
CHAPITRE V. EXPRESSION DE GÈNES ET MODÉLISATION

Analyse de puces à ADN (microarray)

1. Introduction

Loin de correspondre à une technique unique, les puces à ADN regroupent un ensemble très divers de méthodes et de technologies. L'objectif de cette partie n'est donc pas de présenter une liste exhaustive des méthodologies qui sont en constante évolution, mais plutôt de mettre en évidence les étapes essentielles de la conception et de l'utilisation des puces à ADN.

Leur description concerne principalement l'analyse du transcriptome, qui est au centre de notre problématique. De plus, compte tenu de l'immense quantité d'informations disponibles, certains aspects ont été volontairement limités, en raison de notre problématique biologique, aux particularités concernant les modèles bactériens.

Les nombreuses étapes de la technologie des puces à ADN sont à l'origine d'une variabilité expérimentale qui peut affecter la qualité des résultats (Nadon et Shoemaker, 2002) et qu'il est essentiel de prendre en compte (Leung et Cavalieri, 2003). Ces sources de variabilité sont donc également décrites et discutées au sein des différents paragraphes.

2.1.1 Historique

Le développement des puces à ADN et à gènes s'inscrit dans une évolution technologique marquée par plusieurs étapes clés :

1. 1974 - Hybridation moléculaire

Le concept fondamental des puces à ADN repose sur le principe d'hybridation décrit par Edwin Southern. Ce mécanisme stipule que deux fragments d'acides nucléiques complémentaires peuvent s'associer ou se dissocier de manière réversible, selon les conditions de température et de salinité du milieu.

2. 1977 - Northern blotting

Alwine et ses collaborateurs introduisent la méthode de northern blotting, permettant l'étude de l'expression des gènes. Cette méthode est un précurseur des technologies actuelles de détection génétique.

3. 1987 - Concept des puces à ADN

L'idée des puces à ADN voit le jour, bien avant leur première application concrète. Ce concept repose sur l'utilisation d'une approche miniaturisée pour analyser les interactions entre acides nucléiques.

4. 1991 - Reverse dot blot

Lennon et Lehrach introduisent le système de reverse dot blot, qui propose une configuration simplifiée pour détecter des séquences nucléotidiques spécifiques, préparant ainsi le terrain pour la miniaturisation des technologies génomiques.

5. 1992 - Expression différentielle

Liang et Pardee développent une méthode d'étude de l'expression différentielle des gènes, ouvrant de nouvelles perspectives pour analyser des variations d'expression dans différents contextes biologiques.

6. 1993 - Premières puces à ADN

Les premières puces à ADN font leur apparition. Elles combinent les avancées précédentes et exploitent une approche pluridisciplinaire intégrant des technologies électroniques, chimiques, et informatiques.

7. Après 2001 - Explosion des publications

Avec l'évolution rapide des technologies et l'intérêt croissant pour la génomique, les publications consacrées aux puces à ADN connaissent une augmentation exponentielle, consolidant leur rôle dans les domaines de la recherche et du diagnostic.

Ainsi, les puces à ADN, bien que basées sur des concepts établis dès les années 1970, sont le fruit d'une innovation continue et interdisciplinaire qui a transformé les approches en biologie moléculaire et en génomique.

Le marché mondial des puces à ADN et à gènes connaît une croissance rapide et significative, avec des taux de développement remarquables ces dernières années. Cette tendance devrait se poursuivre de manière substantielle sur la période de prévision, de 2022 à 2030 (Figure 1).

Global Dna Gene Chip Market Size and Scope

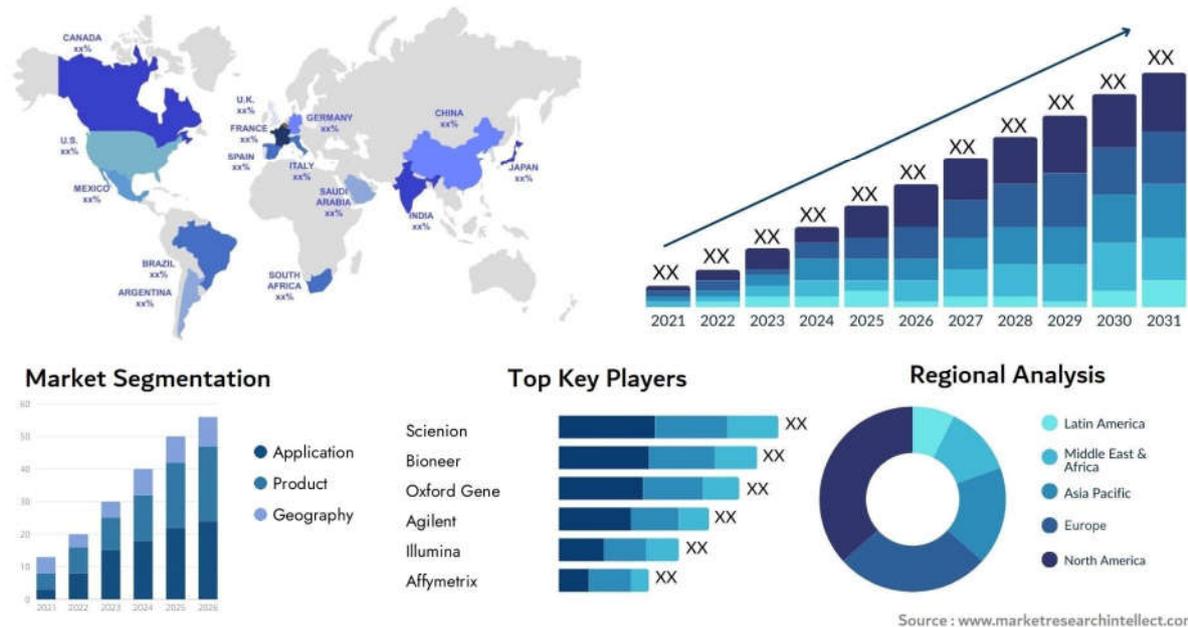


Figure 1 : Taille du marché des puces à ADN et à gènes

Historiquement les macroarrays, les microarrays et les « véritables » puces à ADN correspondent à trois méthodes différentes d'analyse. Les macroarrays utilisaient des clones d'ADN complémentaire (ADNc) disposés sur des membranes de nylon (avec un espacement de l'ordre du millimètre) en association avec des cibles radioactives. Les microarrays, plus miniaturisés, comportaient quelques milliers de gènes représentés par des produits PCR déposés tous les 200 à 400 microns sur une lame de verre et des cibles marquées par fluorescence. Enfin, les « véritables » puces à ADN associaient à chacun des gènes d'un organisme un ensemble d'oligonucléotides synthétisés in situ. La première de ces puces à ADN s'appelait la « Gene Chip™ HIV PRT ». Commercialisée en 1998 par Affymetrix (Santa Clara, CA, USA), elle avait été conçue pour l'analyse des mutations de la transcriptase inverse et de la protéase du virus HIV. La même année a vu le développement de la première puce à oligonucléotides dédiée à une bactérie, contenant un sous-ensemble de cent gènes de *Streptococcus pneumoniae*.

Aujourd'hui ces trois distinctions n'ont plus vraiment lieu d'être, d'autant plus que ces techniques sont utilisées de façon croisée, comme le montre l'exemple de puces à ADN utilisant des produits PCR et des cibles radioactives. Les terminologies « puce à ADN » et « microarray » sont donc employées de façon indifférente. Les termes de « biopuce » ou « microréseau » sont également employés dans la littérature française.

3. Principe des puces à ADN et analyse du transcriptome

L'idée conceptuelle de la puce à ADN est très simple. Il s'agit de greffer sur une surface de quelques centimètres carrés des fragments synthétiques d'ADN (les sondes) espacés de quelques micromètres et représentatifs de chacun des gènes étudiés. Ce micro dispositif est ensuite mis au contact des acides nucléiques à analyser, au cours de l'étape d'hybridation. Ces acides nucléiques, appelés cibles, correspondent aux ARNm ou aux ADNc qui ont été préalablement couplés à un marqueur fluorescent ou radioactif. Ce contact entre cibles et sondes conduit à la formation d'hybrides qualifiés par leurs coordonnées, et quantifiés grâce à la lecture des signaux radioactifs ou fluorescents. En ce qui concerne la terminologie associée à la technique des puces à ADN, il est important de rappeler que les puces ne sont qu'un northern blotting inversé, où la sonde est fixe alors que la cible marquée est en solution. Cependant cette différence est à l'origine d'une confusion entre les termes « cible » (probe) et « sonde » (target). Une nomenclature a donc été recommandée et semble aujourd'hui bien respectée. Les sondes correspondent aux acides nucléiques fixés sur la puce, alors que les cibles représentent l'ensemble des acides nucléiques libres étudiés dipeptide.

La répétition d'une telle réaction à l'une ou l'autre extrémité du dipeptide produira une chaîne d'acides aminés qui, finalement, constituera la protéine (figure 2). La séquence des acides aminés formant la protéine s'appelle la structure primaire. La nature de la chaîne latérale influence la structure tridimensionnelle finale de la protéine. En effet, les chaînes latérales des acides aminés non-polaires sont hydrophobes et ont tendance à se regrouper en un cœur hydrophobe au sein de la protéine et à laisser les acides aminés hydrophiles à l'extérieur de celle-ci où ils sont accessibles au solvant. Ceci est illustré à la figure 3.

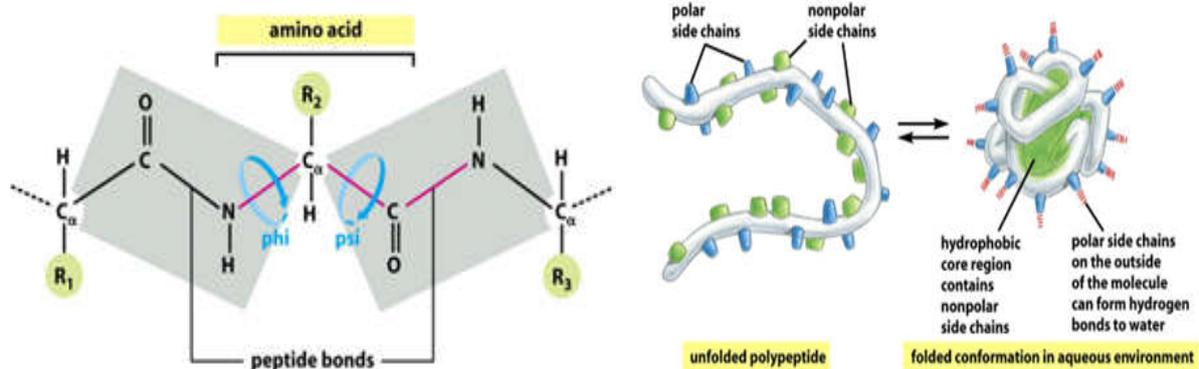


Figure 2. Liens peptides entre acides aminés Figure 3. Cœur hydrophobe d'une protéine

4. Structures

La structure primaire d'une protéine, c'est-à-dire sa séquence en acides aminés, détermine sa structure tertiaire. Elle contient donc toute l'information nécessaire au repliement de celle-ci. Le deuxième niveau de structure est appelé structure secondaire et définit les différentes structures tridimensionnelles locales que l'on peut observer sur des sous-ensembles d'acides aminés de la structure primaire. Parmi les différentes structures secondaires, les plus connues sont les hélices α et les brins β . Ces deux structures secondaires résultent de liaisons hydrogène entre un atome d'oxygène du groupe carboxyle d'un acide aminé et un atome d'hydrogène du groupe amine d'un autre acide aminé de la chaîne principale. Ces structures ne font intervenir que les atomes de la chaîne principale, ce qui explique leur grande fréquence d'observation.

L'analyse du transcriptome nécessite de mesurer les niveaux d'expression des gènes. Cette mesure peut être réalisée par une évaluation absolue de l'expression des gènes. Ce type de mesure est réservé aux puces commercialisées par la société Affymetrix qui sont conçues de façon un peu particulière. Pour chaque gène, une série de dix à vingt sondes, réparties sur toute la séquence du gène, est représentée sur la lame. À chacune de ces sondes PM (Perfect Match) est associée une sonde MM (MisMatch) dont la séquence est identique à la séquence des sondes PM, excepté une mutation ponctuelle située en position centrale. Cette sonde MM permet de quantifier la part du signal aspécifique (bruit de fond) associé à la sonde PM. Le calcul du niveau d'expression d'un gène est relativement complexe, mais peut être considéré, en première approximation, comme une moyenne pondérée des différences (PMMM) de chaque paire de sondes associées à ce gène.

Cependant, pour l'ensemble des puces qui sont classiquement utilisées l'étude du niveau d'expression des gènes est basée sur la détermination des variations de niveau d'expression d'un organisme dans deux conditions différentes. Il s'agit donc d'une mesure différentielle de l'expression des gènes. En pratique, pour réaliser cette mesure, les cibles constituant deux échantillons d'étude sont marquées au moyen de deux fluorochromes différents. Elles sont ensuite mélangées et hybridées sur la puce (Figure 4). Après hybridation, la mesure des intensités des signaux de fluorescence relatifs à chacun des fluorochromes permet finalement de calculer pour chaque gène le rapport des intensités, qui évalue son expression différentielle. Ces deux types de mesures ne sont cependant pas complètement disjoints. En effet, contrairement à ce que laisse penser la description de ce principe, les cibles

marquées avec deux fluorochromes différents ne sont pas en compétition pour se lier sur une même sonde car les sondes sont toujours en large excès par rapport à la quantité de cibles qui est déposée. Une étude expérimentale montre même que l'intensité du signal obtenu pour un fluorochrome n'est pas affectée par la présence d'une cible marquée avec l'autre fluorochrome. Cette étude valide donc la possibilité de travailler sur les signaux aussi bien que sur les rapports des intensités et autorise la comparaison d'échantillons qui ne sont pas hybridés sur une même lame. Basées sur ce principe, de nombreuses puces ont été développées avec des jeux de sondes spécifiques pour étudier le transcriptome d'organismes divers. Actuellement, des chercheurs essayent de concevoir une puce universelle contenant des sondes qui représentent l'ensemble des combinaisons possibles de séquences d'ADN. Le problème majeur est de définir la taille optimale des sondes. Un modèle a été réalisé pour concevoir une puce dédiée à la fois à la Souris et à la levure. Une première puce universelle appelée UMAS (Universal Micro-Array System) a même vu le jour. Elle contient pour sondes toutes les combinaisons possibles d'hexamères et son utilisation combinée avec une étape enzymatique de fractionnement des cibles permettrait de générer des profils d'expression pour n'importe quel organisme.

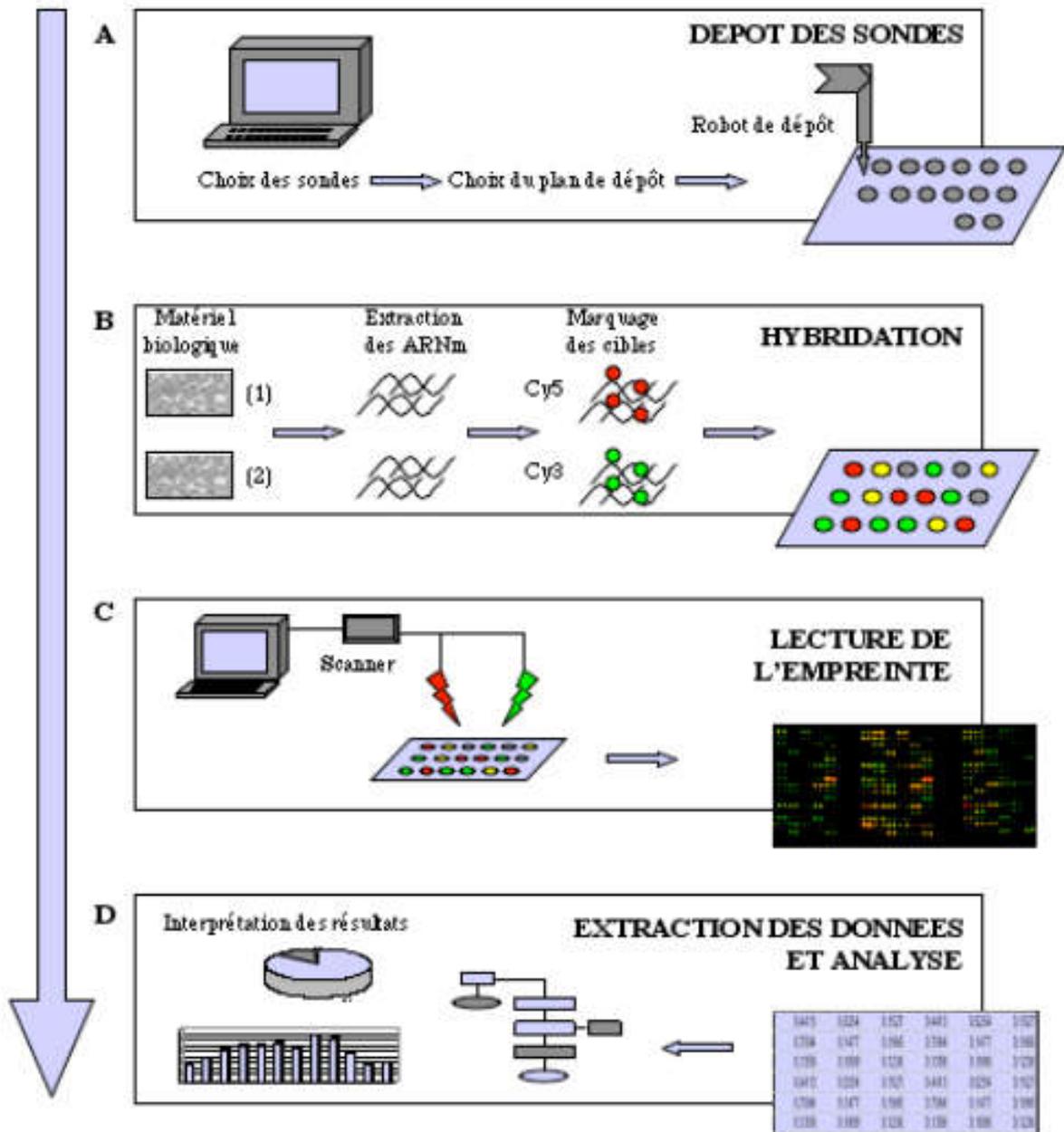


Figure4 : Principe de la technologie des à puces à ADN

(A) Les séquences des sondes sont déterminées de façon à optimiser leur spécificité et leur sensibilité. Les sondes synthétisées sont déposées par un robot sur la surface de la lame selon un plan défini. (B) Les ARNm sont extraits des échantillons biologiques à comparer, marqués avec deux fluorochromes différents puis mélangés avant hybridation. (C) La lecture des lames est réalisée avec un scanner (microscope à fluorescence) couplé à un photomultiplicateur (PMT). (D) L'image est alors analysée de façon à quantifier le signal. Les données sont ensuite normalisées, analysées et interprétées.