



La puce à ADN en médecine et en science

**Peter Lee
Thomas J. Hudson**

P. Lee, T.J. Hudson : Centre génomique de Montréal, Université McGill, 1650, Cedar avenue, Montréal, Québec, H3G 1A4 Canada.

► La révolution génomique offre de nouveaux outils pour l'étude de processus biologiques complexes à l'échelle pan-génomique. Cet article passe en revue les différents principes et applications de l'une de ces technologies émergentes, celle des puces à ADN. Il existe actuellement deux procédés majeurs de fabrication de puces à ADN : (1) le dépôt direct de molécules d'ADNc ; (2) la synthèse *in situ* d'oligonucléotides sur une surface solide. Ces deux types de procédés, bien que présentant des différences qualitatives, offrent tous deux la possibilité d'un grand nombre d'applications nouvelles, à la fois fondamentales et cliniques, en permettant d'étudier simultanément plusieurs milliers de gènes et de découvrir rapidement de nombreux polymorphismes fonctionnels au niveau génomique. ◀

Le projet de séquençage du génome humain est en plein développement, et une première version préliminaire sera probablement disponible dès le printemps de cette année (voir l'article de Jean Weissenbach et Marcel Salanoubat, p. 10 de ce numéro). L'avènement de ce programme de séquençage, et des projets similaires concernant d'autres organismes, révolutionne la recherche en biomédecine, tant par l'élaboration de nouvelles technologies d'analyses de l'ADN que par la création d'immenses banques de données informatiques. Ces ressources extraordinaires amènent la communauté scientifique à se poser de nouvelles questions et permettront vraisemblablement d'élucider des mécanismes moléculaires complexes. Traditionnellement, les biologistes ont utilisé des approches réductionnistes afin de disséquer un problème. La portée de chacune des questions posées est ainsi bien souvent limitée, tous les efforts étant tournés vers la résolution de points très spécifiques. Nous analysons ainsi de façon minutieuse, à l'échelle moléculaire, chaque composante d'un processus biologique de base. Cependant, les

techniques de biologie moléculaire classique trouvent leurs limites lorsqu'elles sont appliquées à l'élucidation de processus complexes. Les fruits des divers projets de séquençage de génomes offrent de nouveaux outils afin d'étudier ces processus à l'échelle du génome. Dans cet article, nous examinerons l'un de ces outils, dont l'utilisation est de plus en plus répandue, la puce à ADN ou *microarray*.

Le principe de l'hybridation moléculaire permettant de détecter la présence d'acides nucléiques est maintenant bien établi. Les *Southern blots* et les *Northern blots* font depuis longtemps partie des techniques de base de tous les laboratoires de recherche en biologie. Le jumelage de plusieurs technologies a permis la miniaturisation de ces techniques d'hybridation, permettant ainsi de déceler des milliers de molécules d'acide nucléique de façon simultanée sur des matrices solides mesurant quelques centimètres carrés. Deux procédés majeurs de fabrication de puces à ADN sont couramment utilisés : (1) le dépôt direct d'ADNc sur lamelle de verre activée ; ou (2) la synthèse *in situ* d'oligonucléotides par photolithographie (figure 1).

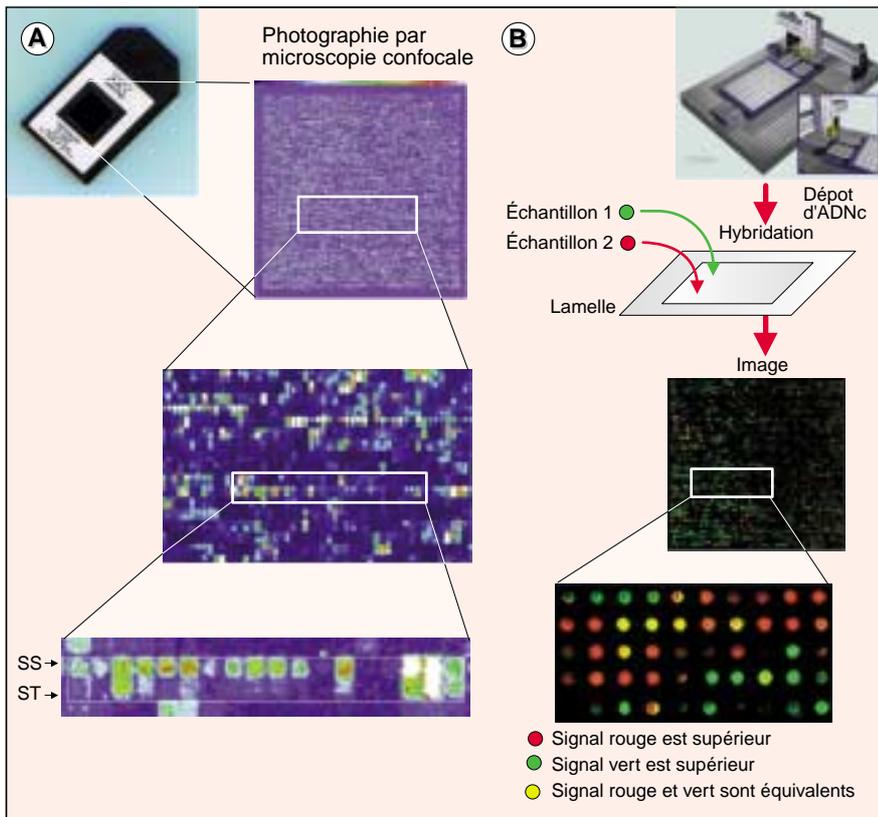


Figure 1. **Puces à ADN.** **A.** Puce à ADN produite par Affymetrix. La puce à ADN est contenue dans une plaquette de plastique contenant une chambre d'hybridation. Après l'hybridation d'un échantillon d'ADN ou d'ARN marqué par un fluorophore, la surface contenant les 60 000 à 400 000 oligonucléotides est analysée par microscopie confocale et photographiée. Le grossissement de l'image permet de visualiser chaque espèce d'oligonucléotide retrouvé sur une surface de 20 µm x 20 µm. Dans l'exemple du bas, un gène est représenté par une série d'oligonucléotides de 25 nucléotides dérivés de la séquence du gène (SS: sonde spécifique). Chaque oligonucléotide possède son propre contrôle d'hybridation, obtenu par la synthèse d'un second oligonucléotide dont la séquence varie d'un seul nucléotide en position centrale (ST: sonde témoin). La concentration de l'ARN est mesurée par la moyenne des différences des oligonucléotides SS et SC. **B.** Puces à ADN préparées en parallèle à l'aide d'un micropipetteur robotisé qui dépose des ADNc sur la surface de la puce. Deux échantillons d'ARN provenant de différents tissus ou traitements sont marqués par des fluorophores différents (Cy-3 vert et Cy-5 rouge). La quantité relative de chaque gène est déterminée par le rapport d'émission de chaque fluorophore à des longueurs d'ondes différentes. La partie B est adaptée du site internet (<<http://cmgm.Stanford.EDU/pbrown/>><http://cmgm.Stanford.EDU/pbrown/>) avec la permission de Joseph DeRisi.

Dépôt de sondes sur puce à ADN

Le premier type de puce à ADN consiste en une lamelle de verre (identique à celle utilisée en microscopie traditionnelle) sur laquelle des milliers d'ADNc sont déposés à l'aide d'un micropipetteur robotisé. Grâce à cette technique, chacun des gènes (de fonction connue ou inconnue) est repré-

senté par un seul point sur la lamelle. En général, deux échantillons d'ARN (sous forme d'ADNc obtenus par transcription inverse) sont co-hybridés sur la puce à ADNc. Les deux échantillons marqués par un fluorophore différent (Cy-3 vert ou Cy-5 rouge) s'hybrident simultanément avec les molécules complémentaires sur la puce. L'intensité du signal lumineux mesurée aux deux longueurs d'ondes correspondant

aux différents fluorophores est alors mesurée à l'aide d'un microscope confocal. Le rapport de fluorescence rouge/vert est ainsi déterminé et permet de comparer les taux d'expression relatifs de chacun des gènes pour les deux échantillons d'ADNc. Un excès du gène X dans l'échantillon marqué en rouge produira un signal rouge; un excès du gène Y dans l'échantillon marqué en vert produira un signal vert; enfin, une expression équivalente du gène Z dans les deux échantillons produira un signal jaune. L'un des avantages de cette analyse comparée repose sur le fait que le rapport rouge/vert n'est pas influencé par la qualité de la goutte déposée par le pipetteur robotisé.

Synthèse d'oligonucléotides sur puce à ADN

Le second type de puce à ADN, proposé par la société Affymetrix, est constitué d'oligonucléotides synthétisés directement sur un substrat solide par photolithographie. Dans ce procédé, une lumière dirigée sur des sites spécifiques de la puce active la réaction d'oligo-synthèse [1, 2]. La synthèse d'un oligonucléotide de 25 paires de bases occupe un carré de 20 µm x 20 µm et contient plus de 10⁷ copies de cette molécule. La surface d'une puce est d'environ 1,28 cm², et peut contenir 400 000 oligonucléotides différents! Une puce à ADN destinée à des études d'expression contient pour chaque gène un ensemble d'oligonucléotides mimant la séquence du gène, souvent choisis dans sa région 3', réduisant ainsi les risques d'hybridations croisées avec des séquences homologues de ce gène. Des oligonucléotides, dont la séquence varie pour une seule base, sont également ajoutés, ce qui permet de confirmer que le signal obtenu pour chacun des gènes est bien spécifique. Contrairement à la puce à ADN décrite plus haut, celle produite par ce procédé permet l'hybridation d'un seul échantillon marqué à la fois. L'intensité de l'hybridation est également mesurée par microscopie confocale.

Comparaison des puces à ADN construites par ces deux procédés

Les puces à ADN produites par micropipetteurs offrent une grande

souplesse d'utilisation car il est facile pour le chercheur d'en modifier le contenu. En outre, elles sont relativement peu coûteuses, et, pour les amateurs, les modalités de construction d'un système robotisé sont disponibles sur Internet (Tableau I). La préparation et l'optimisation de ces puces à ADN ne sont cependant pas simples, l'assemblage de milliers de gènes sur une puce nécessitant la validation et la purification de nombreux ADNc. Les puces à ADN produites par Affymetrix permettent d'étudier plus de 45 000 gènes humains, plus de 30 000 gènes murins et environ 6 000 gènes de levures. Elles ont cependant le désavantage d'être très coûteuses et leur contenu n'est pas modulable. Il est à souhaiter qu'avec le temps, elles deviennent plus fonctionnelles.

Ces deux technologies sont d'ores et déjà très prometteuses. D'autres types de puces, comme les biopuces électroniques (Nanogen), dans lesquelles des circuits électriques miniatures sont utilisés afin de diriger les tests moléculaires à la surface de la puce,

sont maintenant en développement. Cette technologie permettra entre autres la séparation de cellules par affinité ainsi que le développement d'autres tests moléculaires.

Applications des puces à ADN

Analyses d'expression de gènes

Les premières puces ont servi à évaluer l'expression simultanée de milliers de gènes dans des systèmes biologiques bien connus, tels que celui du métabolisme respiratoire et de fermentation chez la levure [3], le cycle cellulaire de la levure (figure 2) [4, 5], la sporulation [6] et la stimulation de fibroblastes par le sérum [7]. Ces premiers travaux ont permis de valider la technologie. La comparaison des résultats obtenus par les puces à ADN avec ceux préalablement obtenus par d'autres approches a démontré une concordance pour les gènes dont l'expression était déjà connue dans ces systèmes biologiques. De plus, pour les milliers de nouveaux résultats obtenus,

il existe une conformité « intuitive » avec des connaissances provenant d'autres processus cellulaires.

Analyses de voies biochimiques

Les puces à ADN permettent l'analyse de voies métaboliques spécifiques. Ainsi, Fambrough *et al.* [8] ont étudié la cascade de signalisation des récepteurs à activité tyrosine kinase dans les cellules NIH-3T3. Cette famille de récepteurs membranaires est conservée de la levure jusqu'à l'homme et joue un rôle fondamental dans la transmission de signaux de la membrane au noyau cellulaire. L'une des analyses utilisant des puces à ADN a été réalisée avec des cellules exprimant un récepteur muté du PDGF (*platelet derived growth factor*), sur lequel les résidus tyrosine potentiellement spécifiques de certaines cascades de signalisation ont été modifiés. Or, cette analyse a révélé une redondance fonctionnelle considérable entre les différents mutants, qui n'avait pu être anticipée par les seuls modèles exploités anté-

Tableau I

RESSOURCES INFORMATIQUES SUR INTERNET

URL	Description	Référence
http://cmgm.stanford.edu/pbrown/explore/	Cycle métabolique de la levure	[3]
http://genome-www.stanford.edu/Saccharomyces/	Base de données sur le cycle cellulaire de la levure	[4]
http://genome-ww.stanford.edu/cellcycle/	Analyse du cycle cellulaire de la levure	[5]
http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/	Projet de micropuces au NHGRI. Distribution de logiciels ArrayDB et ArrayViewer	[19]
http://web.wi.mit.edu/young/expression/	Données d'expression pan-génomique	[20]
http://cmgm.stanford.edu/pbrown/sporulation/index.html	Programme transcriptionnel de sporulation	[6]
http://genome-www.stanford.edu/serum/	Base de données d'expression sur la stimulation par le sérum	[7]
http://rana.stanford.edu/clustering/	Logiciel Cluster	[16]
http://waldo.wi.mit.edu/MPR/	Logiciel GeneCluster	[17]
http://www.affymetrix.com/technology/index.html	Aperçus des puces à ADN Affymetrix	
http://cmgm.stanford.edu/pbrown/mguide/index.html	Guide de fabrication de robot pour puces à ADN	

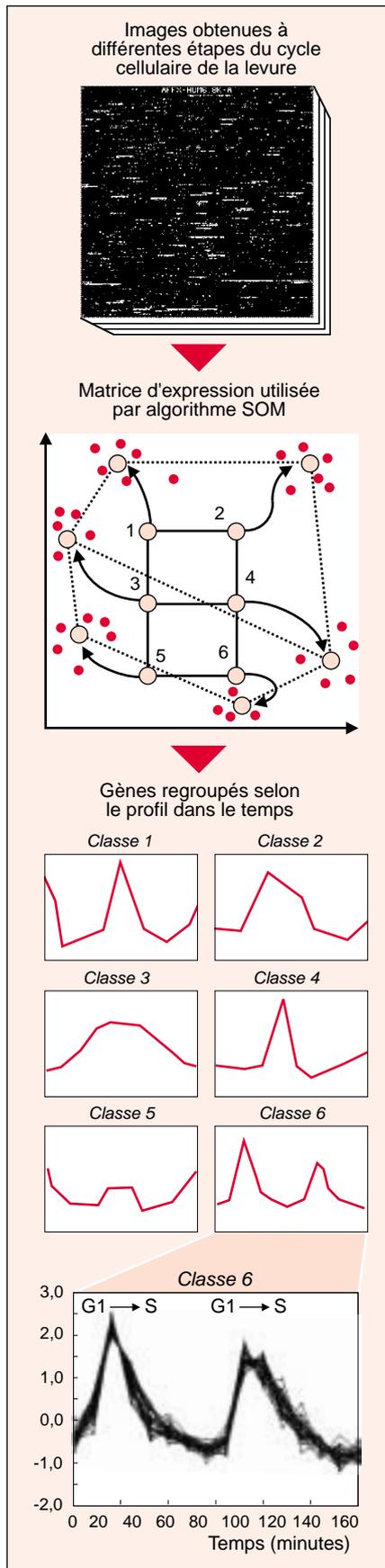


Figure 2. Analyse d'expression du cycle cellulaire de la levure. Les données provenant de multiples mesures d'expression par puces à ADN à différentes étapes du cycle cellulaire de la levure engendrent un déluge d'informations. Un algorithme nommé Self-Organizing Maps (SOM), inventé par Kohonen [21], a été utilisé par Tamayo [17] pour l'étude des données génomiques. Cette méthode permet de regrouper les gènes ayant un profil d'expression semblable. Brièvement, les données sont d'abord localisées dans un espace (ou matrice d'expression) et ensuite groupées selon leur proximité dans cet espace, signifiant qu'elles ont des profils d'expression semblables. Dans l'exemple provenant du cycle cellulaire, plusieurs classes de gènes ont été décelées (de fonctions connues et inconnues), ayant des profils d'expression démontrant une périodicité semblable pendant le cycle cellulaire. L'exemple donné à la partie inférieure de la figure représente les gènes ayant une expression maximale durant la transition de la phase G1 à la phase S du cycle cellulaire de la levure. (Adapté de l'article de Tamayo et al. [17] avec la permission de Proc Natl Acad Sci USA.)

rièvement. L'étude a aussi permis de démontrer que ces récepteurs peuvent remplir d'autres fonctions, telles que la stimulation de la production d'interleukine(s) pour l'un des récepteurs mutés. Il est possible que certaines de ces réponses soient réprimées dans le cas des récepteurs normaux, et l'utilisation de puces à ADN aurait alors permis de mettre en évidence l'existence de voies biochimiques alternatives.

Validation de mécanismes d'action de médicaments

Les puces à ADN peuvent également être utilisées pour étudier le mécanisme d'action d'un médicament. En principe, un médicament qui agit par inhibition spécifique d'un seul gène ou de son produit devrait engendrer un effet identique à celui résultant de l'inactivation de ce gène par délétion ou par mutation. Marton et al. [9] ont utilisé une puce à ADN contenant l'ensemble des gènes de la levure afin de démontrer l'existence d'une corrélation significative entre le profil

obtenu lors d'une stimulation médicamenteuse antimicrobienne et le profil d'expression d'une levure portant un gène muté et impliqué dans le métabolisme d'action de ce médicament. Ce principe peut être exploité pour la création d'une base de données contenant un grand nombre de profils d'expression, provenant à la fois de cellules stimulées par des médicaments et de souches contenant différentes mutations. Ces données offrent un moyen de « décoder » les profils complexes d'expression de groupes de gènes modulés par différentes classes de médicaments.

Classification phénotypique et prédiction

En clinique, la distinction entre une leucémie myéloïde aiguë (LMA) et une leucémie lymphoïde aiguë (LLA) est d'une importance cruciale afin de bien orienter le traitement thérapeutique. Néanmoins, les techniques usuelles de cytopathologie sont complexes et requièrent une très grande expertise. Golub et al. [10] ont tenté de définir différentes classes de leucémies sur la base du profil d'expression de près de 6 000 gènes humains. L'analyse d'une soixantaine de leucémies présentant des cytopathologies connues a permis de distinguer plus d'une centaine de gènes ayant un profil différent dans les cas de LMA et de LLA. Bien qu'aucun de ces gènes (telle la myéloperoxidase) n'ait un mode d'expression identique pour toutes les leucémies d'une même classe, l'analyse du groupe composé des 50 à 100 gènes les plus différents entre les deux classes de leucémies a une valeur prédictive significative (pratiquement de 100 % dans les cohortes de validation ultérieure, lorsque le taux de confiance est jugé satisfaisant). De plus, ces analyses ont permis d'identifier un troisième type de leucémies non identifiables par la seule approche cytopathologique. Les analyses par les puces à ADN ont donc une application immédiate en clinique, et présentent en outre la possibilité d'identifier de nouvelles classes de phénotypes pathologiques.

Analyses d'ADN génomique

L'étude des variations génomiques est d'une grande utilité en recherche bio-

médicale. Une grande partie de la variabilité interindividuelle observée au sein d'une même espèce, en particulier la susceptibilité à certaines maladies, est due à des différences (ou polymorphismes) existant au niveau de la séquence de l'ADN génomique. En raison de leur potentiel et de la rapidité d'analyse qu'elles présentent, les puces à ADN offrent un avantage considérable sur les techniques déjà existantes pour aborder ces variations.

Études génomiques comparées

Behr *et al.* [11] ont étudié plusieurs souches du bacille de Calmette et Guérin (BCG) originaires de l'Institut Pasteur au cours du xx^e siècle (figure 3). A partir de la séquence complète de *M. tuberculosis*, un assemblage de sondes génomiques

couvrant l'ensemble de ce génome a été déposé sur une puce à ADN. Le génome de chaque souche de BCG étudiée a été amplifié, marqué par un fluorochrome et hybridé sur la puce à ADN. Cela a permis d'identifier des divergences au niveau génomique et de les corrélérer de façon temporelle avec l'historique de la dissémination mondiale de ces souches. Cette étude suggère l'existence d'un lien entre l'efficacité du vaccin contre le BCG et la prééminence des souches classifiées par ces marqueurs génomiques. A l'avenir, ce type d'analyse sera sûrement appliquée à d'autres pathogènes.

Reséquençage : détection de SNP

Grâce aux puces à ADN, une séquence connue peut être séquencée de nouveau dans l'intention de découvrir des polymorphismes affectant un seul nucléotide (SNP, *single nucleotide polymorphisms*). Cette technologie utilise le principe de séquençage par hybridation (SPH) [12]. Brièvement, une séquence connue peut être caractérisée par un assemblage d'oligonucléotides chevauchants. Pour cette application par-

ticulière, il faut donc créer une puce à ADN contenant tous les 25 mers chevauchants, définissant la séquence à interroger ainsi que trois amorces contenant les trois permutations possibles pour le nucléotide central de l'amorce (c'est-à-dire qu'un T sera remplacé par un A, un C et un G). Cette méthode a été utilisée par Wang *et al.* [13] pour reséquencer 2,3 Mb du génome chez 7 individus et a ainsi permis d'identifier 3 241 SNP humains. Cette approche a également été utilisée afin d'identifier des polymorphismes dans un gène de susceptibilité au cancer du sein (*BRCA1*) [14] ainsi que pour plus d'une centaine d'autres gènes impliqués dans des processus vasculaires, métaboliques et endocriniens [15]. Enfin, le reséquençage de génomes de pathogènes (tel celui du VIH) permet d'identifier des mutations fonctionnelles, ouvrant ainsi la voie vers des thérapies spécifiques et déterminées en fonction des polymorphismes génomiques du pathogène.

Génotypage

La détermination du génotype pour un SNP donné exploite la haute spé-

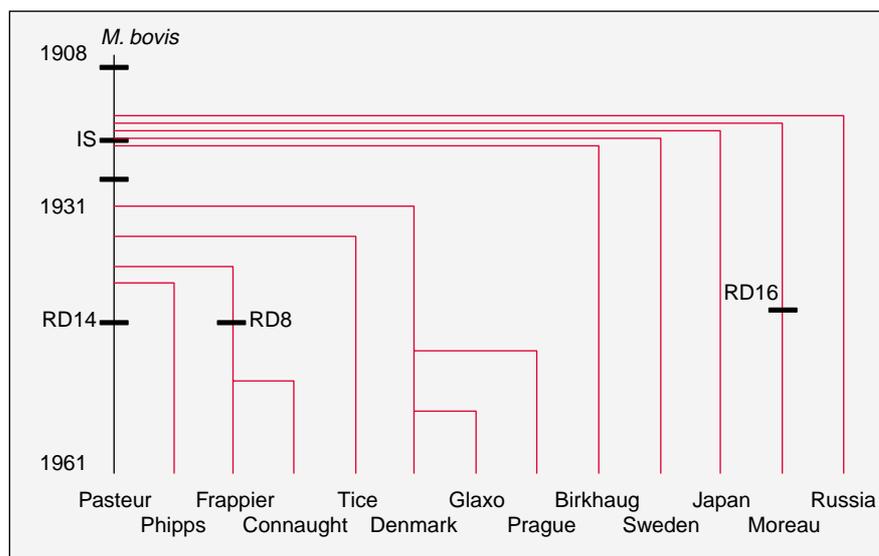
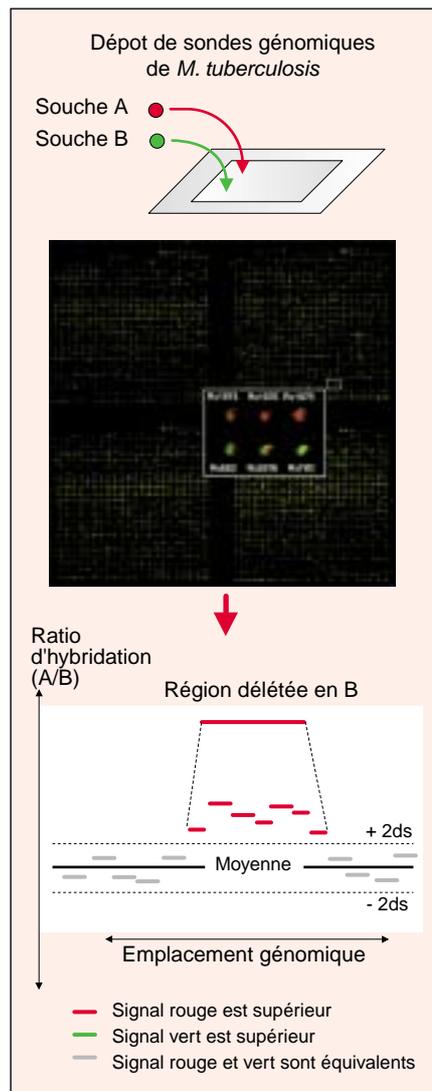


Figure 3. **Analyse génomique de souches BCG.** Une puce à ADN contenant des sondes génomiques de *M. tuberculosis* a été utilisée pour interroger le génome de plusieurs souches de BCG [11]. Dans cet exemple, les comparaisons entre deux souches ont permis d'identifier de petites régions génomiques affichant un excès de couleur rouge, signifiant une délétion au sein du génome de la souche B marquée en vert. L'analyse complète de plusieurs souches a mis en évidence plusieurs microdélétions (contenant parfois un ou plusieurs gènes). (Adapté de l'article de Behr *et al.* [11] avec la permission de Science.)

cificité de la puce à ADN à discriminer une complémentarité parfaite *versus* une complémentarité imparfaite causée par une seule paire de bases non appariées. La puce à ADN de génotypage (figure 4) contient donc des amorces spécifiques de chacun des allèles, outre des amorces témoins [13]. Une puce à ADN a la capacité de contenir des milliers de polymorphismes. Les applications possibles pour ces puces à ADN sont nombreuses : criblage génomique pour études de liaison génétique, études d'association avec des milliers de gènes candidats, caractérisation d'anomalies cytogénétiques, etc.

Informatique

La gestion du déluge d'informations produites par ces technologies représente un défi énorme pour les biologistes. L'un des facteurs limitants est donc la capacité d'analyse de l'information, pour laquelle de nouvelles méthodes sont nécessaires. Une nouvelle génération d'outils informatiques est actuellement en développement : ainsi, pour réaliser les études d'expression, il existe des algorithmes capables d'identifier des groupes de gènes partageant des profils d'expression semblables (*cluster* [16], *gene cluster* [17]). D'autres méthodes sont destinées à trouver des relations entre des gènes ayant des expressions identiques, comme la recherche de motifs semblables au niveau des séquences localisées en 5' ou 3' du gène (par exemple, Yeast Toolset AlignACE pour la levure [18]). La création de ces nouvelles bases de données est indispensable afin de pouvoir regrouper, visualiser et partager les profils d'expression obtenus par les puces à ADN (ArrayViewer, ArrayDB [19]). Le Tableau 1 dresse la liste des ressources informatiques disponibles sur Internet et qui concernent les puces à ADN.

Ainsi, moins de cinquante ans après la découverte de la structure de l'ADN, nous entrons résolument dans l'ère post-génomique. Ces technologies fascinantes permettront d'explorer l'univers cellulaire dans toute sa complexité. Ayant gardé longtemps les apparences d'un roman de science-fiction, la révolution génomique est désormais une réalité captivante ■

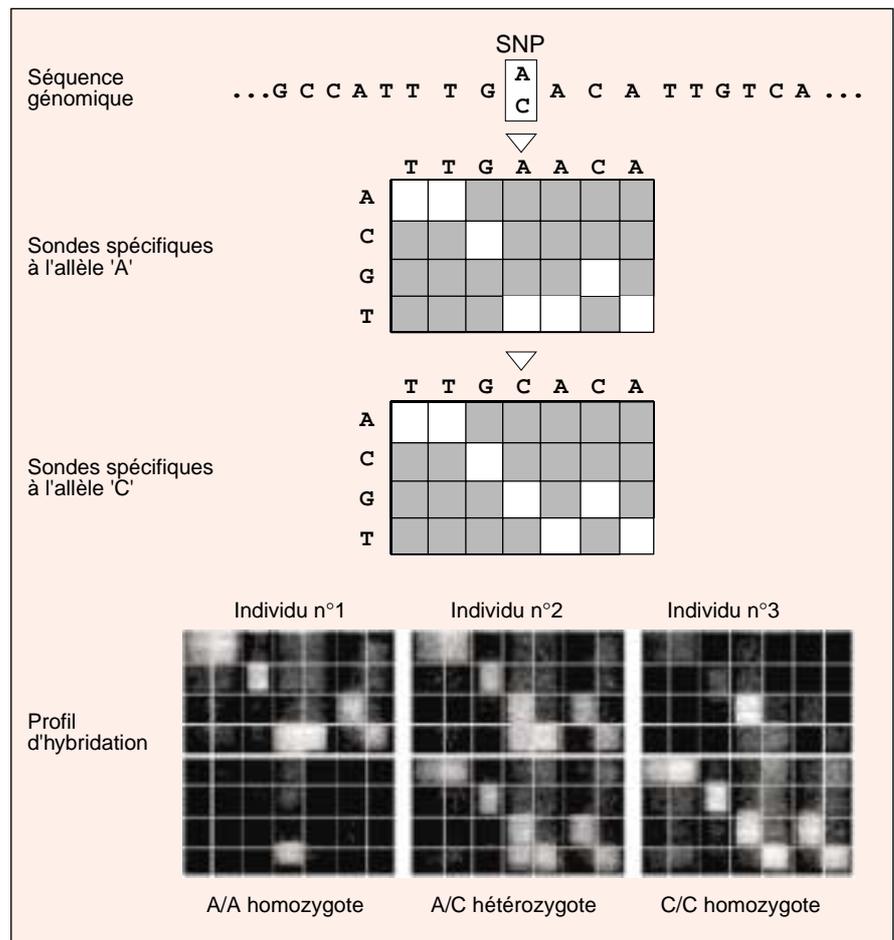


Figure 4. **Génotypage de SNP.** Pour réaliser le génotypage d'un SNP, tel que celui qui apparaît en haut de cette figure, deux séries d'oligonucléotides doivent être présents sur la puce, chaque série représentant un allèle différent. Chaque colonne contient des oligonucléotides successifs qui sont complémentaires de la séquence interrogée et dont la base centrale est substituée par un A, C, G, ou T dans les quatre rangées. Les sondes spécifiques aux allèles A et C apparaissent en blanc, tandis que les sondes témoins apparaissent en gris. Le génotype d'un individu peut être déterminé par l'analyse de la variation du signal d'hybridation provenant d'un produit PCR marqué contenant le SNP. Le signal d'hybridation de trois individus avec les génotypes AA, AC et CC est présenté à la partie inférieure de la figure. Une micropuce peut d'ailleurs interroger des milliers de SNP en parallèle. (Adapté de l'article de Wang et al. [13] avec la permission de Science.)

Remerciements

Les auteurs remercient Marcel Behr, Jean-Paul Comet, Marie-Claude Vohl et Claire Goguen pour leurs précieux commentaires.

RÉFÉRENCES

1. Chee M, Yang R, Hubbell E, et al. Accessing genetic information with high-density DNA arrays. *Science* 1996; 274: 610-4.
2. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996; 14: 1675-80.

3. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; 278: 680-6.
4. Cho RJ, Campbell MJ, Winzeler EA, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 1998; 2: 65-73.
5. Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998; 9: 3273-97.
6. Chu S, DeRisi J, Eisen M, Mullholland J, Botstein D, Brown PO, Herskowitz I. The transcriptional program of sporulation in budding yeast. *Science* 1998; 282: 699-705.

RÉFÉRENCES

7. Iyer VR, Eisen MB, Ross DT, *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* 1999; 283: 83-7.
8. Fambrough D, McClure K, Kazlauskas A, Lander ES. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* 1999; 97: 727-41.
9. Marton MJ, DeRisi JL, Bennett HA, *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* 1998; 4: 1293-301.
10. Golub TR, Slonim D, Tamayo P, *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 2000 (sous presse).
11. Behr MA, Wilson MA, Gill WP, *et al.* Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999; 284: 1520-3.
12. Drmanac S, Kita D, Labat I, *et al.* Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nat Biotechnol* 1998; 16: 54-8.
13. Wang DG, Fan JB, Siao CJ, *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; 280: 1077-82.
14. Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 1996; 14: 441-7.
15. Cargill M, Altshuler D, Ireland J, *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; 22: 231-8.
16. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; 95: 14863-8.
17. Tamayo P, Slonim D, Mesirov J, *et al.* Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999; 96: 2907-12.
18. Van Helden J, Andre B, Collado-Vides J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol* 1998; 281: 827-42.
19. Ermolaeva O, Rastogi M, Pruitt KD, *et al.* Data management and analysis for gene expression arrays. *Nat Genet* 1998; 20: 19-23.
20. Holstege FC, Jennings EG, Wyrick JJ, *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 1998; 95: 717-28.
21. Kohonen T. *Self-organizing maps*. Berlin: Springer Verlag, 1997.

TIRÉS À PART

T.J. Hudson.

MS2000

Summary

DNA chips in medicine and science

The Human Genome Project is changing our conception of modern biology. Recent advances in technology are now enabling us to observe complex processes on a genome-wide scale. This review examines the emerging technology of DNA microarrays. Notwithstanding the differences related to manufacture characteristics and properties of the two major technologies used today, DNA microarrays offer the potential to simultaneously investigate thousands of genes. Expression DNA chips containing gene probes rely on the expression profile of collections of genes to investigate complex biochemical pathways, validate drug targets, and classify cell phenotype. Microarrays may be used to detect variations in DNA sequences and correlate these with phenotypes – as in genome scans for linkage studies, mutations detection, large-scale association studies, and analyses of drug responses. Numerous applications related to modern medicine in the areas of diagnostics and drug management are rapidly emerging.