

Introduction aux puces à ADN

Le concept de puce à ADN repose sur une technologie multidisciplinaire intégrant la biologie, la nanotechnologie, la chimie des acides nucléiques, l'analyse d'images et la bioinformatique. Grâce à cet outil, il est possible de mesurer le niveau d'expression de plusieurs milliers de gènes simultanément, et les applications (*e.g.* détermination des familles de gènes co-régulés, recherche de systèmes de régulation, ...) sont en plein essor dans un grand nombre de domaines comme la pharmacologie, la médecine ou l'environnement.

L'analyse biologique des premières biopuces repose sur l'étude du transcriptome (étude des ARN messagers) d'une cellule ou d'un organisme. Si de nouvelles applications pour les puces à ADN ont également été développées (biopuces pour le diagnostic, pour la génomique comparative, pour l'identification de régions d'ADN régulatrices après ImmunoPrécipitation de la Chromatine (ChIPchips)), seules les biopuces transcriptome sont effectivement abordées dans ce chapitre.

1- De la cellule au transcriptome

De la cellule à l'ADN

La cellule est l'unité de base biologique. Chez les organismes les plus simples que sont les procaryotes (unicellulaires), le matériel génétique n'est pas compartimenté dans un noyau vrai mais est libre dans le cytoplasme. Pour les organismes plus complexes, les Eucaryotes (uni- ou pluricellulaires), l'information génétique est compartimentée dans un noyau (présence d'une membrane nucléaire). L'ADN, présent dans ce noyau, est une molécule qui peut être gigantesque avec un enchaînement linéaire de millions de nucléotides (l'adénine (A), la cytosine (C), la guanine (G) et la thymine (T)). L'appariement des paires de nucléotides de type Watson et Crick (A et T d'une part, C et G d'autre part) concourt à l'organisation de l'ADN en double hélice, dont chaque brin a alors une séquence parfaitement complémentaire à l'autre (Figure A-1). Cette complémentarité est à la base de la plupart des techniques d'analyse en biologie moléculaire. Pour une espèce donnée, la séquence de la molécule d'ADN, *i.e.* l'enchaînement des nucléotides, est globalement semblable, sauf cas de polymorphismes.

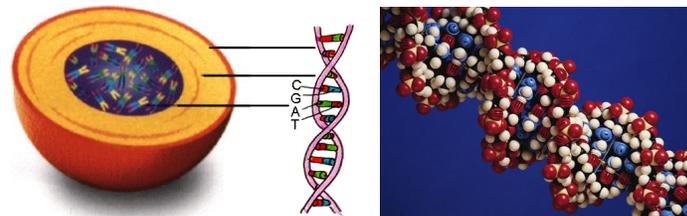


Figure A-1 : organisation de la molécule d'ADN

De l'ADN au génome

L'unité de base du stockage de l'information génétique est le gène, petite séquence d'ADN utilisée par la cellule pour synthétiser les protéines de structure, des enzymes nécessaires à son fonctionnement (respiration, division cellulaire, croissance, échanges, métabolismes...) ou des ARN (ARNt, ARNr et divers petits ARN). Les gènes sont de taille variable (de quelques centaines à plusieurs milliers de nucléotides), et sont disséminés le long du double brin d'ADN de chacune des cellules, souvent morcelés chez les organismes Eucaryotes en de nombreux fragments codants (exons), séparés par des séquences non codantes (introns). L'inventaire génique est alors défini comme l'ensemble de ces gènes, dont le nombre peut varier suivant l'individu, de 6200 chez la levure à une trentaine de milliers chez la souris ou

l'homme. Le génome est donc de façon très simplifiée, l'ensemble de la séquence des chromosomes comprenant les gènes, des séquences répétées et des séquences fonctionnelles telles que les télomères (extrémités des chromosomes eucaryotes) ou les origines de réplifications des chromosomes par exemple.

Du génome au transcriptome

En fonction de leurs besoins, les cellules utilisent à un instant donné une partie des gènes pour réaliser la synthèse des protéines nécessaires aux grandes fonctions cellulaires. Le passage du gène (contenu dans le noyau) à la protéine (synthétisée dans le cytoplasme) s'effectue en deux grandes étapes de transcription et de traduction, à l'aide d'un intermédiaire essentiel : l'ARN messager. La transcription correspond à la synthèse d'un brin d'ARN à partir de la double hélice d'ADN. L'ARN est synthétisé à partir du brin matrice (non codant) et a donc la même séquence que le brin non matrice (codant), la seule différence résidant dans le remplacement de la base T par la base U (uracile). Chez les Eucaryotes, l'ARN obtenu dans le noyau est appelé ARN pré-messager ou précurseur car il subira des maturations avant son transfert vers le cytoplasme (passage de la membrane nucléaire) où il sera traduit en protéines par les ribosomes. La traduction interprète chaque triplet de nucléotides (codon) de l'ARN messager en un acide aminé selon le code génétique universel : un codon donné correspond à un acide aminé donné, et plusieurs codons peuvent coder pour le même acide aminé (dégénérescence du code génétique). Ainsi, lors de la traduction, les ribosomes décodent l'information génétique pour permettre la synthèse de la protéine correspondante (Figure A-2). Au niveau des ribosomes, il existe une reconnaissance codon anti-codon entre la séquence de l'ARNm (codon) et la séquence anti-codon de l'ARNt (ARN de transfert) qui apporte l'acide aminé correspondant. Les ribosomes se déplacent sur l'ARNm pour permettre la synthèse complète de la protéine. La rencontre d'un codon non sens (aucun ARNt correspondant donc pas d'acide aminé) constitue le signal de fin de la traduction.

Signalons simplement que chez les Procaryotes, les gènes sont organisés en opérons (un ARNm porte plusieurs gènes), qu'il n'existe pas les mécanismes de maturation (absence d'introns) et enfin que la transcription et la traduction sont couplées du fait de l'absence de membrane nucléaire.

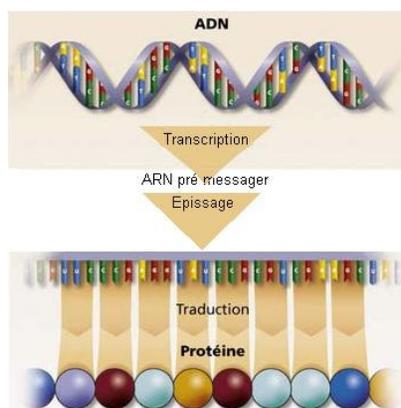


Figure A-2 : du gène à la protéine

Sachant qu'à partir d'un même gène, plusieurs copies d'ARN messagers peuvent être produits à des niveaux différents en fonction de l'activité de la cellule, le transcriptome reflétera donc le niveau d'expression de tous les gènes à un temps t pour une condition physiologique donnée. Il est le reflet instantané de l'activité cellulaire, et peut donc varier d'un type cellulaire à l'autre (neurone, cellule de la peau...), au cours du temps ou des différentes phases du cycle cellulaire, en fonction de conditions environnementales imposées aux

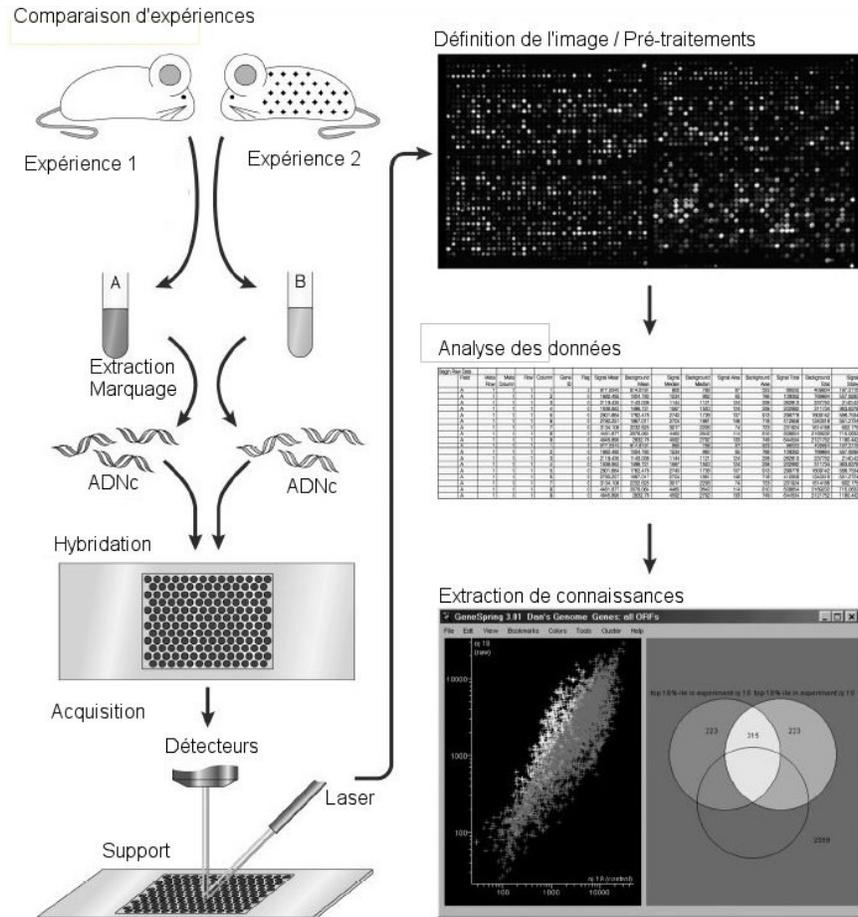
cellules, ou en fonction de l'état sain ou pathologique de ces dernières. Si Jansen *et al.* [JANSEN01] ont montré que l'analyse du transcriptome donnait une assez bonne vision du jeu de protéines présent dans la cellule, il est cependant important de noter qu'il n'existe pas nécessairement de corrélation directe entre le niveau d'expression d'un gène (quantité d'ARNm) et la quantité de protéines produite. D'une manière plus générale, pouvoir comparer le transcriptome de différents types cellulaires, dans différentes conditions, ou pouvoir analyser l'ensemble du transcriptome d'une cellule à divers stades de son cycle cellulaire ou dans des conditions pathologiques, doit permettre d'une part de mieux comprendre le fonctionnement cellulaire sur le plan fondamental, et offre d'autre part beaucoup d'intérêts en termes d'applications potentielles.

L'analyse de l'abondance de chaque ARNm d'une cellule, d'un type tissulaire ou d'un organisme est depuis quelques années possible à grande échelle grâce à des techniques d'hybridation. Cette mesure repose sur la propriété d'hybridation de l'ADN simple brin à son brin complémentaire de façon très stringente dans des conditions très contrôlées, et ce même si ce brin complémentaire n'est présent qu'en petites quantités au milieu d'autres fragments d'ADN. Depuis sa première apparition en 1975 (Southern Blot [SOUTHERN75]), la technique d'hybridation ADN-ADN généralement utilisée pour la cartographie des gènes a été adaptée à l'étude des ARN (Northern Blot). Dans la technique du Northern Blot, ce sont les ARN extraits qui après séparation sur un gel d'agarose dénaturant sont transférés sur une membrane (nitrocellulose ou nylon). Par la suite, une sonde marquée spécifique du gène étudié permettra par hybridation complémentaire (ADN-ARN ou ARN-ARN) de caractériser le messager correspondant (taille, niveau d'expression). Ainsi, dans cette technique les cibles se trouvent sur le support solide et les sondes sont généralement marquées par de la radioactivité pour permettre la révélation de l'hybridation. Ces dernières années, cette technologie qui ne permettait l'étude de l'expression que d'un gène à la fois a évolué vers le développement d'un outil de post-génomique très puissant permettant de suivre de façon simultanée l'expression de tous les gènes d'une cellule ou d'un organisme donné. Il s'agit de la technologie très innovante des puces à ADN.

2- Les puces à ADN

Principe

Les puces à ADN [LANDER99] reposent sur le principe de complémentarité des brins de la double hélice d'ADN. Ce sont des supports de verre ou de silicium, de petite taille (de l'ordre de 25×75 mm) sur lesquels sont synthétisés directement ou greffés après synthèse (plus économique) des milliers de séquences d'ADN appelées sondes, caractéristiques d'autant de gènes. Les puces sont ensuite mises en contact avec l'ensemble des transcrits issus d'une cellule, les cibles, marqués par une molécule fluorescente (fluorochrome). Le mécanisme d'hybridation implique que les cibles marquées reconnaissent, parmi les sondes de la puce, celles qui leur sont complémentaires, et s'y appariant. Après cette étape, les hybrides sondes/cibles sont imagés *via* un scanner, repérés et quantifiés grâce à leur fluorescence. En déposant un nombre important de sondes, on peut ainsi analyser un grand nombre de gènes simultanément. On peut même, en utilisant deux fluorochromes tels que les cyanines (*e.g.* un rouge Cy5 et un vert Cy3) comparer les niveaux d'expression relatifs de deux transcriptomes différents sur une seule puce (sain/pathologique, début/fin de cycle cellulaire,...) : pour chaque fluorochrome, deux images distinctes sont obtenues et superposées *in silico* pour analyser le différentiel d'expression (Figure A-3).



distance entre deux dépôts consécutifs sur la lame est de l'ordre de 250 μm dans les deux directions.

Concernant les biopuces oligonucléotides, la synthèse des sondes est chimique ce qui évite les premières étapes laborieuses de PCR et de contrôles rencontrées lors de la conception des biopuces ADNc. Ces oligonucléotides joueront donc le rôle « d'étiquettes » spécifiques identifiant les différents gènes.

Les puces Affymetrix (<http://www.affymetrix.com>) très haute densité utilisent une technologie permettant une synthèse directe, *in situ*, des oligonucléotides sur le support solide (semi-conducteurs, masquage et photolithographie). Ainsi, 65536 oligomères peuvent être obtenus sur 1,6 cm^2 en 32 étapes.

Pour les autres biopuces oligonucléotides, la synthèse se fait avant le dépôt sur le support solide. Plusieurs modes de dépôt peuvent être rencontrés, soit par contact direct (aiguille creuse ou pin and ring) soit par la technique « jet d'encre » (biopuces Agilent).

La préparation des cibles et l'hybridation

Des milliers de transcrits différents sont présents dans les cellules à un moment donné, et leur abondance relative est révélatrice de l'activité cellulaire à cet instant. La préparation des cibles consiste tout d'abord à extraire du milieu cellulaire ces transcrits puis leur intégrer un marqueur fluorescent (rouge ou vert) qui permettra d'évaluer et de quantifier l'appariement sonde/cible. Le marquage, soit direct (incorporation d'un nucléotide fluorescent), soit indirect (incorporation d'un nucléotide modifié amino allyl dUTP puis couplage du fluorochrome dans un second temps), se fait suite à la transcription inverse des ARNm permettant l'obtention d'un brin d'ADNc fluorescent. Du fait de la complémentarité des nucléotides, le dépôt des cibles marquées sur la puce déclenche l'appariement des séquences sondes/cibles complémentaires. Cette hybridation, qui dure quelques heures en milieu liquide, est suivie d'un lavage du substrat qui permet d'éliminer les cibles non fixées, ou fixées non spécifiquement. Après séchage la puce est passée au scanner pour repérer les hybridations.

La chaîne d'acquisition de l'image de la puce

La puce est alors révélée par un lecteur (scanner) muni de lasers qui permettent d'exciter les molécules de fluorochromes et de détecter par un microscope confocal le signal émis dans chaque spot. Dans le cas du marquage avec deux fluorochromes (vert Cy3 et rouge Cy5), une image numérique est acquise pour l'échantillon marqué avec le Cy3 et une en Cy5. Un spot de couleur verte indique un gène dont le niveau d'expression est plus élevé dans l'échantillon marqué avec le Cy3 que celui marqué avec le Cy5, et inversement pour un spot de couleur rouge. Le spot apparaît jaune lorsque le gène est exprimé de manière identique dans les deux échantillons comparés. L'analyse des données numériques issues de l'acquisition est effectuée par un logiciel qui prétraite, segmente et quantifie les différents niveaux d'activité dans les spots.

3- L'analyse des images de puces

L'analyse d'image est un aspect central des expériences menées à l'aide des puces à ADN. Dans le cas d'un double marquage, le but de cette phase est de quantifier, de manière relative, le niveau d'expression des gènes. Cette mesure, basée sur un rapport d'intensité entre les deux niveaux des fluorochromes détectés, est fonction de nombreux paramètres dépendant des méthodes utilisées, des mesures expérimentales et des conditions biologiques. Globalement, cette étape d'analyse a un impact considérable sur l'interprétation biologique des données, et repose sur quatre phases (Figure A-4).

La localisation des spots

Il s'agit ici de déterminer les coordonnées de chaque spot de la puce. Cette étape est normalement effectuée à l'aide d'une grille théorique définie lors du plan de dépôt des sondes. Pour localiser un spot sur une image, c'est-à-dire faire correspondre un modèle idéal de puce avec une image acquise, un nombre important de paramètres doit être estimé (espaces entre les spots, espaces entre blocs d'une puce...). De même, les mouvements de translation des spots ou des blocs de spots, liés à la variation de position des aiguilles du spotter, sont à évaluer. Le repérage des spots doit être simple et rapide à réaliser. L'automatisation de cette tâche permet une accélération considérable de l'analyse. Enfin, la définition des grilles doit être aussi juste que possible. En effet, de la précision de cette étape dépend l'efficacité des mesures ultérieures.

La segmentation des spots

Il s'agit ici de classer les pixels de l'image en deux classes, "fond" et "signal". Ceci sous-tend une analyse du signal au niveau de chaque spot et un découpage de l'image en différentes régions, chacune ayant des propriétés propres. En raison des propriétés biologiques des dépôts et de la physique de la chaîne d'acquisition, le bruit acquis est fortement non uniforme dans l'image et varie en fonction d'un certain nombre de paramètres. De plus, à ce stade de l'analyse, la taille variable, la forme complexe et les irrégularités des spots compliquent la tâche de segmentation, mais doivent cependant être déterminés de manière semi-automatique voire automatique.

L'extraction d'indices

Maintenant que les pixels "signal" sont identifiés, il s'agit d'extraire un ou plusieurs indices pertinents permettant de quantifier le niveau d'expression de chaque gène. Ce niveau correspond à une mesure relative des intensités de fluorescence en rouge et vert. Un filtrage des données permet de sélectionner des variations significatives (tri des spots sur un barème de critères), à partir desquelles un rapport de fluorescences est calculé. La valeur de ce ratio indique l'induction ou la répression du gène.

Dans le cas de la comparaison de données issues de plusieurs expériences (données temporelles, sain/pathologique...), il faut de plus nécessairement normaliser les données avant quantification pour éliminer les artefacts dûs par exemple au protocole expérimental.

La classification des données

Enfin, les données préparées par les étapes précédentes permettent de regrouper les gènes par familles ayant des comportements semblables, en analysant leurs niveaux d'expression. Cette étape de classification permet de répondre à des problématiques biologiques importantes allant de la recherche de gènes corégulés au cours du développement ou en réponse à des perturbations environnementales, à l'identification des liens entre des variations d'expression et d'autres données biologiques en passant par la recherche automatique de systèmes de régulation.

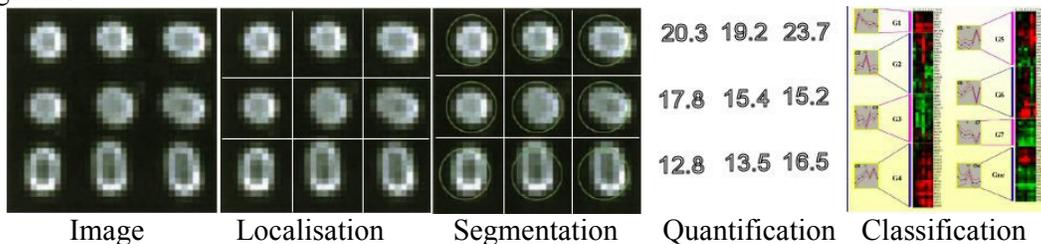


Figure A-4 : schéma récapitulatif des étapes de traitement des images de puces à ADN

4- Les applications

L'utilisation des outils d'analyse des puces à ADN est aujourd'hui de plus en plus intensive (environ 12500 articles parus en 2002 sur l'étude du transcriptome [Scheel02]), et ceci dans de nombreux domaines (Figure A-5), parmi lesquels :

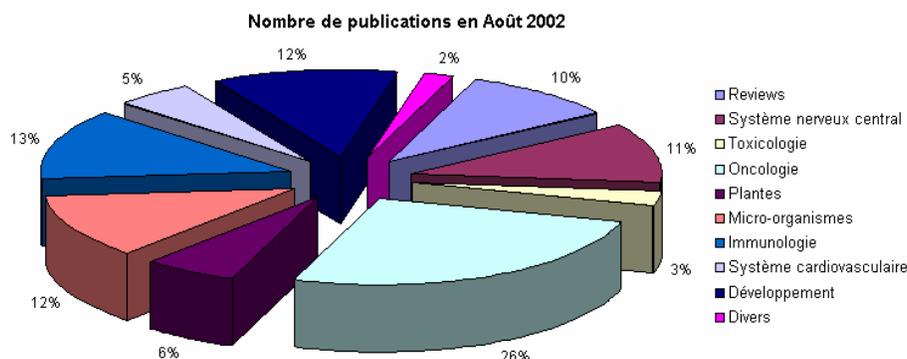


Figure A-5 : domaines de publication des études sur les puces à ADN.

Etude de la dynamique du transcriptome

Bien sûr, un des premiers rôles des puces à ADN a été d'étudier l'expression de milliers de gènes en parallèle, souvent selon plusieurs conditions. De très nombreuses études ont été réalisées sur divers organismes (*e.g.* la levure *Saccharomyces cerevisiae* [SPELLMAN98], *Streptomyces coelicolor* [Huang01] dont dérivent 60% des antibiotiques actuellement utilisés), en fonction de diverses conditions expérimentales (cycle cellulaire, influence du milieu extérieur,...), pour mettre en évidence des gènes co-régulés, des liens entre des variations d'expression et d'autres données biologiques ou des systèmes de régulation.

Nous collaborons actuellement avec une équipe de biologistes de l'Université Blaise Pascal de Clermont-Ferrand (laboratoire de biologie des protistes, équipe Génomique Intégrée des Interactions Microbiennes, Pierre Peyret, UMR 6023 CNRS) sur une telle étude, concernant le pathogène *Encephalitozoon cuniculi* [KATINKA01].

Recherche et action de médicaments

Les puces ADN sont un outil de choix dans la recherche et la caractérisation de nouvelles molécules à visée thérapeutique [DEBOUK99]. En effet, la plupart des médicaments agissent en inhibant leur molécule cible, et une mutation du gène correspondant devrait donc avoir un effet similaire sur le transcriptome de la cellule. La société Affymetrix propose ainsi depuis fin 1997 une puce de 800 oligonucléotides permettant la détection des 18 mutations et des 10 polymorphismes répertoriés dans deux gènes associés à un système enzymatique (cytochrome P450) qui joue un rôle essentiel dans la cinétique de dégradation par le foie de nombreux médicaments. De même, Marton *et al.* [MARTON98] 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.

Dans ce domaine, les perspectives à moyen terme vont du diagnostic prédictif à la thérapie personnalisée, comme précisé dans un rapport récent [PROVENCE02] de la Direction générale de l'Industrie, des Technologies de l'information et des Postes sur la société de l'information (Figure A-6)

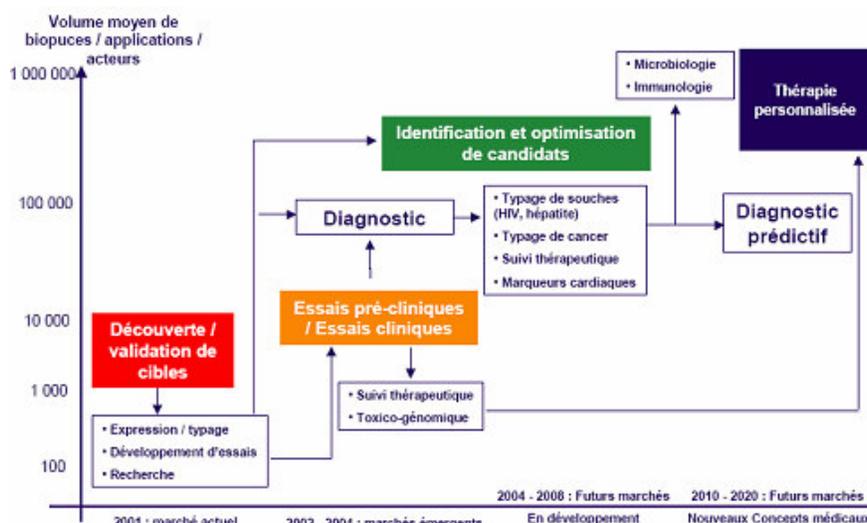


Figure A-6 : évolution des puces à ADN dans les applications pharmaceutiques.

Recherche de mutations et de remaniements chromosomiques

Pour chaque maladie héréditaire, la palette des mutations possibles est en général très étendue, ce qui rend la recherche de la mutation responsable lors du diagnostic très complexe. De nombreuses techniques de criblage ou de caractérisation de mutations connues ont été développées, mais restent aujourd'hui peu performantes. Les techniques de criblage permettent de rechercher une variation de séquence, caractérisée ensuite par séquençage. Une recherche directe de la mutation est aussi possible par séquençage, mais son coût est encore très élevé. La recherche des mutations est donc par excellence un champ d'activité où les puces apportent une révolution [CRONIN96]. De même, certains auteurs proposent de dédier des puces à ADN à l'étude de remaniements chromosomiques (*e.g.* aneuploïdie), par exemple pour classer différents types de rhabdomyosarcomes (plus fréquente des tumeurs des tissus mous en pédiatrie, qui représente 10% des tumeurs solides chez l'enfant) [LU01].

Oncologie

L'exemple précédent illustre un des principaux intérêts des puces à ADN pour la classification des tumeurs, le diagnostic et le pronostic en cancérologie. Des travaux ont plus généralement mis à jour une signature d'expression fortement associée au risque d'évolution métastatique dans le cancer du sein [Van't VEER02], en étudiant le profil d'expression de tumeurs primaires de ce cancer. Cette même pathologie a de même été étudiée par Mei *et al.* [MEI00], qui proposent de détecter à l'aide des puces à ADN haute densité des allèles dangereux du gène de susceptibilité au cancer du sein BRCA1.

Environnement et agriculture

Dans une chaîne de production agroalimentaire, les puces à ADN sont des outils particulièrement efficaces pour reconnaître l'origine des espèces animales qui composent les produits alimentaires, mais également pour détecter et contrôler dans des semences des séquences provenant d'organismes génétiquement modifiés [KUIPER01]. Les applications dans l'environnement concernent quant à elles l'analyse microbiologique de l'eau de consommation,

et plus généralement le contrôle qualité en milieu industriel, pour renforcer la sécurité des procédés et des produits. Ainsi, la Lyonnaise des Eaux et bioMérieux ont par exemple décidé d'unir leurs compétences pour mettre au point, dans le cadre du projet Aquagen, une technique d'analyse de l'eau potable par puces à ADN, qui apporte aux consommateurs une garantie renforcée en matière de contrôle de la qualité de l'eau.

Bibliographie

- [CRONIN96] CRONIN M, MIYADA C, FUCINI R, KIM S, MASINO R, WESPI R., Detecting cystic-fibrosis mutations by hybridization to DNA-probe, *Biologicals*, **24**:209, 1996.
- [DEBOUK99] DEBOUK C., GOODFELLOW P., DNA microarrays in drug discovery and development, *Nature Genetics*, **1**: 48–50, 1999.
- [HUANG01] HUANG J., LIH C., PAN K., COHEN S., Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays, *Genes Development*, **15**:3183-3192, 2001.
- [JANSEN01] JANSEN R., GREENBAUM D., GERSTEIN M., Relating whole-genome expression data with protein-protein interactions, *Genome Research*, **12**:37-46, 2002.
- [KATINKA01] KATINKA M., DUPRAT S., CORNILLOT E., MÉTÉNIER G., THOMARAT F., PRENSIER G., BARBE V., PEYRETAILLADE E., BROTTIER P., WINCKER P., DELBAC F., EL ALAOU H., PEYRET P., SAURIN W., GOUY M., WEISSENBACH J., VIVARÈS C., Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, **414**:450-453, 2001.
- [KUIPER01] KUIPER H., KLETER G., NOTEBORN H., KOK E., Assessment of the food safety issues related to genetically modified foods, *Plant Journal*, **27**:503-528, 2001.
- [LANDER99] LANDER E., Array of hope, *Nature Genetics*, **21**:3-4, 1999.
- [LU01] LU Y., WILLIAMSON D., CLARK J., WANG R., TIFFIN N., SKELTON L., GORDON T., WILLIAMS R., ALLAN B., JACKMAN A., COOPER C., PRITCHARD-JONES K., SHIPLEY J., Comparative expressed sequence hybridization to chromosomes for tumor classification and identification of genomic regions of differential gene expression, *Proceedings of National Academy of Sciences*, **98**:9197-9202, 2001.
- [MARTON98] MARTON M., DERISI J., BENNETT H., IYER V., MEYER M., ROBERTS C., STOUGHTON R., BURCHARD J, SLADE D, DAI H, BASSETT D. JR, HARTWELL L., BROWN P., FRIEND S., Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Medicine*, **4**:1293-301, 1998.
- [MEI00] MEI R., GALIPEAU P., PRASS C., BERNO A., GHANDOUR G., PATIL N., WOLFF R., CHEE M., REID B., LOCKHART D., Genome-wide detection of allelic imbalance using human SNPs and high density DNA arrays, *Genome Research*, **10**:1126-1137, 2000.
- [PROVENCE02] PROVENCE M., Etude biocapteurs et biopuces, *Rapport de la Direction générale de l'Industrie, des Technologies de l'information et des Postes sur la société de l'information*, 2002.
- [SCHEEL02] SCHEEL J, VON BREVERN MC, HORLEIN A, FISCHER A, SCHNEIDER A, BACH, A. Yellow pages to the transcriptome, *Pharmacogenomics*, **3**:791-807, 2002.

- [SOUTHERN75] SOUTHERN E., Detection of specific sequences among DNA fragments separated by gel electrophoresis, *Journal of Molecular Biology*, **98**:503-517, 1975.
- [SPELLMAN98] SPELLMAN P., SHERLOCK G., ZHANG M., IYER V., EISEN M., BROWN P., BOTSTEIN D., FUTCHER B., Comprehensive identification of Cell-Cycle-regulated genes of the Yeast *Saccharomyces cerevisiae* by microarray hybridization, *Molecular Biology of the Cell*, **9**:3273-3297, 1998.
- [VAN'T VEER02] VAN'T VEER, L., Gene expression profiling predicts clinical outcome of breast cancer, *Nature*, **415**: 530-536, 2002.