o fins i tot d'espècies diferents.

En el mateix número de la revista *Nature* en què es va publicar el primer esborrany de la seqüència del genoma humà (Lander *et al.*, 2001) es parla també dels SNP (o SNIP, de l'anglès *single nucleotide polymorphisms*). No era la primera vegada que es descriu: ja es coneixien com a RFLP, també de l'anglès (*restriction fragment length polymorphism*), i no són més que variacions d'una sola lletra en la seqüència referència del nostre genoma que s'observen quan comparem dos cromosomes entre si, sempre que estiguin presents en almenys un 1 % d'una població determinada. Cal tenir en compte que alguns d'aquests SNP els podem trobar en la part codificant dels gens, els anomenats cSNP, que poden canviar l'aminoàcid normal per un altre en la proteïna amb les implicacions funcionals que això suposa (Brooker, 1999; Miller i Kwok, 2001).

De totes maneres, cadascuna de les persones d'aquest planeta, amb l'excepció dels bessons monozigòtics, té un genoma únic, i encara que quan comparem dues persones, aquestes tindran un 99,9 % d'identitat entre

si, els separen milions de diferències. Un total d'1,42 milions de SNP eren detectats ja el 2001, quantitat que correspondria a una densitat d'un SNP per cada 1,91 kilobases (Lander *et al.*, 2001). Actualment aquestes xifres s'han incrementat fins a un total de deu milions de SNP o polimorfismes d'una lletra, i aproximadament un canvi cada tres-centes bases. Aquestes xifres s'han obtingut gràcies a l'esforç internacional cristallitzat en el projecte Hapmap (o mapatge d'haplotips —l'haplotip és la combinació de diversos SNP en un mateix cromosoma—) (Couzin, 2002). L'objectiu d'aquest projecte ha estat determinar el patrons comuns de variació de la seqüència de DNA en el genoma humà, mitjançant la caracterització de les variants SNP, les seves freqüències i la correlació entre aquestes, obtingudes amb l'estudi de mostres de DNA de diferents poblacions del món, com són Àfrica, Àsia i Europa. Una de les conseqüències d'aquest projecte ha estat el desenvolupament de tecnologia puntera que ha permès el genotipatge de totes aquestes variants en un temps rècord i amb una exactitud altíssima.

La disponibilitat d'aquest ampli catàleg de SNP obre la possibilitat del seu estudi com a marcadors genètics en l'estudi d'associació a malalties complexes (com poden ser la diabetis, el càncer, l'arteriosclerosi, etc.) en estudis d'identificació genètica (paternitat, procedència de mostres biològiques) i estudis evolutius.

Les tecnologies desenvolupades en la detecció dels SNP són moltes i variades. Totes intenten discriminar els dos al·lèls del polimorfisme i poder genotipar individus, és a dir, separar entre gent que porta les dues còpies «normals» o més abundants —els anomenem *homozigots normals* o *comuns*—, individus que són portadors de dues còpies de l'al·lel polimòrfic, mutat o rar, que són els *homozigots rars* o *polimòrfics* i, per acabar, els individus que en porten una còpia de ca-

da, és a dir, els *heterozigots*. Amb les anàlisis també podem calcular la proporció d'ambdós allels en la població: el més abundant s'anomena *normal*, *referència* o *major*; l'altre és el *minoritari*, *mutant* o *menor*.

A continuació, exposem una de les moltes possibilitats tecnològiques que actualment es tenen per detectar SNIP. Per entendre-la utilitzarem conceptes explicats en el capítol anterior. Es tracta de la tècnica de lligació i extensió, que és el mètode més específic i relativament fàcil d'optimitzar, però és el més lent i el que requereix el major nombre d'oligonucleòtids o sondes per portar-se a terme. Per facilitar-ne la comprensió, recomanem seguir les imatges que podeu trobar en el CD adjunt.

En aquesta tècnica, s'utilitza l'enzim DNA-ligasa, capaç de realitzar unions entre dues molècules de DNA. Quan dos oligonucleòtids adjacents són units al DNA diana, seran lligats només si aquests s'aparellen perfectament en el punt d'unió; és a dir, si l'aparellament no és perfecte, la DNA-ligasa no podrà unir els dos oligonucleòtids adjacents. La base del mètode està en el disseny d'aquests oligonucleòtids. El primer i el segon (P1 i P2 al dibuix) donen sensibilitat a la tècnica, ja que són els que indiquen si la hibridació amb el DNA diana és perfecta o si hi ha una base que balla. El tercer oligonucleòtid (P3) és més llarg i té tres regions amb seqüències diferents: la primera s'uneix al DNA diana de manera específica, just al davant dels primers oligonucleòtids (P1 i P2); la segona és una regió central, també amb una seqüència específica de nucleòtids, que després servirà per unir-se en un punt concret del xip de DNA; la tercera, anomenada *universal*, és comuna a tots els oligonucleòtids P3.

Un cop s'ha produït la lligació entre els dos nucleòtids es procedeix a l'amplificació per PCR dels productes resultants de la lligació. Per això s'utilitzen un altre cop els encebadors P1 i P2, però aquest cop marcats

amb un fluorocrom, i també s'utilitza l'encebador P3. Fixem-nos que en una reacció es posen tots els encebadors, perquè en realitat no sabem quin funcionarà. Si el pacient era homozigot per a l'allel normal llavors haurà lligat l'oligonucleòtid P1 i s'amplificarà amb l'encebador P1; en cas contrari, ho farà amb els encebadors P2, i si és heterozigot tindrà una barreja de productes amplificats amb els encebadors P1 i P2.

A continuació es procedeix a desnaturalitzar els productes amplificats i a la hibridació en el xip del DNA. Aquest xip és una peça metàl·lica hexagonal que conté una fibra òptica d'uns 0,5 cm de costat, on hi ha dipositades una gran quantitat de gotes de DNA: unes 30.000. Aquestes plataformes, anomenades en anglès *microarrays* ('arreglaments petits') són un prodigi de l'enginyeria que han permès la miniaturització dels processos d'anàlisi (Fan, 2006). En cada gota del xip de DNA podem trobar una petita quantitat de DNA amb la seqüència complementària de la regió específica central de l'oligonucleòtid P2. Precisament gràcies a aquesta seqüència específica, cada DNA amplificat s'unirà a un punt determinat del xip de DNA, cosa que després ens permetrà la seva localització. En realitat, els DNA s'hibriden en cinc o sis punts diferents de l'hexàgon, per tal de tenir mesures per quintuplicat i estar segurs de la consistència de l'observació. Això fa que en cada xip puguem analitzar mil cinc-cents variacions de nucleòtids per persona. Fixem-nos que des del punt de vista de la informació pràctica, si estem avaluant els SNIP de mil cinc-cents gens diferents, el resultat del xip equival a «seqüenciar» mil cinc-cents gens de cop!

Per acabar, tenim el procés de detecció de les fluorescències, que es podria fer a ull, mitjançant la utilització d'un microscopi de fluorescència i avaluant el resultat de cada gota. Però això seria molt costós i es podrien fer molts errors, ja que les gotes estan ex-

tremadament juntes dins el xip. Per evitar aquests problemes s'utilitzen uns lectors automàtics d'alta resolució que informen en una pantalla d'ordinador del resultat global del xip i els resultats de cada gota. El producte final és un informe detallat de la presència de mil cinc-cents SNIP d'una persona.

## ES POT LLEGIR LA SEQÜÈNCIA D'UN GEN DE MOLTES PERSONES ALHORA

Les tècniques de genòmica tenen moltes aplicacions. Per exemple, de vegades en la pràctica clínica ens trobem amb mutacions molt representades en certes malalties i, en aquests casos, el fet de trobar tals mutacions és un diagnòstic clar de la malaltia. Fins ara s'havia d'amplificar el gen i portarlo a seqüenciar, però si es volen fer cribratges ràpids a un cert nombre de pacients, de vegades s'utilitzen estratègies genòmiques.

En les presentacions del CD es pot trobar un esquema del protocol de determinació. De manera resumida direm que primer cal extreure el DNA dels pacients, després es col·loca en un microxip i després s'analitza amb una sonda de DNA fluorescent que té la mutació que es vol analitzar. Si els pacients presenten la mutació, la sonda s'unirà als DNA i donarà senyal fluorescent, i a l'inrevés si no la tenen.

## ES POT SABER SI UN GEN EXPRESSA EL SEU mRNA

Sovint, en molts projectes de recerca o en proves de diagnòstic clínic, es vol saber si en un gen s'està fabricant l'mRNA o la proteïna corresponent. Entre altres aplicacions, l'estudi de l'expressió dels gens pot ser útil per:

- Conèixer en quins teixits s'està expressant un determinat gen.

- Estudiar quins gens s'expressen quan s'administra un fàrmac.

- Determinar si una proteïna vírica o bacteriana està present en l'organisme.

Si es vol saber si un gen s'està expressant, generalment s'avalua la quantitat de proteïna que codifica. De totes maneres, en molts treballs d'investigació s'estudien els nivells de mRNA, ja que pot donar-se el cas que un gen es transcriu però no es tradueixi.

En valorar els nivells d'un mRNA concret d'una cèl·lula apareix un problema complicat: es vol buscar una agulla en un paller. La cèl·lula té molts mRNA i cal utilitzar una tècnica que permeti detectar-ne un de sol. Aquesta tècnica rep el nom de *transferència northern* (Alwine *et al.*, 1977). El peculiar nom anglès donat a aquesta tècnica no ilustra molt el seu funcionament, el qual resumim a continuació amb l'ajut dels esquemes presentats en el CD.

En primer lloc, es prenen unes quantes cèl·lules en un tub i es trenquen les membranes per alliberar-ne el contingut. A continuació, s'eliminen les proteïnes presents mitjançant l'addició de fenol; d'aquesta manera s'obté una solució més o menys pura de molècules de RNA. Aquest tipus de molècules tenen càrrega negativa, de manera que si s'apliquen directament a un gel d'electroforesi d'agarosa (vegeu el capítol anterior), en un parell d'hores s'hauran separat en funció del seu pes molecular.

Posteriorment es transfereixen els RNA separats des del gel d'agarosa a una membrana d'un derivat de paper (nitrocellulosa) per fer-los més accessibles a la sonda específica que s'afegeix a continuació. Aquesta sonda és la clau de la tècnica: es tracta d'una molècula de DNA complementària a l'RNA buscat. Aquesta sonda es troba marcada amb biotina unida a les citosines, les quals podran ser detectades posteriorment amb l'ajut d'un anticòs. Quan s'afegeix la sonda

a la membrana, lliscarà per sobre d'aquesta buscant una seqüència complementària, i si la troba s'hi enganxarà. Es renta bé la membrana i a continuació s'afegeix l'anticòs antibiòtic per detectar els punts d'unió de la sonda. Aquest anticòs porta unida una proteïna (com per exemple la fosfatasa alcalina) que hidrolitza un compost químic incolor i el torna acolorit. Al final el que tenim és l'aparició de bandes de color a sobre de la membrana que ens indicaran la presència de l'mRNA cercat. En lloc de biotina, la sonda de DNA pot estar marcada radioactivament, i aquesta és una altra estratègia per poder visualitzar-la.

### ES POT SABER QUINS GENS DEL GENOMA EXPRESSEN EL SEU RNA

Mentre que la transferència *northern* ens permet estudiar la presència d'un sol gen, la transcriptòmica estudia la presència de moltes molècules de mRNA, alhora que permet estudiar les variacions dels mRNA (o transcrits) entre teixits diferents o persones diferents. Les tècniques transcriptòmiques es basen en les tècniques genòmiques descrites anteriorment, i es presenten a continuació de manera resumida.

Tal i com es pot seguir en les il·lustracions, existeixen xips de DNA de la grandària d'un portaobjectes, on hi ha trenta mil microgotetes, perfectament cartografiades, cadascuna de les quals conté un gen diferent del genoma humà; podem dir, per tant, que aquest xip de DNA conté tot el genoma humà codificant. Si extraiem l'RNA d'un pacient, el marquem amb un fluorocrom i el llancem sobre el xip, els diferents RNA s'hibridaran amb els corresponents gens fixats al xip. Quan amb l'ajut d'un lector de fluorescència analitzem els resultats, podrem comprovar que hi ha gens que tenen fluorescència (i que per tant expressen l'mRNA) i gens que no. Fins i tot es pot quantificar una mica la

fluorescència i deduir el grau d'expressió d'un gen concret. D'aquesta manera obtenim el perfil d'expressió de tot el genoma! Si resulta que estàvem analitzant un teixit malalt, podem comparar el seu perfil d'expressió amb el perfil d'un teixit sa i observar-ne les diferències. No us heu preguntat mai quins gens estan desregulats en un determinat tipus de càncer? Potser aquests seran la causa del creixement desmesurat del teixit. Doncs aquestes anàlisis permeten respondre aquest tipus de preguntes.

Una millora d'aquesta tècnica també es troba representada de manera esquemàtica en el CD adjunt. Es tracta de la hibridació diferencial en la qual els dos teixits (el teixit problema i el control) es marquen amb fluorocroms diferents i s'hibriden al mateix xip. Si un gen del xip té color verd és que només s'expressa en el teixit 1, si té color vermell només s'expressa en el teixit 2 i si té color groc és que s'hi han unit RNA dels dos colors i, per tant, s'expressa en els dos teixits.

### ES POT SABER SI UN GEN EXPRESSA PROTEÏNES

Per saber si un gen s'està expressant, és a dir, si està fabricant la proteïna corresponent, es necessita disposar d'anticossos que reconeguin i s'uneixin específicament a la proteïna que volem. La seva obtenció és un procés rutinari que es pot veure esquematitzat en les diapositives. En primer lloc, es purifica la proteïna que es vol investigar i s'injecta a un animal d'una espècie diferent (normalment s'utilitzen conills). L'animal desenvolupa una reacció immunitària i fabrica anticossos contra la proteïna injectada. Després d'unes poques setmanes s'extreu sèrum sanguini de l'animal i es purifica, i d'aquesta manera s'obtenen grans quantitats d'anticossos contra la proteïna desitjada. Els anticossos poden distingir proteïnes molt similars i es poden utilitzar

per determinar la quantitat total d'una proteïna en qualsevol barreja.

L'immunoassaig és una tècnica que permet la detecció de determinades substàncies d'origen biològic presents en l'organisme encara que sigui en quantitats extremadament petites. Generalment s'utilitza per al diagnòstic de malalties: detectar cèl·lules tumorals, hormones, virus o bacteris. L'immunoassaig més famós és la prova de l'embaràs.

L'immunoassaig ELISA (de l'anglès *enzyme linked immunosorbent assay*, 'enzim-immunoassaig sobre fase sòlida') té lloc en unes plaques especials amb diversos pous per realitzar múltiples deteccions de manera simultània. En el fons dels pous es troba un polímer especial al qual s'unirà l'anticòs. Les etapes d'aquest assaig es poden seguir en l'esquema de la presentació animada.

De vegades, la molècula que es vol detectar en un pacient és un anticòs, com en el cas de les al·lèrgies o les infeccions víriques. Per exemple, en les anàlisis sanguínies rutinàries per detectar el VIH no es busca el virus directament, sinó que es pretén determinar si la persona analitzada ha fabricat anticòsos contra el virus. En aquest cas la molècula que s'utilitza com a «trampa» per caçar els anticòsos és un fragment d'una proteïna vírica. Aquest fragment es pot sintetitzar químicament i unir-lo al fons dels pous d'una placa d'ELISA, a la qual s'afegeix una mica de sèrum del pacient. Després de realitzar un rentatge, s'afegeix el segon anticòs marcat, que detectarà la presència d'anticòsos humans. De nou, la presència de color en la placa d'ELISA confirmarà el diagnòstic; en aquest cas ens dirà que el pacient és seropositiu, és a dir, que en la seva sang s'ha detectat la presència d'anticòsos contra el VIH.

Els ELISA generalment s'utilitzen per detectar proteïnes presents en mostres més o menys senzilles, com les mostres d'orina o sèrum sanguini. Però quan la mostra a ana-

litzar és molt més complexa, per exemple, quan existeix un gran nombre de proteïnes i en una gran concentració, s'utilitza una tècnica de immunoassaig més sensible i específica: la transferència *western*. Es tracta, doncs, un altre cop, d'una tècnica que permet trobar una agulla en un paller immens (Towbin *et al.*, 1979).

La similitud del nom amb el de la tècnica descrita en l'apartat anterior ja suggereix que les passes a seguir seran molt semblants. Bàsicament, es prenen unes quantes cèl·lules (procedents d'un cultiu de teixit, per exemple) en un tub i mitjançant un procediment físic (pressió, ultrasons, xoc osmòtic...) es trenquen les membranes per alliberar el contingut cellular, i se n'obté un homogeneïtat. A continuació, aquest homogeneïtat es barreja amb un detergent aniònic (normalment l'SDS, dodecil sulfat sòdic) que s'unirà a les proteïnes i les desnaturitzarà i carregarà negativament. Després d'uns minuts, la barreja es col·loca en un gel d'electroforesi d'acrilamida i s'hi aplica un camp elèctric. Les proteïnes avançaran cap al pol positiu, però les més grosses es retardaran en comparació de les petites. D'aquesta manera, les proteïnes es distribuïran al llarg del gel segons el seu pes molecular. Aquests tipus d'electroforesi rep el nom de SDS-PAGE (de l'anglès *sodium dodecyl sulfate polyacrylamide gel electrophoresis*, 'electroforesi en gels de poliacrilamida en presència de dodecil sulfat sòdic'), a causa dels dos principals components que porta: l'acrilamida i l'SDS.

Un cop separades les proteïnes, es transfereixen a una membrana (que en aquest cas sol ser de niló) per fer-les més accessibles a l'anticòs específic que s'afegeix a continuació. Un cop afegit, la membrana s'agita de manera que l'anticòs llisqui per sobre seu, i si troba la proteïna corresponent s'hi unirà. El complex proteïna-anticòs format sobre la membrana pot detectar-se mitjançant l'addició d'un segon anticòs que



reconeixerà la presència dels anticossos primaris (normalment, aquest segon és un anticòs fet en cabra que reconeix anticossos de conill en general). Per poder visualitzar totes aquestes interaccions, el segon anticòs porta unit un enzim que produeix un compost acolorit com en el cas de la tècnica d'ELISA.

## ES POT SABER QUINS GENS DEL GENOMA EXPRESSEN PROTEÏNES

La proteòmica es pot definir com l'estudi del proteoma o conjunt de proteïnes expressades en la cèl·lula en un moment donat. És una àrea de recerca amb un creixement fort en l'era postgenòmica, i té com a objectiu fonamental la identificació dels nivells d'expressió de proteïnes, l'estudi d'interaccions entre proteïnes i la determinació de possibles modificacions posttraduccional.

L'anàlisi proteòmica permet als investigadors estudiar el contingut proteínic global d'un compartiment subcel·lular, cel·lular, teixit o organisme en un moment determinat i sota unes certes condicions ambientals. És una aproximació molt adient per comparar diferents estats d'un sistema biològic; per exemple, sa *versus* malalt, cèl·lules control *versus* estressades, teixit alimentat *versus* teixit en dejú, etc.

Les estratègies de la proteòmica es poden agrupar en dos grans passos: la separació de les barreges complexes de proteïnes i la identificació posterior de cada una. El conjunt de proteïnes es pot separar fent servir gels de SDS-PAGE, l'electroforesi 2D (electroforesi en dues dimensions) o la cromatografia líquida multidimensional. La identificació de proteïnes es fa bàsicament amb dues tècniques complementàries: l'aplicació de l'espectrometria de masses per a la identificació de proteïnes i la seqüenciació automàtica de l'extrem N-terminal de proteïnes mitjançant degradació d'Edman. A

continuació detallarem algunes d'aquestes tècniques.

L'electroforesi 2D (dues dimensions) és una tècnica que permet analitzar mesclades complexes de proteïnes procedents de mostres biològiques. D'entrada, cal un tractament de la mostra per obtenir-ne exclusivament la fracció proteínica i eliminar possibles interferències, com són les sals, entre d'altres. Inicialment, les proteïnes se separen en una primera dimensió segons el punt isoelèctric (pI) mitjançant isoelectroenfocament en gels que tenen un gradient de pH. En la segona dimensió, les proteïnes se separen d'acord amb la massa molecular en un gel estàndard SDS-PAGE (vegeu l'apartat anterior). A continuació, es duu a terme la tinció dels gels (amb tincions específiques de proteïnes, com el Comassie o la tinció de plata), s'adquireixen les imatges dels gels amb un escàner d'alta resolució i s'analitzen amb un programari especialitzat. Així es poden comparar gels i observar canvis en el patró de taques: variació d'intensitats, aparició o desaparició de taques, desplaçaments d'algunes bandes, etc.

Si es vol saber a quina proteïna correspon una banda concreta del 2D, es pot aplicar l'espectrometria de masses. Per dur a terme aquest procés és retalla la banda, es neteja de contaminants i es digereix enzimàticament amb una proteasa (per exemple, la trombina). Un cop digerida, els pèptids resultants s'ionitzen amb un làser i es fan viatjar a través d'un tub atrets per un ànode que al final té un detector; aquest sistema d'anàlisi es coneix amb el nom de MALDI (de l'anglès *matrix assisted laser desorption-ionization*, 'alliberament-ionització facilitada per làser'). El temps de vol que tarda el pèptid a viatjar fins al detector (conegut com a TOF, de l'anglès *time of flight*, 'temps de vol') és dependent de la massa molecular del pèptid: els més grans arribaran més tard i els més lleugers arribaran abans. Al final del procés una pantalla d'ordinador

ens representarà els espectres de MALDI-TOF, els quals indiquen el nombre i la grandària dels fragments generats a partir de la digestió d'una proteïna purificada.

Cal tenir en compte que les proteases utilitzades per a la degradació de les proteïnes tallen seqüències específiques d'aminoàcids i, per tant, podem predir el pes molecular dels fragments resultants: és el que es coneix com a *empremta dactilar de les proteïnes*. Podríem fer el símil que són com els enzims de restricció del DNA, que donen fragments del pes esperat si es coneix la seqüència del gen. Per tant, el pas final de les tècniques de proteòmica és comparar els cromatogrames del MALDI-TOF amb els cromatogrames predits per ordinador, que es troben en diferents bases de dades electròniques. Al final l'ordinador ens dirà a quina proteïna pertanyen els diferents pèptids detectats pel MALDI.

Si la determinació de l'empremta peptídica no és suficient per identificar una proteïna determinada, es poden obtenir petites seqüències dels extrems N-terminals dels pèptids mitjançant la degradació d'Edman. Aquestes seqüències parcials obtingudes d'un o diversos pèptids solen ser suficients per identificar una proteïna mitjançant la recerca de les seqüències obtingudes en bases de dades informàtiques.

## CONCLUSIÓ

Les tècniques d'anàlisi d'un gen particular han donat pas a estudis globals, i en aquests moments disposem de perfils d'expressió de gens en condicions molt variades. Sabem com responen els genomes en presència de sal, davant variacions de temperatura, de pH, d'infeccions, d'agents tumorals, de situacions d'excés o manca de nutrients, etc. Fixem-nos que la informació que s'obté és radicalment diferent a la de la genètica clàssica: aquí la pregunta no és

«s'expressa un determinat gen en presència de sal?», sinó «què fa el genoma sencer davant un canvi en la pressió osmòtica?». Així doncs, amb un sol experiment sabem les quantitats d'expressió de tots els gens d'un determinat genoma.

En qualsevol cas, la informació que s'extreu d'aquests tipus d'escomeses és limitada i cal que s'interpreti amb molta cura. En primer lloc, a causa de les limitacions de la tècnica, ja que la miniaturització dels processos sovint porta a errors difícils de detectar. Per minimitzar aquests errors, recentment les revistes científiques més importants demanen que els experiments de perfils d'expressió (els experiments de transcriptòmica, per exemple) es presentin per triplicat i que s'eliminin els resultats que no coincideixin en els tres experiments. També, les variacions subtils, aquelles que potser tenen una importància biològica clau, sovint no poden ser detectades mitjançant les escomeses experimentals de les «òmiques», i cal fer experiments individuals de transferències *northern* i *western* per a cadascun dels gens que es volen analitzar.

Amb el temps s'anirà millorant la sensibilitat i la reproductibilitat dels resultats. Mentrestant, la quantitat ingent d'informació que s'està obtenint en els laboratoris no s'analitza en detall (resultaria impossible) però s'emmagatzema en servidors especials, on passen a formar part d'enormes bases de dades que estan al servei de la comunitat científica. D'aquesta manera, quan algú està interessat en l'estudi d'un gen concret, el primer que fa és mirar si ha donat algun resultat interessant en experiments de transcriptòmica o proteòmica en altres laboratoris. En definitiva, estem en un moment en què es generen grans quantitats d'informació, fet que sens dubte produirà una explosió del coneixement molecular que tenim de la vida.

## BIBLIOGRAFIA

- ALWINE, J. C.; KEMP, D. J.; STARK, G. R. (1977). «Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes». *Proc. Natl. Acad. Sci. USA*, 74: 5350-5354.
- BROOKES, A. J. (1999). «The essence of SNP». *Gene*, 234: 177-186.
- COUZIN, J. (2002). «Human genome. HapMap launched with pledges of \$100 million». *Science*, 298: 941-942.
- COX, T. M.; SINCLAIR, J. (1998). *Biología molecular en medicina*. Madrid: Editorial Médica Panamericana.
- FAN, J.; GUNDERSON, K. L.; BIBIKOVA, M.; YEAKLEY, J. M.; CHEN, J.; WICKHAM GARCIA, E.; LEBRUSKA, L. L.; LAURENT, M.; SHEN, R.; BARKER, D. (2006). «Illumina universal bead arrays». A: KIMMEL, A. [ed.]. *DNA Microarrays*. Nova York: Academic Press (col. *Methods in Enzymology*; 410), 57-73. [Tot el volum fa referència als diferents tipus de microxips]
- LANDER, E. S. [et al.] (2001). «Initial sequencing and analysis of the human genome». *Nature*, 409: 860-921.
- MILLER, R. D.; KWOK, P. Y. (2001). «The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine». *Hum. Mol. Genet.*, 10: 2195-2198.
- RIGHETTI, P. G.; GELFI, C.; D'ACUNTO, M. R. (2002). «Recent progress in DNA analysis by capillary electrophoresis». *Electrophoresis*, 23: 1361-1374.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. (1977). «DNA sequencing with chain-terminating inhibitors». *Proc. Natl. Acad. Sci. USA*, 74: 5463-5467.
- TOWBIN, H.; STAEBELIN, T.; GORDON, J. (1979). «Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications». *Proc. Natl. Acad. Sci. USA*, 76: 4350-4354.
- VENTER, J. C. [et al.] (2001). «The sequence of the human genome». *Science*, 291: 1304-1351.