



par Bertrand JORDAN

médecine/sciences 1998 ; 14 : 1097-102

## Voyage au pays des puces

Les « puces à ADN » continuent à faire l'objet d'une attention soutenue de la part des industriels et des médias. Chacun vante son approche : *microarrays* sur lame de verre pour Synteni, *Molecular Dynamics* et bien d'autres, centaines de milliers, bientôt millions d'oligonucléotides que la firme *Affymetrix* fait pousser *in situ* par des techniques proches de celles employées pour la fabrication des microprocesseurs, sans oublier diverses versions des bonnes vieilles membranes Nylon qui ont sans doute encore un bel avenir devant elles à en juger par le nombre de firmes qui les commercialisent sous des formes variées. Dans ce secteur où l'industrie est présente en force, la communication commerciale prend parfois le pas sur l'information scientifique et, malgré le paragraphe légal de rigueur sur les *forward-looking statements*<sup>1</sup>, il est souvent difficile de savoir où en sont effectivement ces différentes approches. Je vais faire le point, en me limitant à la mesure du niveau d'expression de gènes et en m'appuyant sur les informations glanées lors d'un récent colloque<sup>2</sup>. Pourquoi donc s'intéresser autant à ces mesures d'expression ? Deux rai-

sons s'imposent. Savoir à quel niveau s'exprime un gène dans différents tissus, différentes situations physiologiques ou pathologiques, c'est indubitablement faire un pas vers la compréhension de sa fonction. Information limitée, l'expression ne dit pas tout, et ses indications sont parfois ambiguës ou trompeuses : nous connaissons tous des gènes à expression ubiquitaire, et qui pourtant jouent un rôle crucial dans un seul tissu<sup>3</sup>. Elle n'en reste pas moins précieuse. La deuxième raison est technique : ces données sont à l'heure actuelle les seules que nous soyons capables d'obtenir pour de grands ensembles de gènes, les seules qui puissent être rassemblées à l'échelle des dizaines de milliers d'entités révélées par le séquençage de banques d'ADNc ou de génomes entiers. Après le génome, le transcriptome<sup>4</sup> !

Le monde industriel a accumulé dans les coffres-forts virtuels de ses *proprietary databases* des millions de sé-

quences partielles d'ADNc, déchiffrées par des entreprises comme *Incyte* ou *Human Genome Sciences*, qui s'ajoutent au million d'EST humains disponibles dans dbEST [1] ; il a compris (et quelques *success-stories* l'ont démontré aux incrédules) que la valeur de cette information pouvait être décuplée par la connaissance des patrons d'expression. L'étude conjointe des séquences, avec toutes les ressources d'une bio-informatique de pointe, et des tissus dans lesquels s'expriment les gènes correspondants, ainsi que la comparaison de ces profils entre tissu normal et pathologique, aboutissent à la définition de nouvelles protéines thérapeutiques dont certaines sont déjà en phase d'essai clinique [2].

## Recherche cibles désespérément...

Plus généralement, ces travaux révèlent de nouvelles « cibles » (*targets*), protéines jouant un rôle-clé dans divers processus physiologiques ou pathologiques (inflammation, coagulation, auto-immunité...). Le but de l'industrie pharmaceutique consiste alors à synthétiser des molécules pouvant agir sur ces entités, et ainsi moduler le processus en cause ; elle dispose pour cela des ressources de la chimie combinatoire, capable de produire en peu de temps des centaines de milliers de composés, et de techniques de criblage sophistiquées (*high-throughput screening assays*) qui peuvent en quelques semaines déterminer quelles molécules interagissent avec la cible et sont donc suscep-

<sup>3</sup> *Songeurs par exemple à l'enzyme adénosine désaminase, exprimée dans la plupart des tissus mais critique pour le bon fonctionnement des lymphocytes et dont l'absence est responsable d'un déficit immunitaire sévère sans autre symptôme majeur.*

<sup>4</sup> *Terme au sens assez flou, qui désigne la connaissance du patron d'expression de l'ensemble des gènes d'un organisme. Mais tout dépend de ce que l'on entend par patron d'expression : suffit-il de quelques tissus de base, ou faut-il avoir étudié les deux cents types cellulaires répertoriés dans l'organisme ? Et quid des variations d'expression dans les situations pathologiques ? Le transcriptome est donc une notion assez mal définie.*

<sup>1</sup> *Ce paragraphe se cache à la fin des communiqués de presse triomphalistes et en module considérablement le sens – en résumé, il explique que ce qui est annoncé peut constituer une extrapolation et ne doit pas être pris à la lettre !*

<sup>2</sup> *Congrès IBC « Biochip Technologies », San Francisco (22 au 24 juin 1998).*

tibles d'avoir une action thérapeutique. Notons à cet égard que le nombre total de cibles contre lesquelles sont dirigés les médicaments existants est évalué à environ quatre cents seulement; les analystes du domaine estiment que la Génomique en a déjà révélé un nombre équivalent, et que leur nombre total sera de plusieurs milliers d'ici deux ou trois années. On comprend que l'industrie s'intéresse de très près à l'analyse à grande échelle des niveaux d'expression qui, avec la connaissance des séquences et la bio-informatique, joue un rôle central dans la mise en évidence de ces entités.

Sur le plan technique, on assiste à la domination des méthodes fondées sur l'hybridation de sondes complexes avec de grands jeux de segments d'ADN fixés sur un support. Le *differential display* [3] et ses variantes ont, malgré leur difficulté expérimentale, leurs artefacts et leurs faux positifs, rendu de grands services dans les laboratoires. Ces systèmes, bien adaptés à la recherche d'un ou deux gènes fortement modulés lors du passage d'une situation à une autre, sont à ce titre pertinents pour maint projet de recherche académique; mais leur débit est trop faible, leur fiabilité trop aléatoire pour des études à grande échelle. L'ingénieuse technique *SAGE* [4], réduisant chaque gène exprimé à une information minimum, une *tag* d'une quinzaine de nucléotides, a tenté de nombreuses équipes; les difficultés de mise en œuvre ont pour le moment empêché sa généralisation<sup>1</sup>. Son impact dans le domaine reste aujourd'hui marginal.

### Les techniques en lice

*Macroarrays*, *microarrays* et «puces à ADN»: toutes ces méthodes reposent sur l'emploi d'une sonde complexe préparée à partir de l'ensemble des ARN messagers d'un tissu et hybridée avec un jeu ordonné (*array*) de milliers de cibles représentant chacune

un gène différent. La grande force de cette approche – comme le souligne depuis bientôt dix ans Hans Lehrach (Berlin, Allemagne) [5] – est leur fondamental parallélisme (au sens informatique du terme). Chaque hybridation donne en effet un renseignement sur chaque gène représenté dans le jeu, même s'il se résume à «non exprimé à un niveau détectable dans ce tissu». Ces informations sont cumulatives, puisque chaque niveau d'hybridation mesuré est relié à une cible, une séquence, un gène précis. De plus, rien ne s'oppose à ce que les collections de gènes mis en jeu s'approchent du nombre total de gènes humains – certaines des expériences actuelles n'en sont pas loin.

#### • Les *macroarrays*

Voyons donc où en sont ces techniques. La plus ancienne (et qui n'a sans doute pas dit son dernier mot) est celle des «membranes à haute densité», qui tendent aujourd'hui à prendre le nom de *macroarrays* par référence à leur version miniaturisée, le *microarray*. Elle utilise des clones d'ADNc ou leurs produits de PCR disposés régulièrement sur des membranes de Nylon, hybridés avec des sondes complexes radioactives produites par rétrotranscription d'une préparation d'ARN messenger ou total. L'hybridation, conduite en conditions d'excès de cible par rapport à la sonde, aboutit à la fixation sur chaque segment d'ADN d'une quantité de l'espèce correspondante de la sonde proportionnelle à son abondance dans le mélange d'ARN de départ; l'ensemble des données est acquis de manière quantitative, généralement par un système à écran phosphore, et mesure ainsi le niveau d'expression de chacun des gènes représentés – plus précisément l'abondance relative de chacun des ARN messagers dans le mélange de départ.

Des robots aptes à la confection de *macroarrays* sont disponibles dans le commerce, les systèmes de détection sont en place dans de nombreux laboratoires, et cette méthode est à la portée de beaucoup d'équipes, d'autant que plusieurs fabricants commercialisent des membranes prêtes à l'emploi portant quelques centaines

ou quelques milliers de gènes [6]. Sous sa forme la plus répandue, cette technique utilise des membranes d'assez grandes dimensions, avec une densité de quelques dizaines de dépôts par cm<sup>2</sup> et un espacement entre les *spots* supérieur à un millimètre, compatible avec une détection par des sondes marquées au <sup>32</sup>P ou, mieux, au <sup>33</sup>P. Une évolution vers la miniaturisation est en cours, avec un espacement réduit à deux ou trois cents microns et une détection colorimétrique [7] ou radioactive (avec des systèmes à haute résolution). L'emploi de sondes fluorescentes est interdit par l'autofluorescence des supports existants. Cette méthode a pour elle sa flexibilité, l'excellente sensibilité absolue de la détection radioactive et sa large gamme dynamique, ainsi que la capacité élevée des membranes Nylon qui permet la fixation de fortes quantités d'ADN cible et abaisse ainsi le seuil de détection. Elle est employée par de nombreuses équipes «académiques» [8-11] et par plusieurs industriels, notamment *Hyseq* [12].

#### • Les *microarrays*

Les méthodes plus récentes sont celles des *microarrays* et des puces à oligonucléotides ou *oligo-chips* (les puces à ADN proprement dites). Un *microarray* comporte quelques milliers de gènes représentés par des produits de PCR qu'a déposés un robot *ad hoc* sur une lame de verre traitée; l'hybridation est effectuée avec une sonde complexe obtenue par rétrotranscription du mélange d'ARN en présence de nucléotides substitués autorisant une détection ultérieure (directe ou non) par fluorescence. L'acquisition des résultats est généralement effectuée par un système de balayage laser équipé d'une optique confocale, les signaux étant mesurés par un photomultiplicateur. Ce système a été mis au point, dans le secteur académique, par l'équipe de Pat Brown à Stanford (CA, USA) qui a publié plusieurs articles très convaincants [13-16] et donne sur son site [17] des renseignements détaillés – y compris un manuel permettant en principe de construire son propre système! L'approche a également été employée par plusieurs groupes in-

<sup>1</sup> Dans des mains expertes, elle donne néanmoins de bons résultats; une entreprise, Genzyme ([www.genzyme.com/sage](http://www.genzyme.com/sage)) se consacre à sa diffusion et propose aux industriels la réalisation de «profils digitaux d'expression» à façon.

dustriels, notamment *Synteni* (entreprise fondée par un collaborateur de Pat Brown), et un système complet devrait bientôt être commercialisé par l'industriel *Molecular Dynamics* à un prix de l'ordre de 1,5 MF. Les *microarrays* ont l'avantage de la compacité: on peut envisager des jeux comportant cinquante ou cent mille cibles. De plus, le double ou éventuellement le triple marquage sont possibles, ce qui permet des comparaisons directes entre sondes différentes.

#### • Les *oligo-chips*

Quant aux *oligo-chips*, ils ont été développés au départ dans une autre perspective, celle de l'encore mythique « séquençage par hybridation » [18-20]. Des milliers d'oligonucléotides de séquence connue sont greffés sur une petite surface de verre ou de silicium; la lecture de la puce, après hybridation avec un ADN marqué par fluorescence, donne des informations sur sa séquence. Les *chips* vendus par *Affymetrix* [21] sont destinés à des applications de recherche de mutations (VIH, p53, cytochrome C). Il s'agit là d'un « quasi-séquençage » dans lequel la puce porte un jeu d'oligonucléotides représentant la séquence « officielle » du gène considéré: l'hybridation avec une sonde provenant d'un prélèvement biologique révèle alors les éventuelles différences entre la séquence présente chez le patient et la séquence habituelle. Mais *Affymetrix* s'intéresse beaucoup à la mesure de niveaux d'expression, et a produit dans ce but des *expression chips* dans lesquels chaque gène est représenté par un jeu d'oligonucléotides répartis le long de sa séquence codante; la viabilité de cette approche a été démontrée [22, 23]. Bien que cette façon d'évaluer des niveaux d'expression puisse paraître inutilement compliquée, elle bénéficie de l'impressionnant potentiel de miniaturisation de la technique: *Affymetrix* fabrique actuellement des puces d'un peu plus d'un centimètre de côté portant 320 000 oligonucléotides différents, susceptibles de mesurer le niveau d'expression de milliers de gènes, et affirme pouvoir passer bientôt à des puces portant des millions

d'éléments. Comme les méthodes de fabrication des *oligo-chips* suivent de près celles de l'industrie des microprocesseurs dont on connaît les prodiges récents en termes de miniaturisation, ces affirmations sont à prendre au sérieux.

#### Problèmes et limites de ces méthodes

Quelques précisions sont néanmoins nécessaires pour apprécier correctement l'état de l'art. Mentionnons d'abord un problème commun aux deux premières méthodes, celui de l'identification des cibles ou, plus précisément, de leur vérification. *Macroarrays* et *microarrays* se fondent en général, pour constituer leurs jeux de gènes, sur les séquences complètes et partielles déjà publiées, et représentent chaque gène par un clone d'ADNc (en fait, son produit de PCR) issu de la collection rassemblée par le consortium IMAGE [1], dans laquelle chaque clone est rattaché à au moins une séquence dans dbEST. Malheureusement, cet assortiment est entaché de multiples erreurs et 10 % à 20 % des clones IMAGE ne correspondent pas à leur séquence<sup>1</sup>... Pour bien faire, chaque clone destiné à un *array* doit donc être séquencé à nouveau, ce qui alourdit considérablement les projets. Les *oligo chips*, eux, échappent à cet écueil; mais ils ne peuvent naturellement inclure que des entités déjà séquencées, au contraire des micro- et macro-*arrays*.

<sup>1</sup> C'est-à-dire que le clone IMAGE obtenu d'un distributeur n'est pas celui qui a été séquencé pour dbEST: erreur d'identification de clone ou de rattachement de fichier qui souligne la difficulté à gérer rigoureusement des collections aussi étendues (un million et demi de clones actuellement, toutes espèces confondues...). Cela met sans doute aussi en évidence un préjugé nord-américain lié à la stratégie des STS, selon lequel seule l'information de séquence est essentielle, l'entité « clone » restant secondaire. Lors du lancement du programme d'EST pour la souris (qui a aujourd'hui produit plus de 300 000 séquences partielles), il n'était même pas prévu de conserver les clones et c'est seulement devant le tollé des utilisateurs que cette phase fut incluse – sans doute à contrecœur, et peut-être pas avec tout le soin nécessaire. Notons aussi l'annonce récente de la contamination de nombreux clones IMAGE par un bactériophage...

Un deuxième point est celui de la sensibilité et de la masse d'échantillon nécessaire pour préparer une sonde complexe. Ces deux paramètres sont interdépendants: la détection est d'autant plus sensible (en termes de détection d'espèces rares) que la concentration de sonde est élevée, puisque les signaux lui sont proportionnels (c'est même le principe de la mesure). Et les quantités mises en jeu à ce niveau sont comparables pour les trois méthodes, de l'ordre du microgramme d'ARN messager. Quantité élevée, qui interdit par exemple d'effectuer une analyse à partir d'une biopsie ou de quelques dizaines de milliers de cellules obtenues par tri cellulaire, et qu'il est urgent de réduire. Regardons de plus près ce résultat à première vue paradoxal, puisque les concentrations de la sonde complexe diffèrent considérablement, les volumes d'hybridation se comptant en millilitres d'un côté (*macroarrays*), en microlitres de l'autre (*microarrays*, *oligo-chips*). Un calcul à partir des données publiées [9, 15, 22] montre pourtant que le seuil de détection (le nombre minimum de molécules d'une espèce spécifique qui doit être présent dans la sonde afin d'obtenir un signal détectable) est du même ordre, soit quelques dizaines de millions (Bertucci, communication personnelle). Cela découle du fait que la quantité de cible fixée sur Nylon est beaucoup plus importante que sur lame de verre<sup>2</sup>, ainsi que de la meilleure sensibilité intrinsèque de la détection radioactive. De plus, l'agitation du milieu d'hybridation dans lequel baigne la membrane évite tout épuisement local de la sonde, contrairement à ce qui se passe pour les *microarrays* souvent hybridés sous une lamelle de microscope avec un volume total d'une dizaine de microlitres. Naturellement, l'amplification de la sonde pra-

<sup>2</sup> Le signal observé est proportionnel non seulement à la concentration de la sonde, mais aussi à la quantité de cible: la probabilité d'hybridation pour chaque molécule de cible est indépendante du nombre de ces molécules, donc les signaux s'ajoutent à la seule condition que les interactions entre ces molécules soient négligeables, ce qui est le cas tout au moins sur support Nylon où les molécules d'ADN occupent moins de 1 % du volume du dépôt.

tiqué par *Affymetrix* [22, 23] recule cette limite ; mais, en l'absence de résultats publiés prouvant qu'elle ne fausse pas les valeurs d'abondance relative mesurées, cette méthode reste contestable.

Venons-en maintenant à la validation des mesures. Il ne suffit pas de collecter des dizaines de milliers de niveaux d'expression. Il convient aussi de garantir la validité des chiffres obtenus, l'absence de divers artefacts qui peuvent les affecter, et de connaître les limites de confiance des différentiels observés. On estime aujourd'hui que, dans des expériences bien conduites (quelle que soit l'approche), une variation d'un facteur deux pour le niveau d'expression d'un gène donné est significative. La comparaison entre une large série de sondes, de tissus ou de conditions demande plus de soins et de références, soit absolues, soit relatives pour aboutir à des données fiables.

Reste enfin l'interprétation détaillée des résultats. La masse d'informations qui résulte de ces expériences est considérable : le niveau d'expression de milliers de gènes est mesuré dans plusieurs tissus et/ou sous diverses conditions. Examiner ces données, les appréhender, les comprendre, n'est pas un problème trivial. De nombreux développements sont encore nécessaires au niveau des logiciels d'analyse et de représentation, afin d'extraire de ces chiffres le maximum de sens – et aussi de les rendre disponibles à la communauté des biologistes.

### Les grandes tendances

Après ces remarques critiques, qui veulent donner de la réalité une vision plus nuancée que les communiqués parfois triomphalistes d'entreprises en compétition féroce sur un marché porteur, reste à appréhender l'avenir de technologies qui continuent d'évoluer rapidement et font l'objet d'investissements considérables. Plusieurs tendances étaient perceptibles lors du colloque déjà cité. Je les regrouperai en trois thèmes : début de « sectorisation » du marché, accent sur l'exploitation des données, émergence possible de méthodes révolutionnaires.

Sectorisation du marché : expression barbare pour indiquer que, du moins pour les industriels concernés, il semble se profiler deux champs d'application, celui des puces complexes présentant un très grand nombre d'éléments et pouvant, par exemple, mesurer l'expression de dix mille gènes simultanément, et celui de systèmes plus simples visant quelques centaines d'entités. Le fait que la firme *Affymetrix* détienne des brevets lui assurant (d'après ses avocats) l'exclusivité d'arrays comportant mille éléments ou plus joue sans doute un rôle ; mais il est vrai aussi que beaucoup d'applications, notamment dans le domaine clinique, peuvent se contenter de suivre une ou deux centaines de gènes, et que le prix des puces complexes (de l'ordre du millier de dollars)<sup>1</sup> est à cet égard dissuasif. Des entreprises comme *Genometrix* [24] font cette analyse et s'équipent de manière à produire massivement des puces simples et à les commercialiser pour quelques dizaines de dollars l'unité, tout en proposant des systèmes de lecture ne faisant pas appel à une optique confocale, donc relativement abordables.

Accent sur l'exploitation des résultats : après l'apparition de systèmes pouvant mesurer des milliers de niveaux d'expression par expérience, les équipes qui ont commencé à les utiliser ont été rapidement submergées par l'abondance des données et ont dû créer des méthodes d'analyse permettant de les traiter et d'aboutir – si possible – à des conclusions biologiques. Le *National Cancer Institute* (NCI au NIH) a lancé un projet sur la pharmacologie moléculaire du cancer [25] dans le cadre duquel l'expression de dix mille gènes (mesurée à l'aide de *microarrays* fabriqués par l'entreprise *Synteni*) est maintenant suivie dans soixante lignées représentant les principaux types de cancer. Cette masse de données (six cents mille valeurs...) est corrélée avec la sensibilité de ces différentes

lignées à un large jeu de molécules. Des analyses de corrélation sophistiquées permettent alors de déceler, par exemple, les cas où la forte expression d'un groupe de gènes dans certaines lignées est associée avec la sensibilité à une certaine drogue. Les gènes en question deviennent alors des cibles potentielles pour ce composé, pouvant éclairer l'étude de son mode d'action...

Un des plus beaux exposés de cette conférence fut celui de Michael Byrne (de l'entreprise *Genetics Institute*, en collaboration avec plusieurs équipes académiques). Il s'agit cette fois de travaux effectués grâce à des *oligo-chips* produits par *Affymetrix*, avec qui *Genetics Institute* collabore depuis trois ans, et qui touchent des thèmes d'intérêt immunologique et médical. Une illustration, l'étude de l'activation de cellules T/NK dans le diabète, menée avec le jeu de quatre chips *Affymetrix* qui permet de suivre l'expression de 6800 gènes humains. Le matériel de départ est privilégié puisqu'il s'agit de lignées de cellules T/NK, établies à partir de deux jumeaux monozygotes dont l'un seulement est diabétique. Les profils d'expression ont été mesurés, dans les deux cas, avant et après stimulation des cellules par un anticorps anti-CD3. Chez la personne non diabétique, une augmentation d'expression très nette est observée pour plusieurs dizaines de gènes – ce qui, au passage, donne pour la première fois une vision exhaustive du processus à ce niveau et met en évidence plusieurs effets nouveaux ; chez son jumeau, la plupart de ces effets sont absents ou fortement réduits. Dans cette situation très bien définie (puisque, naturellement, le *background* génétique est identique), on a donc maintenant une douzaine de candidats, parmi les 6800 gènes humains connus, à une implication dans le mécanisme pathologique. J'évoquerai enfin un dernier exemple d'application large des profils d'expression. Celui-ci déclenche, du moins chez le chercheur « académique », un fort sentiment de frustration, lié au caractère secret des (très intéressantes) données obtenues. L'entreprise *Synteni* a été rachetée par la très performante *Biotech Incyte*

<sup>1</sup> Notons que toutes ces puces sont présentées comme à usage unique, ce qui est naturellement dans l'intérêt de leurs fabricants...

[26], qui a fait sa spécialité de la détermination d'EST et de leur inclusion dans des bases de données perfectionnées – accessibles aux industriels moyennant une redevance d'une dizaine de millions de dollars [27]. *Synteni* mesure actuellement sur une série de quatre *microarrays* le profil d'expression de quarante mille gènes humains<sup>1</sup> dans divers tissus normaux et pathologiques, en conditions de *stress*, après traitement par des drogues... et incorpore ces résultats dans une *UniGEM Expression Database* reliée aux bases de données d'*Incyte*, et dont l'accès sera vendu – certainement très cher – aux compagnies intéressées. On touche là un point déjà souligné [27]: ces informations de nature fondamentale, potentiellement utiles pour toute la Biologie, sont obtenues dans le secteur privé... et n'en sortiront pas. Décidément, la quête de nouvelles cibles amène les laboratoires industriels à entrer en concurrence directe avec la recherche fondamentale académique... mais avec de tout autres moyens !

### Nouveautés et perspectives

Émergence de techniques révolutionnaires enfin: la plus spectaculaire à mon sens était présentée par une équipe de la *Tufts University*. Il s'agit de la réalisation de *microarrays* sur fibres optiques. D'un diamètre total inférieur au millimètre, la fibre comprend dix mille filaments individuels; à l'extrémité de chacun, une cavité de quelques microns dans laquelle se trouve une microsphère conjuguée à un segment d'ADN. Lorsqu'on trempe l'extrémité du faisceau dans quelques microlitres d'une solution contenant un mélange de fragments d'ADN fluorescents, l'hybridation a lieu et la fluorescence correspondante peut être excitée par une illumination laser et mesurée individuellement pour chaque micro-

<sup>1</sup> Cet ensemble comporte la petite dizaine de milliers de gènes actuellement « connus » (dont la séquence complète, le produit et souvent la fonction ont été déterminés), et environ trente mille clusters constitués, comme dans *Unigene* (<http://www.ncbi.nlm.nih.gov/UniGene/index.html>), par regroupement informatique des séquences d'EST obtenues par *Synteni*.

fibre. Les résultats montrés pour cette méthode en cours de développement et dont le principe a déjà été publié fin 1996 [28] sont assez convaincants et en démontrent la faisabilité, même si de nombreux détails pratiques restent à résoudre. En extrapolant un peu, il est possible d'imaginer des fibres encore plus fines autorisant des mesures... à l'intérieur d'une cellule ! Science-fiction certes, mais qui illustre le bouillonnement d'idées et la qualité de l'innovation dans ce secteur. Signalons aussi de nombreux travaux sur la détection de l'hybridation par des méthodes électriques (et non par la mesure de fluorescence), susceptibles d'accroître considérablement la vitesse d'acquisition, de lever les derniers obstacles à une miniaturisation encore plus poussée et d'aboutir à des puces intégrant les segments d'ADN et le dispositif de mesure.

Tirons maintenant quelques leçons de ce périple au pays des *macroarrays*, *microarrays* et puces à ADN. Il est clair qu'aujourd'hui la mesure du profil d'expression de dizaines de milliers de gènes dans des dizaines de conditions différentes est techniquement possible, que son utilité est scientifiquement et industriellement incontestable, et que les outils d'analyse permettant d'en déduire des cascades de signalisation ou de métabolisme commencent à apparaître. Il est évident aussi que la majorité des projets menés le sont dans un cadre et avec des financements industriels. Leurs résultats ne seront pas généralement disponibles, ce qui est fort dommage pour nous tous... Notons tout de même les quelques travaux (comme ceux du NCI) destinés à être rendus publics, et souhaitons que des financements (ainsi que des équipes compétentes et motivées) soient trouvés pour qu'un ensemble cohérent de résultats aboutisse dans des bases de données publiques. Dans un registre plus optimiste, il est permis d'espérer que l'amélioration de ces méthodes débouche sur une accessibilité accrue. Celle-ci est aujourd'hui réduite, peu d'appareils sont disponibles et leur prix est encore prohibitif, mais on peut compter sur une intense concurrence entre les industriels qui, à l'instar de *Geno-*

*metrix*, s'attachent à mettre au point des versions *light* et abordables de cette technologie.

Resterait à intégrer concrètement ce versant de l'expérimentation biologique parmi les approches que nous pratiquons au laboratoire. Cela suppose une double stratégie. D'une part, un investissement massif dans la bio-informatique, de manière à faire le meilleur usage des données qui seront rendues publiques; d'autre part, la mise en place d'équipements bien dimensionnés, permettant de réaliser des *arrays* de quelques milliers de gènes adaptés aux besoins de nos équipes, et de les exploiter de manière efficace. Inutile à cet égard de vouloir faire la course avec *Affymetrix* ou *Synteni*; pas question pourtant de se priver d'employer cette technologie, d'autant qu'elle va devenir de plus en plus accessible. Et, bien sûr, il faudra veiller à éviter de se placer en compétition directe avec la recherche industrielle et chercher, au contraire, une complémentarité fondée sur la connaissance intime de différents systèmes biologiques... ■

### Bertrand R. Jordan

Équipe TAGC, ICIM, Centre d'Immunologie Inserm-Cnrs de Marseille-Luminy, case 906, 13288 Marseille Cedex 09, France.

### RÉFÉRENCES

1. <http://www.ncbi.nlm.nih.gov/dbEST/index.html>
2. Marshall A. HGS launches « first » genomics product in clinic. *Nat Biotechnol* 1998; 16: 129.
3. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992; 257: 967-71.
4. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995; 270: 484-7.
5. Lennon GG, Lehrach H. Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 1991; 7: 314-7.
6. Clontech « Atlas array »: <http://www.clontech.com/clontech/Catalog/Hybridization/Atlas.html> Genome Systems « Gene discovery array »: <http://www.genomesystems.com/GDA/>. Research Genetics « GeneFilters »: <http://www.resgen.com/>

## RÉFÉRENCES

7. Chen JJW, Wu R, Yang PC, *et al.* Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* 1998 (sous presse).
8. Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H. Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm Genome* 1992; 3: 609-61.
9. Nguyen C, Rocha D, Granjeaud S, Baldit M, Bernard K, Naquet P, Jordan BR. Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* 1995; 29: 207-15.
10. Zhao N, Hashida H, Takahashi N, Misumi Y, Sakaki Y. High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. *Gene* 1995; 156: 207-13.
11. Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Samson R, Houlgatte R, Soularue P, Auffray C. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Research* 1996; 6: 492-503.
12. <http://www.hyseq.com/>
13. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270: 467-70.
14. Derisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 1996; 14: 457-60.
15. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 1996; 93: 10614-9.
16. Derisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; 278: 680-6.
17. <http://cmgm.Stanford.EDU/pbrown/>
18. Khrapko KR, Lysov YuP, Khorlin AA, Ivanov IB, Yershov GM, Vasilenko SK, Florentiev VL, Mirzabekov AD. A method for DNA sequencing by hybridization with oligonucleotide matrix. *DNA Seq* 1991; 1: 375-88.
19. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991; 251: 767-73.
20. Southern EM, Maskos U, Elder JK. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics* 1992; 13: 1008-17.
21. <http://www.affymetrix.com/>
22. Lockhart DJ, Dong H, Byrne MC, *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996; 14: 1675-80.
23. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 1997; 15: 1359-67.
24. <http://www.genometrix.com/>
25. Weinstein JN, Myers TG, O'Connor PM, *et al.* An information-intensive approach to the molecular pharmacology of cancer. *Science* 1997; 275: 343-9.
26. <http://www.incyte.com/>
27. Jordan B. Du programme Génome à la « pharmacogénomique ». *Med Sci* 1997; 10: 1176-7.
28. Ferguson JA, Boles TC, Adams CP, Walt DR. A fiber-optic DNA biosensor microarray for the analysis of gene expression. *Nat Biotechnol* 1996; 14: 1681-4.

## TIRÉS À PART

B.R. Jordan.

# SOCIÉTÉ DE BIOLOGIE

Cent-cinquantième de la Société de Biologie

**19 novembre 1998 (14 h - 18 h)**

Séance des membres correspondants de la Société de Biologie

**Harold Kalant** (Professeur à l'Université de Toronto)

Rôle de la Science dans la formulation de la politique sociale sur la drogue

**Christian Doutremépuich** (Professeur à l'Université de Bordeaux)

Polymorphisme de l'ADN, applications médico-légales

**Ba Denian** (Président de l'Académie des Sciences Médicales de Chine)

La recherche en immunologie du cancer en Chine

**Chen Zhu** (Directeur de l'Institut d'Hématologie de Shanghai II)

Le génome humain en Chine

**Susan Hollán** (Professeur à l'Institut National d'Hématologie et d'Immunologie de Budapest)

Déficiences en enzymes glycolytiques et neurodégénérescence

**Gustav Born** (Professeur à l'Institut de Recherche William-Harvey de Londres)

Les effets d'hormones et de médicaments sur la fixation de protéines plasmatiques athérogènes sur les parois artérielles

**Hannes Stähelin** (Professeur à l'Hôpital Cantonal de l'Université de Bâle)

Vitamine D plasmatique et force musculaire chez les sujets âgés

**Francis Karst** (Professeur à l'Université de Poitiers)

Biosynthèse et transport des stérols chez la levure

Collège de France

11, place Marcelin-Berthelot – 75005 Paris, France