

GENÒMICA FUNCIONAL: L'EXEMPLE DEL LLEVAT

JOSEP VILARDELL

Departament de Biologia Cel·lular. Albert Einstein College of Medicine. EUA.

Adreça per a la correspondència: Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Av., The Bronx, NY10461, Estats Units. Adreça electrònica: vilardel@aecom.yu.edu

INTRODUCCIÓ

Fa més de cent anys, des que es van començar a definir les bases de la biologia, que els científics fan servir els anomenats *organismes model* per a estudiar els mecanismes de fenòmens biològics fonamentals, com ara el desenvolupament o l'herència genètica, mecanismes que, com s'ha anat veient amb els anys, han estat força conservats durant l'evolució. Les raons per les quals un determinat organisme esdevé un model són diverses. Avantatges de tipus pràctic, com ara duració del cicle biològic, facilitat de cultiu, o possibilitat d'interpretació dels resultats obtinguts, han estat sens dubte les més importants; tot i que factors comercials o de proximitat evolutiva a la nostra espècie han prevalgut en altres casos. Amb la revolució científica que ha estat el desenvolupament de la biologia molecular, els investigadors han seguit emprant organismes model per a estudiar els mecanismes fonamentals de l'expressió gènica. Molts dels gens estu-

diats tenen homòlegs en humans, i possiblement tinguin funcions similars. Fa uns deu anys, en un nou episodi d'aquesta revolució, es van iniciar projectes de seqüenciació de genomes sencers, alguns ja complets en l'actualitat. Una nova branca de la biologia, la *genòmica*, de l'anglès *Genomics*, es dedica exclusivament a l'estudi funcional i estructural dels genomes. (El terme *genome* va ésser emprat per primer cop el 1920 per H. Winkler, i prové de la fusió de les paraules *gene* i *chromosome*; *Genomics* va ser proposat el 1986 per T. R. Roderick [McKusick, 1997]). Si es té en compte, per exemple, que hi ha al voltant de quatre mil malalties degudes a disfuncions d'un gen en particular (Gelehrter *et al.* 1998) i que molts dels problemes de salut actuals tenen un fort component poligènic (hipertensió, càncer, obesitat, desequilibris psíquics, etc.) la importància de l'estudi de la funció del genoma esdevé òbvia.

Com s'ha dit, el fet que tots els organismes tinguin una història evolutiva compar-

tida és el motiu clau que fa útil l'estudi dels organismes model. D'altra banda, disposar de més d'un model és necessari ja que, generalment, les aproximacions experimentals i la informació obtinguda són diferents i, per tant, complementàries. Afortunadament, aquest fet es va reconèixer en l'estratègia dissenyada per a seqüenciar el genoma humà i es van incloure com a fites importants l'obtenció de les seqüències de diversos organismes model, com ara (entre altres) *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana* i *Mus musculus*.

EL PROJECTE GENOMA DE LLEVAT

El projecte per a seqüenciar el genoma de llevat (*Saccharomyces cerevisiae*) va ésser iniciat per investigadors europeus el 1989. La idea inicial es basava en la coordinació dels laboratoris europeus amb potencial per a dur a terme el projecte, més que en la creació de nous i costosos centres (Dujon, 1996). L'èxit inicial, amb el treball sobre el cromosoma III, va fer que el projecte s'expandís ràpidament. La seqüència sencera del genoma de llevat, de 12.052 kb (el genoma humà, en comparació, té 3.000.000 kb), es va dipositar als bancs de dades internacionals l'abril de 1996 (<http://genome-www.stanford.edu/Saccharomyces>) i va resultar de l'esforç coordinat de més de sis-cents científics, distribuïts en més de cent laboratoris arreu del món. Europa va contribuir amb un 72 % de la seqüència (Sanger Centre, Anglaterra, 17 %), els EUA amb un 22 % (Washington University, 15 %; Stanford University, 7 %), el Canadà (McGill University) amb un 4 % i el Japó amb un 2 % (RIKEN University). L'error és molt baix, al voltant d'una base per cada 10.000 (Hieter *et al.*, 1996). El genoma de llevat és força compacte, amb una mitjana d'un gen cada 2 kb, i un 70 % del DNA codifi-

cant proteïnes. Deixant de banda els gens del rRNA (1-2 Mb), la resta del genoma conté uns sis mil gens, dels quals uns 2.300 (al voltant d'un 40 % del total) havien estat identificats prèviament pels mètodes genètics tradicionals. De la resta de les 3.700 proteïnes (normalment es parla de pautes de lectura obertes o ORF, *open reading frames*) predites, 2.300 no tenien cap similitud amb proteïnes identificades prèviament en altres organismes (tot i que aquestes ORF òrfenes poden contenir dominis típics, codificant, per exemple, regions transmembrana, de localització subcel·lular, etc., informació a tenir en compte en futures anàlisis). Aquest darrer descobriment va generar la recerca de mètodes sistemàtics per a determinar la possible funció d'aquests nous gens (Oliver, 1996b). La llargada mitjana de les ORF és de 1.450 pb, o 483 codons. L'ORF més llarga, de funció desconeguda, conté 4.910 codons, tot i que la gran majoria d'ORF no sobrepassen els 1.500 codons. Es fa difícil predir quin és el gen més petit, però, com a exemple, el gen *RPL41* té una ORF de 26 codons, que codifica la proteïna ribosomal L41, de 24 aminoàcids. (Més informació sobre ORF es pot trobar a <http://www.mips.biochem.mpg.de> i a <http://www.proteome.com>). El genoma de llevat conté regions amb duplicacions i altres ordenacions dels gens (Mewes, 1997; Eisen *et al.*, 1998; <http://rana.stanford.edu/clustering>) que són força interessants des d'un punt de vista evolutiu, si més no, i que apunten a l'atractiva possibilitat d'una duplicació ancestral de tot el genoma del llevat (Wolfe i Shields, 1997).

EL GENOMA DE LLEVAT: RELLEVÀNCIA

Una qüestió a què és difícil escapar és la rellevància que té el conèixer la seqüència completa del genoma de *Saccharomyces cere-*

visiae, sobretot per a l'anàlisi dels organismes multicel·lulars. Tot i que els beneficiats més immediats han estat els investigadors de llevat, aquesta informació ha estat igualment valuosa per a la recerca en organismes multicel·lulars, i ja ha accelerat el ritme dels descobriments biològics en general. Per exemple, la seqüència d'una nova proteïna d'un organisme multicel·lular eucariota pot emprar-se per fer una cerca de la seva contrapartida (o l'homòleg més proper en la família) en llevat. El fet de tenir tota la seqüència genòmica del darrer permet que la cerca es faci sense ambigüitats, ja que es trobaran totes les seqüències relacionades, i es pot seleccionar la d'homologia més gran. Quan la proteïna original prové d'organismes evolutivament allunyats del llevat (com acostuma a ésser el cas) el fet de disposar de tot el genoma és vital, ja que homologies subtils esdevenen molt més convincents. A més, sabent que la cerca és entre totes les proteïnes que especifiquen una cèl·lula eucariota, les dades que es puguin obtenir de l'homòleg de llevat poden aportar informació molt valuosa per a entendre un determinat fenotip o una determinada funció gènica en l'organisme original. Addicionalment, el fet de disposar d'un homòleg en llevat proveeix un excepcional paradigma experimental per a continuar les anàlisis.

El llevat és un organisme model amb un gran potencial, demostrat innombrables vegades, per a conèixer en profunditat la funció de gens que intervenen en els mecanismes biològics intracel·lulars. També ha estat d'utilitat en el desxiframent de mecanismes tan allunyats com són, per exemple, la transmissió de senyal dels receptors nerviosos (Koelle i Horvitz, 1996). La recerca en llevat, sobretot en biologia molecular, combina la genètica clàssica (aïllament i anàlisi fenotípica de mutants) amb la bioquímica i tècniques de DNA recombinant, manipulant segments de DNA clonats i reintroduint-los

als llocs homòlegs dins el genoma de llevat (la freqüència de recombinació homòloga en llevat es prou elevada per a fer aquests tipus d'experiments de manera més senzilla i ràpida que en la gran majoria d'altres sistemes model; Rothstein, 1983). Aquests mètodes han permès l'associació directa entre mutacions definides i fenotips concrets, tant *in vivo* com *in vitro*.

El progrés recent en el camp de les malalties genètiques ha servit també per a evidenciar l'impacte que té en aquesta àrea el disposar de la seqüència del genoma de llevat. S'ha pogut avançar molt en la comprensió de la funció d'alguns gens humans relacionats amb malalties hereditàries gràcies a llur homologia amb gens de llevat de funció coneguda. Aquesta situació permet també la utilització del llevat com a model per a l'estudi de la funció dels gens involucrats. En alguns casos afortunats, el gen humà, expressat en llevat, complementa certes mutacions en el corresponent gen ortòleg de llevat, permetent un estudi de tipus estructural i funcional exhaustiu. El benefici és prou important com per a dissenyar soques de llevat especials per a poder expressar amb èxit gens d'altres organismes. Per exemple, en alguns casos ha estat necessària la substitució de certs components d'estructures multiproteiques de llevat perquè tinguin lloc les interaccions necessàries amb el nou gen exogen, amb el qual han coevolucionat (Hieter *et al.*, 1996). D'altra banda, a causa del major nombre de gens humans (uns 80.000 mentre que són 6.000 en el llevat, tot i que el genoma humà es dos-cents cops més gran), hi ha la possibilitat que certs gens no tinguin homòlegs en llevat. Tanmateix, en aquests casos el gen pot estar relacionat amb un gen de llevat pel fet de ser ambdós membres de la mateixa família de proteïnes, amb funcions anàlogues però no idèntiques i pot ésser possible el modificar cert procés cel·lular per fer-lo més adient.

Aquesta tècnica ja fa temps que s'empra amb èxit. Per exemple, reemplaçant els gens de llevat que intervien en els primers passos de la resposta a feromones per certs gens humans, es va aconseguir que el llevat respongués a una molècula específica pròpia d'humans (King *et al.*, 1990). Aquest tipus de dissenys experimentals proporcionen sistemes per a les anàlisis de les interaccions receptor-ligand, o per a *screenings* especials (on es crivella un nombre molt gran de possibles candidats) emprats en la recerca de nous fàrmacs (Broach i Thorner, 1996), aspecte que està essent aprofitat per moltes companyies farmacèutiques.

Encara no és possible determinar quants gens humans tindran homòlegs en llevat, però els casos identificats ja representen un grup significatiu. Per exemple, els gens responsables de l'atàxia-telangièctasi (ATM [Morrow *et al.*, 1995]) i la síndrome de Bloom (BLM [Ellis *et al.*, 1995]) tenen una similitud de seqüència significativa amb els gens *MEC1* i *SGS1* de llevat, respectivament. Al cromosoma XI hi ha un gen homòleg del responsable, en humans, de l'adreno-leucodistrofia, mentre que un altre comparteix una sèrie de dominis amb el gen responsable de la fibrosi quística (Dujon *et al.*, 1994). La proliferació de seqüències provinents d'organismes model ha estat útil per al mapatge de gens humans responsables de malalties. El projecte XREF (Bassett *et al.*, 1995; <http://www.ncbi.nlm.nih.com/XREFdb>) es dedica al mapatge sistemàtic d'EST en humans (fragments de seqüència expressada o *expressed sequence tags* [Adams *et al.*, 1991] que ja representen més d'un 50 % dels gens humans) que presenten algunes similituds significatives amb seqüències d'organismes model. D'aquesta manera es relacionen funcions de proteïnes d'organismes model amb posicions del mapa genètic humà. Això permet aplicar tècniques basades en el clonatge posicional per a tractar

d'identificar gens responsables de malalties que estiguin localitzats en la mateixa regió del genoma. Un exemple d'aquesta estratègia va ésser el clonatge del gen humà *hMLH1*, responsable de certes formes hereditàries de càncer de còlon i que és homòleg del gen de llevat *MLH1* (Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994). Una altra tècnica per a la cerca d'homòlegs en llevat de gens humans relacionats amb malalties es coneix com *sondeig per homologia* (*homology probing* [Yahraus *et al.*, 1995; Braverman *et al.*, 1997]). Les eines inicials són també el genoma de l'organisme model (llevat en aquest cas) i el banc d'EST humanes. El procediment comença amb la identificació d'un procés o via bioquímica defectuosa que causa la malaltia. Seguidament tots els gens que participen en la via són identificats en llevat, el que permet fer una cerca al banc d'EST per a possibles homòlegs, els quals permetran identificar gens humans candidats com a responsables de la malaltia. Seguint aquest esquema s'han pogut clonar els gens humans involucrats en la biosíntesi del peroxisoma i responsables dels defectes hereditaris lligats a aquest procés, com ara la síndrome de Zellweger.

Finalment, el llevat també pot emprar-se com una eina de *screening* per a descobrir nous gens en altres organismes, ja que alguns dels mutants es poden complementar amb llibreries de cDNA heteròlogues. Adicionalment, el fet de disposar de tota la seqüència del genoma de llevat permet crear, per genètica reversa, molts més mutants en aquells gens que fins ara havien escapat de les tècniques tradicionals de genètica.

Genòmica funcional al llevat

Els diferents projectes de seqüenciació de genomes estan prou avançats (o fins i tot acabats, com el de llevat [<http://genome-www.stanford.edu/Saccharomyces>] o de

Caenorhabditis elegans [The *C. elegans* Sequencing Consortium, 1998; http://genome.wustl.edu/gsc/C_elegans i http://www.sanger.ac.uk/Projects/C_elegans]) com perquè la genòmica, estigui en forta expansió i es comenci a parlar de diferents subdisciplines (Hieter i Boguski, 1997; McKusick, 1997). La genòmica estructural s'ocupa de la fase inicial de l'anàlisi d'un genoma, amb el mapatge, seqüenciació, etc., fins a obtenir la seqüència sencera del DNA que forma el genoma de l'organisme (també hi ha qui assigna el terme *genòmica estructural* a l'estudi de l'estructura de les proteïnes codificades per un genoma sencer [Sung-Hou, 1998]). La genòmica comparativa es centra en l'anàlisi de la distribució de la informació en els genomes i en la comparació d'aquesta distribució entre genomes de diferents organismes. Per exemple, se sap que certes regions cromosòmiques tenen una distribució de gens conservada entre espècies relativament allunyades, com ara ratolins i humans, i això pot ésser de força utilitat a l'hora de mapar gens responsables de malalties. A vegades, la genòmica comparativa és bàsicament una primera anàlisi de les dades que es van recopilant durant les diferents fases de la genòmica estructural, que pot començar, tot i que no sempre és així (Beardsley, 1998), amb un mapatge exhaustiu del genoma.

La genòmica funcional representa una nova fase en l'anàlisi dels genomes, amb un gran potencial per a la producció d'un gran volum d'informació, de molta novetat, que està revolucionant la recerca en sistemes biològics. Al mateix temps, però, s'està requerint un gran esforç intel·lectual de cara a fer un ús efectiu de la gran quantitat de dades aportades per la genòmica estructural. En concret, la genòmica funcional es centra en el desenvolupament i aplicació de tècniques experimentals per a estudiar i identificar la funció dels genomes, fent ús de la in-

formació que prové de la genòmica estructural. En altres paraules, la genòmica funcional estudia els patrons d'activitat del genoma. L'estratègia fonamental és expandir l'abast de la recerca en biologia, passant de l'estudi de gens i proteïnes individuals, a un estudi global de grans col·leccions de gens al mateix temps, de manera sistemàtica i en el context del genoma on es troben. Les aproximacions experimentals acostumen a generar quantitats ingents de dades, de manera que calen anàlisis estadístiques i computacionals, que poden exigir més feina que no pas l'experiment en sí, però que estan produint tota una sèrie de descobriments importants i a vegades inesperats. La biologia computacional té doncs també un nou paper. Fins ara, en la genòmica estructural, es feia necessari l'ús d'ordinadors per al maneig de grans quantitats de dades; ara cal anar un pas més enllà i extreure informació que pot estar soterrada o camuflada. Els primers resultats semblen confirmar l'esperança que la genòmica funcional faci més petita la separació entre seqüència i funció, com també que faci possible l'estudi del comportament dels sistemes biològics.

El llevat va força per davant en el camp de la genòmica funcional, i és, de fet, el model que millor està mostrant el seu potencial. Fins i tot abans que s'hagués completat la seqüència del seu genoma ja es va crear un sistema per a la detecció i anàlisi de gens presents en seqüències de DNA clonades a l'atzar. Aquest sistema es basa en l'estudi, per electroforesi de proteïnes en dues dimensions, del patró de proteïnes expressades en soques de llevat transformades amb plasmidis multicòpia que contenen el DNA que s'investiga (Thoraval *et al.*, 1990). Un dels primers projectes que es va dur a terme un cop desxifrat tot el genoma de llevat va ésser la creació d'un banc de soques, cadascuna amb un gen delecionat (Oliver, 1996a). Un altre ha estat el de substituir cada ORF

per oligonucleòtids específics, que poden detectar-se per PCR o DNA *microarray* (explicat més endavant) en experiments encaminats a detectar soques que sobreviuen en certes condicions (Shoemaker *et al.*, 1996). Aquests bancs estan disponibles per a tots els investigadors que ho desitgin. Quant a l'anàlisi de les poblacions de mRNA s'han ideat tres mètodes per a recopilar dades: els xips d'oligonucleòtids (Fodor *et al.*, 1993; Hacia *et al.*, 1998), els DNA *microarrays* (Schena *et al.*, 1995, 1998) i SAGE (*Serial Analysis of Gene Expression*, o anàlisi en sèrie de l'expressió gènica [Velculescu *et al.*, 1995, 1997]).

Els xips d'oligonucleòtids es basen en la combinació de tecnologies per a semiconductors (fotolitografia) i de síntesi química de DNA, fent possible la síntesi de milers d'oligonucleòtids (*oligos*) de manera ordenada i controlada, damunt d'un suport, generalment vidre. Els principals avantatges rau en el control precís de la seqüència dels oligos que es fan servir i en la possibilitat de fer la síntesi directament sobre el suport, evitant així els problemes que comporta la preparació i manipulació de mostres. Un altre avantatge que cal tenir en compte es l'estandardització, amb un mínim de variació entre xips, gràcies a l'optimització de cada cicle en la síntesi dels oligos. El principal desavantatge és l'alt cost dels xips ja que la síntesi basada en la fotolitografia requereix la fabricació d'una sèrie de platines especials que permeten especificar les seqüències dels oligos del xip (Fodor *et al.*, 1993). Un cop sintetitzat el xip (els científics que van desenvolupar la tecnologia — Fodor i col·legues — han creat la companyia que gairebé té el monopoli de la seva producció: Affymetrix, a Santa Clara, Califòrnia), aquest s'hibrida amb una població de cDNA que s'ha marcat convenientment, generalment amb un marcador fluorescent, i el resultat de la hibridació s'analitza amb

un equipament força sofisticat, basat en un làser i un lector que escobren totes les posicions del xip. L'estudi correcte de les dades obtingudes necessita també d'un programari adient. Emprant aquesta tecnologia s'han dut a terme estudis, entre altres, de l'expressió de tot el genoma de llevat sota diferents condicions. Darrerament s'ha realitzat un experiment molt interessant amb la finalitat d'identificar la relació de dependència entre els gens i els components essencials de la maquinària transcripcional del llevat (Holstege *et al.*, 1998 i <http://www.wi.mit.edu/young/expression.html>), i s'ha trobat que hi ha grups de gens amb diferent resposta a cada un dels elements de la maquinària basal necessària per a l'inici de la transcripció. Aquests resultats revelen un important i inesperat nivell de regulació transcripcional, en addició als sistemes de regulació basats en factors de transcripció específics de gens o grups de gens. A més, es demostra que algunes vies de transducció de senyal poden actuar directament sobre maquinària d'iniciació de la transcripció. Malgrat l'enorme potencial d'aquesta tècnica, el cost la manté encara fora de l'abast de molts laboratoris de recerca (de fet, molts dels que la fan servir ho fan a través de col·laboracions amb les companyies que manufacturen els xips). Tanmateix, també té utilitats de diagnòstic, com ara la cerca de certes mutacions o patrons d'expressió típics de certes malalties. Actualment els xips d'Affymetrix contenen uns 400.000 oligos en una àrea d'uns 1,6 cm² (Schena *et al.*, 1998).

La filosofia dels *microarrays* de DNA és molt semblant a la dels xips d'oligos. La principal diferència és que els fragments de DNA (generalment productes de PCR) no es sintetitzen sobre el suport, sinó que es sintetitza primer cada fragment per separat per després fixar-lo en una posició determinada al suport. El cost és molt inferior al

dels xips d'oligos, però l'estandardització és molt més difícil i la densitat de seqüències per unitat de superfície pot ésser de deu a cent cops menor (Schena *et al.*, 1998). L'estratègia més comuna, si més no en el cas del llevat (que, com s'ha dit, fins al desembre del 1998 era l'únic eucariota totalment seqüenciat) consisteix en fer una PCR de cada gen que s'estudia (unes sis mil si es fa una anàlisi del genoma sencer) fent servir encebadors (*primers*) comercials i un procés automatitzat, i després fixar el producte en un suport de vidre, generalment un porta-objectes de microscòpia òptica. En un *array* de 18x18 mm es pot tenir representat tot el genoma de llevat (DeRisi *et al.*, 1997). Com en els xips d'oligos, l'*array* s'hibrida a una mostra de cDNA marcada convenientment i el grau d'hibridació es mesura emprant procediments sofisticats de densitometria i anàlisi d'imatges per ordinador, anàlegs als emprats en el cas dels xips d'oligos. Emprant marcatges diferencials de cDNA sintetitzats a partir de RNA aïllats en diferents condicions (per exemple, el pas de metabolisme fermentatiu al respiratori), es pot detectar quins gens s'indueixen i/o reprimeixen. Com amb els xips d'oligos, es pot estudiar el patró d'expressió global de tot el genoma sencer, si més no per estimar l'acumulació de mRNA. A part de la resposta genòmica als canvis metabòlics (DeRisi *et al.*, 1997), s'han dut a terme també estudis molt interessants del comportament del genoma durant la mitosi, l'esporeulació, els canvis de temperatura i la transició de metabolisme fermentatiu a respiratori. Aquest tipus d'experiment, impensable fa pocs anys, ha permès, a més, d'organitzar el genoma en grups d'expressió gènica (*cluster analysis*). Els resultats d'aquest estudi, realitzat per Eisen *et al.*, (1998) són també accessibles a <http://rana.stanford.edu/clustering/serum.html>, on la magnitud de la informació disponible per als investigadors serveix per

a evidenciar, entre altres coses, la revolució científica que estem vivint.

A diferència dels sistemes anteriors, basats en hibridacions de sondes a una alta densitat de molècules fixades sobre un suport sòlid, el SAGE analitza quantitativament una determinada població de transcrits (Velculescu *et al.*, 1997). La finalitat del SAGE és una quantificació de l'expressió genòmica, referida al nombre de missatgers de cada gen en cada cèl·lula, definint el que els autors anomenen el *transcriptoma* (de l'anglès *transcriptome*), en contraposició a *genoma*. (També existeix el *proteome* [<http://www.proteome.com>], una base de dades privada, que proveeix informació sobre les proteïnes d'organismes ben caracteritzats genèticament, com és el cas del llevat). El transcriptoma d'un genoma estaria definit pels gens transcrits sota certes condicions que s'estudien. Per desxifrar el contingut del transcriptoma, es sintetitzen seqüències curtes (9-11 pb) de cDNA, corresponent cadascuna a un transcrit. Aquestes seqüències *etiqueta* (*tags*), es concatenen posteriorment per a ésser seqüenciades. De la seqüència es dedueix l'abundància de cada *tag* individual, de la qual es pot calcular la freqüència de cada mRNA. En concret, els pioners de la tècnica, Velculescu i col·laboradors (1997), van analitzar el transcriptoma de llevat sota diferents períodes del cicle cel·lular. Per a això van aïllar les poblacions de mRNA corresponents i, després de sintetitzar cDNA marcat amb biotina en un extrem, van purificar fragments curts (els *tags*) corresponents a la zona 3' de dels missatgers, mitjançant l'ús d'enzims de restricció, i els van concatenar i seqüenciar. Els períodes estudiats van ésser la fase exponencial de creixement, cèl·lules parades en la fase S i cèl·lules parades a la fase G2/M. Es va descobrir que en cada fase, al voltant d'un 75 % de tots els gens (coneguts i predits) s'està expressant, mentre que el nombre de còpies per cèl·lula oscil·la entre

menys de 0,3 (el límit de l'assaig) i dues-centes còpies. Curiosament, dins del grup dels gens altament expressats (mes de seixanta còpies de missatge per cèl·lula), hi ha tres ORF que corresponen a gens desconeguts (tot i que una d'elles s'havia postulat, durant l'anàlisi genòmica, que codifica una proteïna ribosomal desconeguda). Es va observar que la majoria de gens estan representats per una o dues còpies per cèl·lula. Això implica que alguns gens constitutius es transcriurien menys d'un cop per generació, si es té en compte la vida mitjana dels seus mRNA. Una altra observació va ser que menys d'un 1 % dels components de cada transcriptoma és afectat extraordinàriament per un canvi en les condicions. Aquests resultats van obrir una nova perspectiva en l'estudi del llevat. La possibilitat de definir transcriptomes en llevat va donar peu a intentar el mateix en humans, on un experiment similar, basat en el banc d'EST, s'ha dut a terme per a l'estudi de transcriptomes relacionats amb càncer (Zhang *et al.*, 1997).

Fins ara s'han descrit mètodes per a l'anàlisi de la funció genòmica analitzant un producte de la seva expressió, l'RNA missatge. Tanmateix, el llevat ofereix la possibilitat de l'estudi de les interaccions del següent producte, les proteïnes. Aquesta anàlisi es basa en l'assaig dels *dos híbrids* (traduït també com *dobles híbrids*, però generalment conegut pel seu nom anglès, *two hybrid*; Bartel i Fields, 1995). S'ha iniciat el projecte d'una anàlisi exhaustiva dels divuit milions de combinacions possibles (Hudson *et al.*, 1997), i en aquest aspecte val a dir que les noves tècniques de microseqüenciació de proteïnes, de gels de dues dimensions d'alta resolució i d'espectroscòpia de masses (espectroscòpia *nano-electrospray*, Wilm *et al.*, 1996; McCormack *et al.*, 1997) estan sent emprades en l'assignació de pèptids a gens específics, sense ambigüitats, en el context de tota la seqüència del genoma. Una

qüestió interessant que es planteja es la relació entre els grups de proteïnes que possiblement interaccionen entre elles i els *clusters* d'expressió gènica que s'han observat en els estudis descrits anteriorment.

CONCLUSIÓ

La genòmica funcional està beneficiant de dues maneres les aproximacions tradicionals als problemes biològics. Primer, les fa més productives i eficients. Segon, les complementa, donant una visió globalitzadora dels sistemes biològics com les entitats que són. Tot i que s'ha afegit una nova dimensió a la biologia, el coneixement detallat dels mecanismes biològics continuarà necessitant de la genètica, bioquímica, biologia cel·lular i estudis estructurals clàssics. S'ha afegit una nova dimensió a la biologia, però no se n'ha eliminat cap. En general, les dades proporcionades per la genòmica han de proporcionar el marc de treball per a continuar amb anàlisis més particulars.

Nogensmenys, s'està vivint una revolució científica. La genòmica funcional, amb la seqüència dels genomes de molts organismes (incloent-hi l'humà) està canviant la manera de fer recerca en biologia. Moltes disciplines comencen a replantejar-se llurs estratègies experimentals per a adaptar-se a la nova escala genòmica. Noves tècniques s'han posat a punt o s'estan desenvolupant ràpidament. Es tracta de tècniques potents i que treballen amb un volum de dades que tot just comencem a assimilar. Cal adaptar el rigor del mètode científic per a poder avaluar, de manera fiable, la ingent quantitat d'informació que s'està generant. No és estrany trobar que el format tradicional de l'article científic ja no és suficient per a descriure els resultats amb exactitud, i és possible que l'article vagi acompanyat de referències a bases de dades internacionals

o a pàgines web on els laboratoris dipositen les dades (Dolinsky *et al.*, 1998). De fet, el problema de l'accés a les dades generades no està resolt de manera clara (Eisenberg, 1996). La genòmica estructural ha creat la tradició de fer públics i assequibles a la comunitat científica les noves dades que van apareixent, tradició que s'espera que es mantingui, tot i que pot fer-se necessari algun tipus d'estandardització (Bloom, 1998). Un efecte, relativament inesperat, de la capacitat de les noves tècniques, és l'afavoriment de les col·laboracions entre científics de diferents disciplines o que treballen en diferents sistemes, ja que molts cops és l'única manera d'abastar el significat de les dades produïdes.

El llevat, en si mateix i com a model, està oferint les primeres imatges de cap a on pot dirigir-se el futur. Tanmateix, un dels millors llegats pot ésser el reconeixement del mèrit de fer servir diferents organismes per a resoldre problemes complexos.

AGRAÏMENTS

Voldria agrair a Jonathan R. Warner (Albert Einstein College of Medicine) els suggeriments i les respostes a qüestions que han anat apareixent durant l'escriptura d'aquesta revisió; i a Regina Raz (NYU School of Medicine) els suggeriments, la lectura crítica i l'edició del manuscrit.

BIBLIOGRAFIA

- ADAMS, M. D.; J. M. KELLEY; J. D. GOCAYNE; M. DUBNICK; M. H. POLYMERPOULOS; H. XIAO; C. R. MERRIL; A. WU; B. OLDE; R. F. MORENO (1991). «Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project». *Science*, núm. 252, pàg. 1651-1656.
- BARTEL, P. L.; S. FIELDS (1995). «Analyzing Protein-protein Interactions Using Two-hybrid System». *Methods in Enzymology*, núm. 254, pàg. 241-263.
- BASSETT, D. E. JR.; M. S. BOGUSKI; F. SPENCER; R. REEVES; M. GOEBL; P. HIETER (1995). «Comparative Genomics, Genome Cross-referencing and XREFdb». *Trends in Genetics*, núm. 11, pàg. 372-373.
- BEARDSLEY, T. (1998). «An Express Route to the Genome?». *Scientific American*, núm. 279, pàg. 30-32.
- BLOOM, F. E. (1998). «Staying Afloat on the Seas of Data». *Science*, núm. 282, pàg. 1989.
- BRAVERMAN, N.; G. STEEL; C. OBIE; A. MOSER; H. MOSER; S. J. GOULD; D. VALLE (1997). «Human PEX7 Encodes the Peroxisomal PTS2 Receptor and is Responsible for Rhizomelic Chondrodysplasia Punctata». *Nature Genetics*, núm. 15, pàg. 369-376.
- BROACH, J.; J. THORNER (1996). «High-throughput Screening for Drug Discovery». *Nature*, núm. 384 (supplement 6604), pàg. 14-16.
- BRONNER, C. E.; S. M. BAKER; P. T. MORRISON; G. WARREN; L. G. SMITH; M. K. LESCOE; M. KANE; C. EARABINO; J. LIPFORD; A. LINDBLOM (1994). «Mutation in the DNA Mismatch Repair gene Homologue *hMLH1* is Associated with Hereditary Non-polyposis Colon Cancer». *Nature*, núm. 368, pàg. 258-261.
- DE RISI, J. L.; V. R. IYER; P. O. BROWN (1997). «Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale». *Science*, núm. 278, pàg. 680-686.
- DOLINSKI, K.; C. A. BALL; S. A. CHERVITZ; S. S. DWIGHT; M. A. HARRIS; S. ROBERTS; T. ROE; J. M. CHERRY; D. BOTSTEIN (1988). «Expanding Yeast Knowledge Online». *Yeast*, núm. 14, pàg. 1453-1469.
- DUJON, B.; D. ALEXANDRAKI; B. ANDRE; W. ANSORGE; V. BALADRON; J. P. BALLESTA; A. BANREVI; P. A. BOLLE; M. BOLOTIN-FUKUHARA; P. BOSSIER (1994). «Complete DNA Sequence of Yeast Chromosome XI». *Nature*, núm. 369, pàg. 371-378.
- DUJON, B. (1996). «The Yeast Genome Project: What Did We Learn?». *Trends in Genetics*, núm. 12, pàg. 263-270.
- EISEN, M. B.; P. T. SPELLMAN; P. O. BROWN; D. BOTSTEIN (1998). «Cluster Analysis and Display of Genome-Wide Expression Patterns». *Proceedings of the National Academy of Sciences, USA*, núm. 95, pàg. 14863-14868.
- EISENBERG, R. S. (1996). «Intellectual Property Issues in Genomics». *Trends in Biotechnology*, núm. 14, pàg. 302-307.
- ELLIS, N. A.; J. GRODEN; T. Z. YE; J. STRAUGHEN; D. J. LENNON; S. CIOCCI; M. PROYTCHIEVA; J. GERMAN (1995). «The Bloom's Syndrome Gene Product is Homologous to RecQ Helicases». *Cell*, núm. 836, pàg. 55-66.
- FODOR, S. P.; R. P. RAVA; X. C. HUANG; A. C. PEASE; C. P. HOLMES; C. L. ADAMS (1993). «Multiplexed Biochemical Assays with Biological Chips». *Nature*, núm. 364, pàg. 555-556.
- GELEHRTER, T. D.; F. S. COLLINS; D. GINSBURG (1998). *Principles of Medical Genetics*, 2a ed. Baltimore: Williams and Wilkins, pàg. 4.
- HACIA, J. G.; W. MAKALOWSKI; K. EDGEMON; M. R. ERDOS; C. M. ROBBINS; S. P. FODOR; L. C. BRODY; F. S. COLLINS

- (1998). «Evolutionary Sequence Comparisons Using High-density Oligonucleotide Arrays». *Nature Genetics*, núm. 18, pàg. 155-158.
- HIETER, P.; D. E. JR. BASSETT; D. VALLE (1996). «The Yeast Genome a Common Currency». *Nature Genetics*, núm. 13, pàg. 253-255.
- HIETER, P.; M. BOGUSKI (1997). «Functional Genomics: It's All How You Read It». *Science*, núm. 278, pàg. 601-602.
- HOLSTEGE, F. C.; E. G. JENNINGS; J. J. WYRICK; T. I. LEE; C. J. HENGARTNER; M. R. GREEN; T. R. GOLUB; E. S. LANDER; R. A. YOUNG (1988). «Dissecting the Regulatory Circuitry of a Eukaryotic Genome». *Cell*, núm. 95, pàg. 717-728 [Vegeu també <<http://www.wi.mit.edu/young/expression.html>>].
- HUDSON, J. R. JR.; E. P. DAWSON; K. L. RUSHING; C. H. JACKSON; D. LOCKSHON; D. CONOVER; C. LANCAULT; J. R. HARRIS; S. J. SIMMONS; R. ROTHSTEIN; S. FIELDS (1997). «The Complete Set of Predicted Genes from *Saccharomyces cerevisiae* in a Readily Usable Form». *Genome Research*, núm. 7, pàg. 1169-1173.
- KIM, S. H. (1998). «Shining a Light on Structural Genomics». *Nature Structural Biology Synchrotron Supplement*, (Agost), pàg. 643-645.
- KING, K.; H. G. DOHLMAN; J. THORNER; M. G. CARON; R. J. LEFKOWITZ (1990). «Control of Yeast Mating Signal Transduction by a Mammalian Beta 2-adrenergic Receptor and Gs Alpha subunit». *Science*, núm. 250, pàg. 121-123 [fe d'errates publicada al volum 251, pàg. 144].
- KOELLE, M. R.; H. R. HORVITZ (1996). «EGL-10 Regulates G Protein Signaling in the *C. elegans* Nervous system and Shares a Conserved Domain with Many Mammalian Proteins». *Cell*, núm. 84, pàg. 115-125.
- MCCORMACK, A. L.; D. M. SCHIELTZ; B. GOODE; S. YANG; G. BARNES; D. DRUBIN; J. R. 3rd YATES (1997). «Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-femtomole Level». *Analytical Chemistry*, núm. 69, pàg. 767-776.
- McKUSICK, V. A. (1997). «Genomics: Structural and Functional Studies of Genomes». *Genomics*, núm. 45, pàg. 244-249.
- MEWES, H. W.; K. ALBERMANN; M. BAHR; D. FRISHMAN; A. GLEISSNER; J. HANI; K. HEUMANN; K. KLEINE; A. MAIERL; S. G. OLIVER; F. PEIFFER; A. ZOLLNER (1997). «Overview of the Yeast Genome». *Nature*, núm. 387 (suplement 6632), pàg. 7-65 [fe d'errates publicada al volum 387 (suplement 6634), pàg. 737].
- MORROW, D. M.; D. A. TAGLE; Y. SHILOH; F. S. COLLINS; P. HIETER (1995). «TEL1, an *S. cerevisiae* Homolog of the Human Gene Mutated in Ataxia Telangiectasia, is Functionally Related to the Yeast Checkpoint Gene MEC1». *Cell*, núm. 82, pàg. 831-840.
- OLIVER, S. (1996a). «A Network Approach to the Systematic Analysis of Yeast Gene Function». *Trends in Genetics*, núm. 12, pàg. 241-242.
- (1996b). «From DNA Sequence to Biological Function». *Nature*, núm. 379, pàg. 597-600.
- PAPADOPOULOS, N.; N. C. NICOLAIDES; Y. F. WEI; S. M. RUBEN; K. C. CARTER; C. A. ROSEN; W. A. HASELTINE; R. D. FLEISCHMANN; C. M. FRASER; M. D. ADAMS (1994). «Mutation of a *MutL* Homolog in Hereditary Colon Cancer». *Science*, núm. 263, pàg. 1625-1629.
- ROTHSTEIN, R. J. (1983). «One-step Gene Disruption in Yeast». *Methods in Enzymology*, núm. 101, pàg. 202-211.
- SCHENA, M.; D. SHALON; R. W. DAVIS; P. O. BROWN (1995). «Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray». *Science*, núm. 270, pàg. 467-470.
- SCHENA, M.; R. A. HELLER; T. P. THERIAULT; K. KONRAD; E. LACHENMEIER; R. W. DAVIS (1998). «Microarrays: Biotechnology's Discovery Platform for Functional Genomics». *Trends In Biotechnology*, núm. 16, pàg. 301-306.
- SHOEMAKER, D. D.; D. A. LASHKARI; D. MORRIS; M. MITTMANN; R. DAVIS (1996). «Quantitative Phenotypic Analysis of Yeast Deletion Mutants Using a Highly Parallel Molecular Bar-coding Strategy». *Nature Genetics*, núm. 14, pàg. 450-456.
- THE *C. ELEGANS* SEQUENCING CONSORTIUM (1998). (Vegeu <http://genome.wustl.edu/gsc/C_elegans> i <http://www.sanger.ac.uk/Projects/C_elegans> per a una llista sencera dels autors). *Science*, núm. 282, pàg. 2012-2018.
- THORAVAL, D.; M. RÉGNACQ; P. NEUVILLE; H. BOUCHERIE (1990). «Functional Analysis of the Yeast Genome: Use of Two-dimensional Gel Electrophoresis to Detect Genes in Randomly Cloned DNA Sequences». *Current Genetics*, núm. 18, pàg. 281-286.
- VELCULESCU, V. E.; L. ZHANG; B. VOGELSTEIN; K. W. KINZLER (1995). «Serial Analysis of Gene Expression». *Science*, núm. 270, pàg. 484-487.
- VELCULESCU, V. E.; L. ZHANG; W. ZHOU; J. VOGELSTEIN; M. A. BASRAI; D. E. JR. BASSETT; P. HIETER; B. VOGELSTEIN; K. W. KINZLER (1997). «Characterization of the Yeast Transcriptome». *Cell*, núm. 88, pàg. 243-251.
- WILM, M.; A. SHEVCHENKO; T. HOUTHAEVE; S. BREIT; L. SCHWEIGERER; T. FOTSIS; M. MANN (1996). *Nature*, núm. 379, pàg. 466-469.
- WOLFE, K. H.; D. C. SHIELDS (1997). «Molecular Evidence for an Ancient Duplication of the Entire Yeast Genome». *Nature*, núm. 387, pàg. 708-713.
- YAHRAUS, T.; N. BRAVERMAN; G. DODT; J. E. KALISH; J. C. MORRELL; H. W. MOSER; D. VALLE; S. J. GOULD (1995). «The Peroxisome Biogenesis Disorder Group 4 Gene, PXAAA1, Encodes a Cytoplasmic ATPase Required for Stability of the PTS1 Receptor». *EMBO J.*, núm. 15, pàg. 2914-2923.
- ZHANG, L.; W. ZHOU; V. E. VELCULESCU; S. E. KERN; R. H. HRUBAN; S. R. HAMILTON; B. VOGELSTEIN; K. W. KINZLER (1997). *Science*, núm. 276, pàg. 1268-1272.