

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي



جامعة محمد الصادق بن لبيح
كلية علوم الطبيعة و الحياة
المكتبة 1371
رقم الجرد :

01
02
Bl. 04/09

Université de Jijel
Faculté des Sciences Exactes et Sciences de la Nature
Département de Biologie Moléculaire et Cellulaire

Mémoire

De Fin d'Etudes en Vue de l'Obtention du Diplôme des Etudes
Supérieures en Biologie
Option : Biochimie
Thème :

**La transcriptomique : Une approche systématique
pour l'étude de l'expression des gènes**

Membres de Jury :

- Examineur : Dr SIFOUR Mohammed
- Encadreur : Dr RECHRECHE Hocine

Réalisé par :

- BOUHBOR Khayreddine
- BOUKERRA Saad
- MAOUCHI Nassima



Promotion 2009

Remerciement

Nous tenons tout d'abord à exprimer nos remerciements à Dieu, le tout puissant et maître de l'univers qui nous a donné la capacité nécessaire, la forte volonté et la patience afin d'accomplir ce travail, et qui nous a toujours guider vers le bon chemin.

Nous aimerons par ailleurs remercier tous ceux qui nous ont aidé de près ou de loin pour réaliser ce travail, en particulier notre encadreur Dr RECHRECHE Hocine qui nous a proposé ce sujet de recherche et qui nous a encadré par ces conseils et ses efforts durant la préparation de notre mémoire.

Nous remercions aussi notre examinateur Dr SIFOUR, Mohammed d'avoir accepté et de juger le contenu du présent mémoire. Nous plus vifs remerciements et toutes nos reconnaissances vont à tous les enseignants du département de Biochimie de l'Université de Jijel.

Nous ne serions bien sur jamais arrivés là sans l'aide et le soutien de nos familles. Merci à nos parents pour avoir toujours cru en nous, merci de nous avoir soutenue dans cette voie, merci de votre présence, de vos encouragements, de vos conseils, de vos attentions constantes. Merci pour tout, nous espérons de vous rendre le bonheur que vous nous apportez.

Khayreddinne, Nassima et Saad.

Table des matières

Introduction.....	01
Chapitre I. Expression des gènes et transcriptomique.....	03
I.1. Techniques classiques de l'étude de l'expression des gènes.....	03
I.1.1. Northern blot et dot blot.....	03
I.1.2. Criblage différentiel.....	03
I.2. Etude du transcriptome.....	03
I.2.1. Historique.....	03
I.2.2. Techniques à la base de l'émergence de la transcriptomique.....	04
I.2.2.1. Séquençage d'ADN.....	04
I.2.2.2. Bioinformatique.....	05
I.2.2.3. Techniques de PCR.....	06
I.2.3. Développement de l'étude transcriptomique.....	07
Chapitre II. Techniques de l'analyse du transcriptome.....	08
II.1. Introduction.....	08
II.2. Techniques basées sur la PCR.....	08
II.2.1. Display différentiel et AP-PCR.....	08
II.2.2. PCR soustractive.....	08
II.2.3. RT-PCR en temps réel.....	10
II.3. Technique de micro-alignements d'oligonucléotides.....	10
II.3.1. Historique des puces ADN.....	14
II.3.2. Principe des puces ADN.....	14
II.4. Techniques basées sur le séquençage d'ADN.....	15
II.4.1. Micro-alignements d'ADNc.....	17
II.4.1.1 Principe.....	17
II.4.1.2. Les EST, étiquettes de séquences exprimées.....	17
II.4.2. L'analyse en série de l'expression génique (SAGE).....	19
II.4.2.1. Principe.....	19
II.4.2.2. Développement de SAGE.....	21
II.4.3. La Technique de MPSS.....	21
Chapitre III. Quelques applications de l'analyse transcriptomique.....	24
III.1. Introduction.....	24
III.2. Applications dans le domaine de la santé.....	24
III.2.1. Diagnostic des cancers.....	24
III.2.1.1. Cancers du sein.....	24

III.2.1.2. Cancer pancréatique.....	26
III.2.2. Détection d'activité biologique d'extraits végétaux à visée cosmétique ou dermatologique	27
III.3. Applications dans le domaine de la recherche.....	27
III.3.1. Etude du transcriptome en neurobiologie.....	27
III.3.2. Etude transcriptomique chez le <i>Leishmania</i>	29
III.3.3. Mise en évidence de gènes différentiellement exprimés.....	31
III.3.4. Etablissement d'un profil caractéristique d'un état biologique donné.....	31
III.3.5. Caractérisation des voies de régulation	32
III.3.6. Etude du transcriptome de cellules souches embryonnaires (CSE) et hématopoïétiques (CSH).....	33
III.4. Applications dans le domaine agroalimentaire.....	34
III.4.1. Approche transcriptomique de l'engraissement chez le poulet de chair.....	34
III.4.2. Analyse du transcriptome de microtubercules de Pomme de terre.....	35
III.5. Applications dans le domaine de l'environnement.....	35
III.5.1. Identification des bactéries pathogènes des eaux : le projet Aquachip.....	35
Discussion et conclusion.....	37
Références bibliographiques.....	43

Liste des figures

Figure 1. Principe de la technique du Display différentiel.....	09
Figure 2. Principe de la technique de PCR soustractive (Représentational Difference Analysis « RDA » et de Suppressive Subtractive Hybridization « SSH »).....	11
Figure 3. Principe de base d'un système PCR en temps réel.....	12
Figure 4. Profil d'une courbe PCR en temps réel.....	13
Figure 5. Etapes de l'analyse du transcriptome par la technologie de micro-alignements d'oligonucléotides.....	16
Figure 6. Principe de la technique de micro-alignements d'ADNc.....	18
Figure 7. Principe de la technique SAGE.....	20
Figure 8. Principe de la technique MPSS.....	22
Figure 9. Etudes transcriptomiques chez le <i>Leshmania</i>	30

Listes des abréviations et symboles

ACP	Analyse en Composante Principale
ADN	Acide Désoxyribo Nucleique
ADNc	ADN complémentaire
AP-PCR	Arbitrarily Primed Polymerase Chain Reaction
ARN	Acide Ribo Nucleique
ARNm	ARN messenger
CAGE	Cap Analysis of Gene Expression
cDNA-AFLP	Complementary Desoxyribo Nucleic Acid Amplified Fragment Length Polymorphism
CNGs	Conserved Non-Genic sequences
CSE	Cellules Souches Embryonnaires
CSH	Cellules Souches Hématopoïétiques
CT	Threshold Cycle
Cy3	Cyanine 3
Cy5	Cyanine 5
dATP	Désoxyadénosine 5'-triphosphate
dCTP	Désoxycytosine 5'-triphosphate
DD	Différentiel Display
DFTP	Démence Frontotemporale et Parkinsonisme
d(T)	désoxy Thymine
EBI	Institut Européen de Bio-informatique
EST	Expressed Sequence Tag
HGP	Human Genome Project
HUGO	Human Genome Organization
HSS	Hybridation Soustractive Suppressive
Kb	Kilo base
MH	Maladie de Huntington
MP	Maladie de Parkinson
MM	Mis Match
MPSS	Massively Parallel Signature Sequencing
PCR	Polymerase Chain Reaction
PM	Perfect Match
PSP	Paralyse Supranucléaire Progressive
RT-PCR	Reverse Transcribed- Polymerase Chain Reaction
RE	Récepteurs aux Œstrogènes
SADE	SAGE Adaptation for Downsized Extracts
SAGE	Serial Analysis of Gene Expression
SAM	Significance Analysis of Microarray
SNc	Substance Noire compacte
SNP	Signle Nucleotide Polymorphismes
TIGR	The Institute for Genome Research
UMAS	Universal Micro-Array System

INTRODUCTION

La Biologie Moléculaire est l'ensemble des disciplines consacrées à l'élucidation des mécanismes biologiques à l'échelle moléculaire, mais devant mener à une compréhension globale des systèmes étudiés. Cela implique tout d'abord l'investigation des structures et interactions des constituants moléculaires de la cellule (en particulier les protéines, les lipoprotéines [structures membranaires] et les acides nucléiques). Une vue étroite des choses ramène parfois la Biologie Moléculaire à ce qui n'est plus précisément que de la Génétique Moléculaire. En réalité, la Biologie Moléculaire dans le vrai sens du terme ne concerne pas que la Génétique et donc les acides nucléiques, mais bien aussi - entre autres - des approches expérimentales ressortissant à la Biochimie (exemple : régulations enzymatiques), à la physiologie cellulaire (exemple : différenciation) et à diverses pathologies (exemples : maladies métaboliques congénitales).

Le Génie Génétique comprend l'ensemble des techniques de la Biologie Moléculaire. Cette discipline s'est structurée et définie autour de la « traduction » de l'information génétique et de ses mécanismes de régulation. Depuis la découverte de la structure de la molécule d'ADN par F. Crick et J. Watson en 1953 et le décryptage du code génétique par M. Nirenberg et J.H. Matthaei en 1961, la Biologie Moléculaire a vu son histoire s'accélérer. Les techniques de digestion enzymatique ont notamment permis d'isoler, cloner et séquencer les gènes.

Dans les années 60, la notion d'expression d'un gène est devenu un concept binaire faisant appel à la notion de programme génétique : il existe dans toute cellule des gènes dits répresseurs ou activateurs qui commandent ou répriment la production de protéines indispensables à la cellule. Depuis une quarantaine d'années, plusieurs démonstrations expérimentales sont venues remettre en question cette approche déterministe au profit d'une nouvelle théorie dite probabiliste : un gène a simplement une probabilité de s'exprimer ou pas à tout moment.

En 1977, deux techniques de séquençage des acides nucléiques apparaissent à peu près simultanément : la méthode enzymatique de Frédéric Sanger et l'approche chimique de Walter Gilbert et Allan Maxam. Dans les laboratoires, chacun se met à séquencer son « gène ». Depuis la technologie de l'ADN recombinant ; notamment le clonage moléculaire a largement tenu ses promesses, en fournissant des outils largement utilisés par les biologistes et qui permettent une moisson de résultats dans tous les domaines de la Biologie, du contrôle de l'expression des gènes à l'étude de l'évolution. Cependant, l'une des difficultés majeures qui limitait les applications d'une telle technologie un nombre relativement restreint de gènes, était l'insuffisance des quantités disponibles d'ADN étudié. Cet obstacle majeur a pu être levé au début de l'année 1987, grâce à la découverte d'une nouvelle technique ingénieuse qui a révolutionné la manipulation des gènes et accélérée l'essor de la Biologie Moléculaire de façon fulgurante. Cette technique appelée PCR pour *polymérase chaîne réaction* ou la réaction de polymérisation en chaîne, a bouleversée la Biologie Moléculaire et s'est implantée rapidement dans les laboratoires, en permettant de produire de gigantesques quantités de séquences d'ADN spécifique sans passer par une étape de clonage. Cette technique devient rapidement un outil puissant et indispensable au séquençage des génomes.

En réalité l'avènement de la Biologie Moléculaire, l'étude d'un phénomène biologique quelconque consistait à associer ce dernier à la structure et à la fonction d'une molécule ou d'un ensemble de molécules biologiques, en l'occurrence des gènes. Le dogme est qu'une fonction biologique donnée est le résultat de l'expression d'un gène ou un ensemble de gènes en ARN et puis en protéines correspondantes. Cependant, ce type d'approches, dit unitaire, malgré sa pertinence prouvée, reste insuffisant pour une compréhension précise et adéquate des processus biologiques, car il ne prendrait pas en compte l'environnement complexe et multifactoriel de la molécule étudiée.

Ainsi, l'étude systématique et globale est nécessaire pour compléter les résultats d'une démarche unitaire. En effet, depuis une quinzaine d'années des progrès très importants dans certaines disciplines telles que la Biologie Moléculaire, la Chimie Combinatoire, l'Electrochimie et la Bioinformatique ont permis le développement d'approches systématique et globale en Biologie. Parmi ces approches, on peut citer, le display différentiel, l'hybridation soustractive, le criblage différentiel sur filtres à haute densité, les puces ADN et les puces à protéines.

De ce fait, depuis les années 90, ont été mises au point des nouvelles approches économiques, rapides et moins coûteuses, et qui reposent sur l'étude des techniques de digestion, de séparation, de transfert et d'analyse de première génération, ces approches regroupées sous le terme génomique visent à exploiter cette base de connaissance afin d'étudier comment l'ensemble des gènes fonctionnent et interagissent dans les cellules et les organismes qui les portent (Lokhart et Winzeler, 2000). L'achèvement du séquençage du génome humain correspond plutôt au début d'une aventure qu'à l'avènement d'une connaissance. Le prochain enjeu est l'annotation des génomes. En effet, nous ignorons encore beaucoup de la structure des génomes, des mécanismes de régulation des gènes ou encore du fonctionnement des produits des gènes.

Accéder à la fonction cellulaire et physiologique d'un gène nécessite de l'étudier dans le contexte de son interaction naturelle avec l'ensemble des autres gènes exprimés dans la cellule, dans un contexte tissulaire ou pathologique spécifique. Cette possibilité qui semblait totalement utopique il y a quelques années est devenue réalité depuis l'apparition de plusieurs techniques. (Strausberg et al., 1997 ; Duggan et al., 1999 ; Brown et Botstein, 1999 ; Debouck et Goodfellow, 1999 ; Bassett et al., 1999 ; Schena et al., 1998). La capacité d'identifier les gènes au niveau des acides nucléiques, et plus particulièrement des ARNm, permet non seulement d'aider à l'annotation des gènes mais aussi de mettre en évidence leurs niveaux et modes d'expression dans des conditions données. Par conséquent, l'approche transcriptomique a rapidement été privilégiée pour étudier les mécanismes de l'expression des gènes.

Le rapport présenté ici vise à mettre en évidence la transcriptomique, une approche systématique pour l'étude de l'expression des gènes. Il est organisé en une introduction générale de biologie moléculaire, trois chapitres : expression des gènes et transcriptomique, techniques de l'analyse du transcriptome et quelques applications de l'analyse transcriptomique et finalement une discussion et conclusion.

I.1. Techniques classiques de l'étude de l'expression des gènes

Avant le développement de nouveaux outils de l'étude du transcriptome, il existait déjà des techniques pour mesurer les niveaux d'expression différentiels des gènes. Certaines de ces techniques sont toujours utilisées parallèlement à l'utilisation des nouvelles techniques.

I.1.1. Northern blot et dot blot

Northern blot est une technique qui permet de détecter la présence d'ARN messagers (ARNm) spécifiques mais également des ARN non codants comme les petits ARN et les ARN ribosomiaux à l'aide de sondes marquées. Les ARN messagers d'un échantillon sont séparés par électrophorèse. La mise en présence du résultat de l'électrophorèse avec une sonde radioactive d'ADN complémentaire (ADNc) de l'ARNm recherché entraîne la détection ou non d'un ARN. La présence de l'ARN est révélée par autoradiographie. Cette technique permet de mesurer l'expression relative d'au plus 20 gènes à la fois (Alwine et al., 1977). Une autre technique est le dot blot dont le principe est à l'origine des puces ADN, demande une quantité de matériel relativement considérable à cause de la taille des filtres (Kafatos et al., 1979).

I.1.2. Criblage différentiel

Une technique de détection des gènes qui sont exprimés uniquement sous certaines conditions. Elle consiste à isoler et à comparer les ARNm amplifiés par PCR (*polymerase chain reaction*) provenant d'au moins deux populations cellulaires. Le criblage se fait *via* des gels d'électrophorèse et se fonde donc sur les différences de longueurs d'ARNm. Les ARNm sont ensuite identifiés grâce au séquençage de l'ARNm prélevé. Cette technique présente des inconvénients comme de nombreux faux positifs générés par la PCR (environ 50%) et un biais en faveur des ARN abondants. L'expression différentielle obtenue doit donc être validée par une autre technique (Liang et Pardee, 1992).

I.2. Etude du transcriptome

I.2.1. Historique

La transcriptomique, terme décrivant l'ensemble des approches permettant d'analyser le transcriptome, qui se définit comme le complément transcrit d'un génome, c'est-à-dire l'ensemble des ARNm transcrits dans un organe donné, donc dans des conditions physiologiques bien définies. La Génétique Fonctionnelle s'intéresse à l'expression des gènes, et entre autres aux ARN messagers (ARNm), première étape de la cascade conduisant des gènes aux protéines puis aux caractères phénotypiques, par l'intermédiaire des différents métabolismes (Hérault et al., 2003).

L'étude du transcriptome est une partie fondamentale de « l'ère de la post-génomique », qui fait suite à « l'ère du séquençage » et à « l'ère de la génomique ». Elle a pris naissance à la fin des années 1990. Elle permet l'étude dynamique des gènes d'organismes modèles identifiés par la génomique. En 1996, lors de l'obtention du

génomique complet de la levure, 6200 phases ouvertes de lecture sont identifiées. Alors que cet organisme était fortement étudié auparavant, seulement un quart de ces gènes pouvait être associé à une fonction connue ou putative à partir d'homologies de séquence. Ainsi, malgré la connaissance de la séquence d'un organisme, une grande partie du monde génomique reste encore inexplorée (Brown et Botstein, 1999 ; Lashkari et al., 1997). Depuis la fin des années 90, une nouvelle ère de la biologie porte sur une vision dynamique globale du fonctionnement cellulaire la post-génomique.

I.2.2. Techniques à la base de l'émergence de la transcriptomique

I.2.2.1. Séquençage d'ADN

L'essor fulgurant des études de génomique fonctionnelle, faisant suite aux avancées des projets internationaux de séquençage des génomes amorcés à la fin des années 80, suscite aujourd'hui un très grand intérêt expérimental et conceptuel (Brent, 2000). Depuis une dizaine d'années, le nombre de gènes identifiés a explosé suite au lancement de programmes de séquençage à très haut débit. De nouvelles technologies engendrées par cette révolution sont utilisées pour quantifier les ARN messagers, collectivement appelés transcriptome, présents dans un échantillon biologique (culture cellulaire, tissu ou individu). Bien que les ARNm ne constituent qu'une étape de l'expression des gènes, leur abondance est souvent corrélée à l'activité des protéines codées et leur quantification en parallèle est plus aisée à conduire que celle des protéines (Lockhart et Winzeler, 2000).

Pour les organismes dont le génome a été séquencé, l'analyse parallèle de l'expression d'un grand nombre de gènes peut être réalisée via l'identification d'étiquettes provenant des transcrits extraits de l'échantillon biologique étudié. La technique SAGE (*Serial Analysis of Gene Expression*) utilise les méthodes de séquençage classique pour déterminer la séquence nucléotidique de ces étiquettes (Velculescu et al., 1995 ; Velculescu et al., 1997). Dans cette approche, le niveau d'expression relatif d'un gène particulier est évalué en fonction du nombre d'occurrence des étiquettes provenant de ce gène, par rapport au nombre total d'étiquettes analysées.

Les progrès considérables réalisés au cours de ces cinq dernières années dans le séquençage du génome humain, ainsi que celui de nombreux virus et bactéries, associés à la mise au point de nouveaux outils (développement de la PCR en temps réel par exemple) amènent de plus en plus de laboratoires à travailler sur les produits d'expression des gènes. À l'échelle qualitative utilisée depuis de nombreuses années (présence ou non d'un transcrit spécifique) s'est récemment ajoutée une échelle quantitative, ouvrant de nouveaux champs d'applications en génétique, cancérologie, virologie, infectiologie et hématologie (Bolufer et al., 2000 ; Sidiropoulos et al., 2001).

Depuis quelques années, bénéficiant des avancées produites par le séquençage du génome humain et des progrès technologiques, se développent de nouveaux outils permettant une analyse moléculaire à l'échelle pangénomique, fournissant 100 à 1 000 fois plus de données que les moyens conventionnels. L'approche la plus développée concerne l'étude du transcriptome d'un tissu, qui mesure simultanément et de façon

quantitative le niveau d'expression - au niveau de l'ARN - de milliers de gènes dans ce tissu (Velculescu et al., 1995).

Dans la foulée du séquençage de génomes complets, la conception et l'exploitation de nouvelles approches expérimentales à haut débit, en combinaison avec de nouvelles méthodes mathématiques et informatiques, ouvrent la voie au décryptage des réseaux de régulation contrôlant les processus cellulaires. Parmi les différents niveaux de régulation impliqués, les mécanismes de contrôle de la transcription sont particulièrement importants et étudiés de manière intensive. Les outils de génomique fonctionnelle les plus répandus - les puces ADN - visent précisément à caractériser les niveaux d'ARN messagers produits au sein des cellules, des tissus ou des organismes étudiés, dans différentes conditions de culture, situations pathologiques ou contextes génétiques (Jordan, 2000).

I.2.2.2. Bioinformatique

La Bioinformatique est un domaine de la Science où la Biologie, l'informatique et les technologies de l'information sont associées pour former une seule discipline. L'objectif visé est de permettre la découverte de nouveaux concepts biologiques et la création d'une perspective globale dans laquelle les principes unificateurs de la biologie sont discernables. Au début de la révolution génomique, la Bioinformatique consistait à créer et à maintenir des bases de données pour le stockage de l'information biologique comme les séquences de nucléotides et d'acides aminés. Le développement d'une telle base de données implique non seulement l'aspect conception mais aussi l'implémentation d'interfaces conviviales complexes permettant d'accéder aux données disponibles et de pouvoir également soumettre de nouvelles. Ces dernières années avec les projets de séquençage de génomes entiers, la masse de données biologiques a considérablement accru. Aux fonctions de stockage, de consultation et de soumission s'ajoutent les fonctions d'analyse et d'interprétation des résultats.

Durant les dernières années, des avancées majeures dans le domaine de la biologie moléculaire couplées avec des avancées dans les technologies génomiques ont conduit à une explosion de la masse de l'information biologique générée par la communauté scientifique. Des exemples d'avancées sont les projets de séquençage de génomes entiers et les études utilisant les microarrays ou biopuces pour caractériser les gènes. Les technologies de génomiques fonctionnelles comme la transcriptomique et la protéomique permettent de déterminer les fonctions de gènes des organismes étudiés. Ces technologies conduisent à la génération de tableaux de données ayant souvent des milliers de lignes et des centaines de colonnes. D'où la nécessité d'un espace mémoire suffisant pour le stockage, une bonne organisation et un indexage soigné afin d'assurer l'intégrité de l'information disponible.

L'utilisation des puces ADN permet la mesure simultanée de très nombreux gènes dans de nombreuses conditions. Il en découle un très grand nombre de données qu'il faut ensuite gérer, analyser et visualiser de la manière la plus informative possible afin d'en extraire le maximum de connaissances sur le processus biologique étudié. Ces impératifs représentent autant de nouveaux défis pour la communauté scientifique. La technique rend nécessaire l'utilisation de bases de données pour structurer, stocker et protéger

l'information obtenue lors d'expériences de mesure d'expression (Ermolaeva et al., 1998). Un effort international sous l'égide de l'Institut européen de bio-informatique (EBI) a permis d'établir les informations nécessaires au stockage dans ces bases de données standardisées. Il est tout aussi important de définir la bonne procédure de normalisation des données. Tous ces points devraient alors permettre la comparaison des données de profils d'expression provenant de différents laboratoires et/ou de différentes technologies.

Parallèlement, de nombreuses méthodes d'analyse bio-informatique des résultats de mesure d'expression ont été mises en œuvre. Il s'agit par exemple de l'estimation des seuils de confiance pour l'expression différentielle (Chen et al., 1997 ; Hilsenbeck et al., 1999) ou de mesures de similarité - ou de distance - entre les éléments que l'on souhaite comparer sur la base des profils d'expression (Eisen et al., 1998). Différents algorithmes ont été utilisés pour l'analyse des profils d'expression, avec une approche non supervisée comme le *clustering* hiérarchique (Eisen et al., 1998), le *K-means clustering* (Tavazoie et al., 1999), le *self-organizing map* (Tamayo et al., 1999), ou avec une approche supervisée comme l'analyse de voisinage (*neighborhood analysis*) (Golub et al., 1999).

Dans le même temps, des modèles d'interactions pour l'étude des réseaux géniques ont été explorés (D'haeseleer et al., 1999). Il s'agit de réseaux d'interactions à cinétique discrète utilisant des niveaux d'expression booléens, discrets ou continus (Thieffrey et Thomas, 1998), ou de systèmes dynamiques utilisant des équations différentielles linéaires ou non (Chen et al., 1999).

I.2.2.3. Techniques de PCR

Depuis la découverte de la *Polymerase Chain Reaction* par Kary Mullis en 1986 et qui lui valu le prix Nobel de chimie en 1993, les techniques de PCR sont rapidement apparues comme des outils indispensables en Biologie Moléculaire. La quantification en temps réel, et avec elle un élargissement des applications à de très nombreux domaines tels que le diagnostique clinique et l'agroalimentaire, fut certes proposée en 1992 par Higuchi R., a connu un récent et spectaculaire développement ces dernières années. Dans le domaine de l'expression génique, les techniques de RT-PCR sont largement utilisées pour estimer le niveau d'expression d'un gène, analyser l'expression des variantes d'épissage génique et valider les résultats des puces ADN (Tse et Capeau, 2003).

Les technologies de PCR et de RT-PCR s'inscrivent dans le cadre plus large de la problématique de l'identification et de la quantification des acides nucléiques. La quantification en temps-réel est plus fiable que la PCR conventionnelle, dans la mesure où l'ensemble du profil d'amplification est connu, permettant d'identifier rapidement des réactions dont l'efficacité d'amplification dévie, par exemple en présence d'un inhibiteur de la polymérase.

La PCR et la RT-PCR quantitative peuvent se résumer en quatre mots: spécificité, Sensibilité, bonne reproductibilité et linéarité pour une dynamique de huit ordres de grandeur (Tse et Capeau, 2003). Ajouter à cela le fait d'être rapide, automatisable et en tube fermé, et vous obtenez une technologie d'un immense intérêt et bénéfice pour la Médecine et la Biologie.

I.2.3. Développement de l'étude transcriptomique

Depuis une vingtaine d'années, plusieurs techniques de Biologie Moléculaire ont été développées afin d'étudier le transcriptome (Liang et Pardee, 1992 ; Adams et al., 2000). Les premières approches proposées, le *Southern blot* et le *Northern blot*, permettent d'identifier et localiser une séquence particulière (sonde d'ARNm ou ADNc) dans un génome entier (cible) ou tout autre mélange complexe d'ADN. Ces techniques se limitent à l'analyse d'un petit nombre de gènes à la fois et ne permettent pas d'appréhender la complexité du phénomène de la transcription. Plus récemment, la technique SAGE (*Serial Analysis of Genes Expression*), permet d'identifier et quantifier, simultanément, le niveau d'expression de plusieurs milliers de gènes, dans un type cellulaire donné (Velculescu et al., 1995). Cette méthode consiste à réaliser un inventaire des transcrits par séquençage en série de courts fragments d'ADNc (9 à 14 pb) ou *sequence tags*. Cette méthode est très sensible mais aussi très longue à mettre en œuvre, coûteuse et se limite à l'évaluation des niveaux d'expression des gènes. Parallèlement à la méthode SAGE, s'est développée la technologie des puces ADN (Schena et al., 1995 ; Lockhart et al., 1996), moins coûteuse et surtout plus évolutive en terme d'applications. En effet, les puces ADN permettent non seulement de visualiser, simultanément, le niveau d'expression de plusieurs milliers de gènes dans un type cellulaire et un contexte physiologique et/ou pathologique particulier ; mais aussi d'étudier la séquence des gènes dans un échantillon, les mutations ou le polymorphisme (Mantripragada et al., 2004). Elles sont donc rapidement devenues les outils privilégiés pour l'analyse du transcriptome.

Le transcriptome requière des analyses statistiques qui dépendent généralement des questions posées. Or, la communauté de biologistes n'est généralement pas formée aux statistiques. Un des défis majeurs de la communauté biologique et bioinformatique est donc d'adopter une vision plus statistique et d'interagir avec des statisticiens (Vingron, 2001). Le biologiste de laboratoire et le théoricien ont besoin de faire des efforts concertés afin de concevoir des expériences qui peuvent à la fois être réalisées et analysées.

Parallèlement, le nombre grandissant de données de transcriptome librement accessibles ouvre la voie à la méta-analyse des puces. Comme cette technique dépend fortement de la technique employée mais surtout des facteurs biologiques expérimentaux, les résultats obtenus à partir d'un seul jeu de données sont difficilement généralisables à un phénomène global comme, par exemple, la compréhension du cancer. Depuis 2002, différentes études ont porté sur l'analyse de plusieurs jeux de données obtenus à partir de différentes plates-formes afin d'identifier des résultats reproductibles et sans doute généralisables (gènes impliqués dans un processus, groupes de gènes aux profils d'expression cohérents). Par ailleurs, la méta-analyse peut permettre l'accès à des résultats plus globaux comme l'analyse de régularités d'expression au sein du génome indépendamment des conditions expérimentales. Pour les bactéries, l'organisation chromosomique de l'expression n'avait pas fait l'objet de méta-analyse du transcriptome. Seule une étude du transcriptome d'*Escherichia coli* a été effectuée dans ce but (Jeong et al., 2004). L'interprétation des résultats a été réalisée en fonction des facteurs expérimentaux (dans ce cas deux conditions expérimentales) : l'implication de protéines (les gyrases) dans le repliement du chromosome bactérien.

CHAPITRE II

TECHNIQUES DE L'ANALYSE DU TRANSCRIPTOME

II.1. Introduction

Depuis le début des années 1990, un grand nombre de techniques d'analyse du transcriptome ont été développées. Certaines méthodes se basent sur la PCR, les autres basées sur l'hybridation et le séquençage sont à l'heure actuelle beaucoup plus largement utilisées. Un grand nombre de méthodes d'analyse ont été développées, nous présenterons les plus utilisées d'entre-elles.

II.2. Techniques basées sur la PCR

Des techniques d'analyse du transcriptome basées sur la PCR ont été développées au début des années 1990 : le *Differential Display* (Liang et Pardee, 1992), l'AP-PCR « *Arbitrarily Primed Polymerase Chain Reaction* » (Welsh et McClelland, 1990), la PCR soustractive et la RT-PCR en temps réel.

II.2.1. Display différentiel et AP-PCR

Le display différentiel commence par une transcription inverse des ARNm en ADNc, en utilisant des amorces oligo d(T). Ces ADNc sont ensuite amplifiés par PCR en utilisant des amorces oligo d(T) et différentes amorces arbitraires s'hybridant en 5'. Puis, les fragments amplifiés sont séparés par électrophorèse. Ceci est réalisé sur différentes populations d'ARNm et permet donc d'obtenir des fragments dont l'expression est différente entre ces populations. Ces fragments sont ensuite séquencés, ce qui permet d'obtenir des fragments de gènes différentiellement exprimés entre les conditions étudiées. Cependant, cette méthode présente deux principaux inconvénients : elle n'est pas quantitative et elle génère de nombreux faux positifs (Green et al., 2001). Cette technique est néanmoins encore utilisée car elle est relativement facile et rapide à mettre en œuvre (Fig. 1). Une autre technique conceptuellement très similaire au display différentiel est l'AP-PCR, mis à part, le fait que les fragments amplifiés ne sont pas spécifiquement situés à l'extrémité 3' des transcrits car des amorces arbitraires sont utilisées à la place des amorces oligo d(T) du display différentiel.

II.2.2. PCR soustractive

Une troisième technique d'analyse du transcriptome basée sur la PCR est encore utilisée actuellement : il s'agit de la PCR soustractive. Le principe général de cette technique consiste à rétrotranscrire les ARNm en ADNc, à les digérer avec une enzyme de restriction puis à les lier à des adaptateurs différents pour les deux conditions étudiées. Les fragments obtenus sont ensuite amplifiés par PCR en utilisant des amorces complémentaires des adaptateurs, puis les deux ensembles de fragments (correspondants aux deux situations étudiées) sont hybridés entre eux. Une PCR utilisant des amorces complémentaires aux adaptateurs utilisés pour une des deux conditions permet ensuite de n'amplifier exponentiellement que les ADNc présents uniquement dans cette condition. Ce principe est à la base des techniques de *Representational Difference Analysis* (Lisitsyn et al., 1993) et de *Suppressive Subtractive Hybridization* (Gurskaya et al., 1996). Comme les deux premières techniques présentées, la PCR soustractive est également rapide et relativement facile à mettre en œuvre, mais elle génère des faux

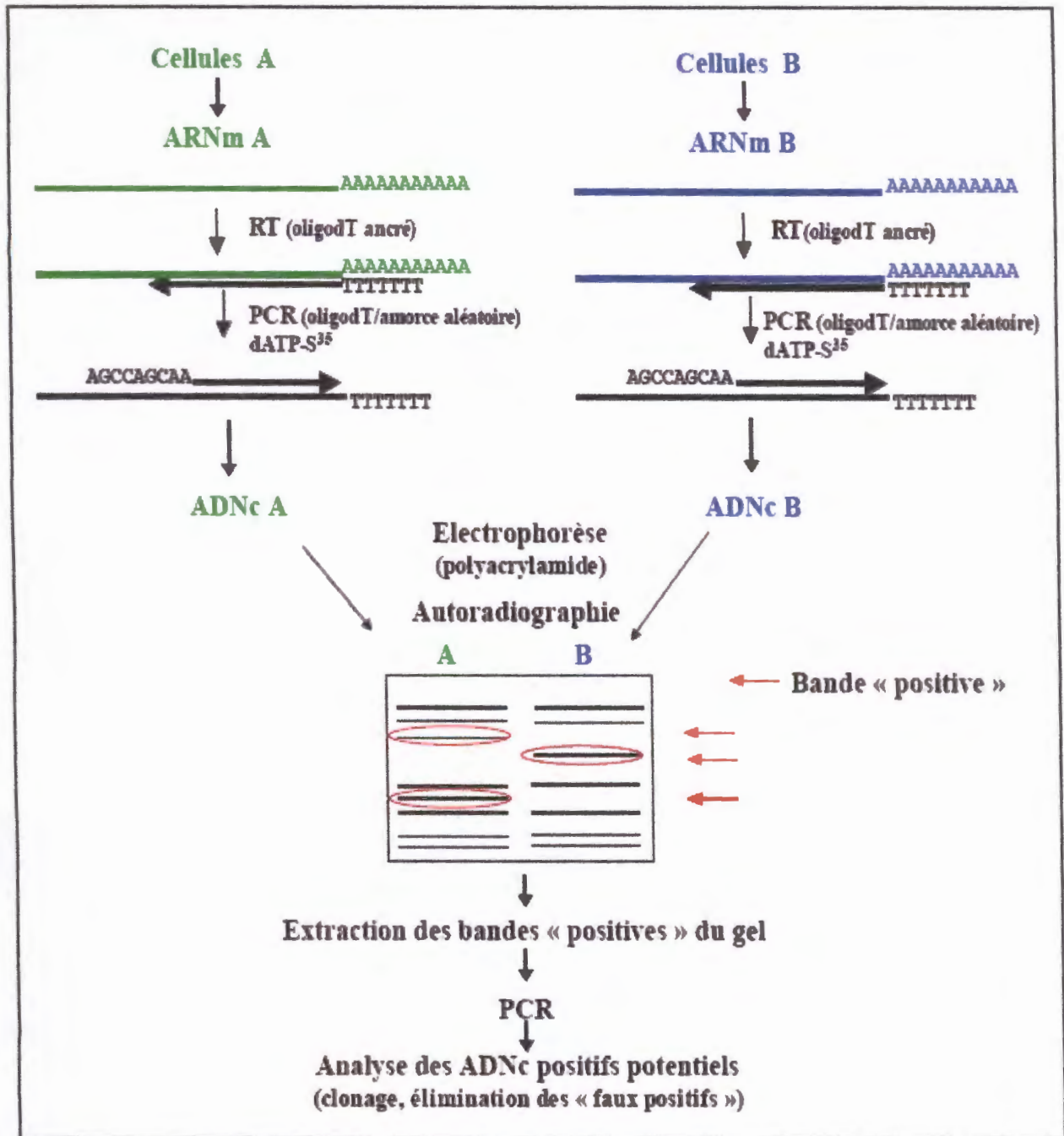


Figure 1. Principe de la technique du display différentiel.

positifs. En outre, elle ne permet de mettre en évidence que des ARNm présents dans une condition donnée, et pas dans une autre condition analysée. Les données obtenues ne sont donc pas quantitatives et deux banques sont nécessaires pour mettre en évidence à la fois les gènes surexprimés et sous-exprimés entre deux conditions (Fig. 2).

II.2.3. RT-PCR en temps-réel

Actuellement, une technique est couramment utilisée afin de mesurer le niveau d'expression d'un gène : la RT-PCR quantitative (*Reverse Transcribed Polymerase Chain Reaction*), C'est-à-dire l'analyse quantitative de l'expression des gènes par RT-PCR en temps réel qui est une méthode alternative aux techniques de microarray pour l'étude du transcriptome. C'est une méthode très sensible, précise et reproductible qui permet l'analyse ciblée d'un nombre de gènes préalablement sélectionnés (la détection des ARNm présents en un seul exemplaire, elle n'est cependant pas utilisée pour mesurer l'expression de l'ensemble des gènes d'un organisme simultanément). Elle permet d'identifier de nouveaux gènes impliqués en pathologie et d'établir des index moléculaires représentatifs de profils spécifiques lésionnels.

L'étude globale de l'expression des gènes ou « transcriptome » permet de déterminer la « signature » d'une cellule ou d'un groupe de cellules dans un état physiologique ou pathologique donné, à un moment donné. Ces méthodes ont permis de définir des profils d'expression spécifiques associés à différents processus physiopathologiques et même de redéfinir certains cadres anatomocliniques (DeRisi et al., 1996 ; Alizadeh et al., 2000). L'analyse du transcriptome fait principalement appel aux puces ADN (ou microarrays) capables de mesurer et de visualiser rapidement des différences d'expression entre les gènes et ceci à l'échelle d'un génome complet (Schena et al., 1995 ; DeRisi et al., 1997). Si le principe de cette technologie est simple, sa mise en œuvre et son utilisation en sont plus complexes.

La RT-PCR en temps réel couple une RT-PCR classique à une méthode de quantification fluorescente (Higuchi et al., 1993). Le système de cette technique utilise le laser de fluorescence qui subit une répartition de l'excitation sur l'ensemble des échantillons par un multiplexer, puis une caméra CDD mesure la fluorescence émise (Fig. 3). La cinétique de quantification est basée sur la détection « en temps réel » du signal fluorescent dont l'intensité est proportionnelle à la quantité de produit PCR généré au cours de l'amplification. La quantification du signal repose sur le concept de « *threshold cycle* » (Ct) (le cycle seuil) où le Ct correspond au plus petit nombre de cycles pour lequel l'intensité du signal fluorescent est supérieure au bruit de fond (Gibson et al., 1996) (Fig. 4).

II.3. Technique de micro-alignements d'oligonucléotides

La grande approche d'analyse de l'expression génique utilise les micro-alignements d'oligonucléotides qui est basée sur l'hybridation, connus également sous leur nom commercial de puces ADN, Genechips[®] de la firme Affymetrix (Lockhart et al., 1996 ; Lipschutz et al., 1999). Les puces ADN aussi appelées biopuces, microréseaux, microarrays, chips ou biochips.

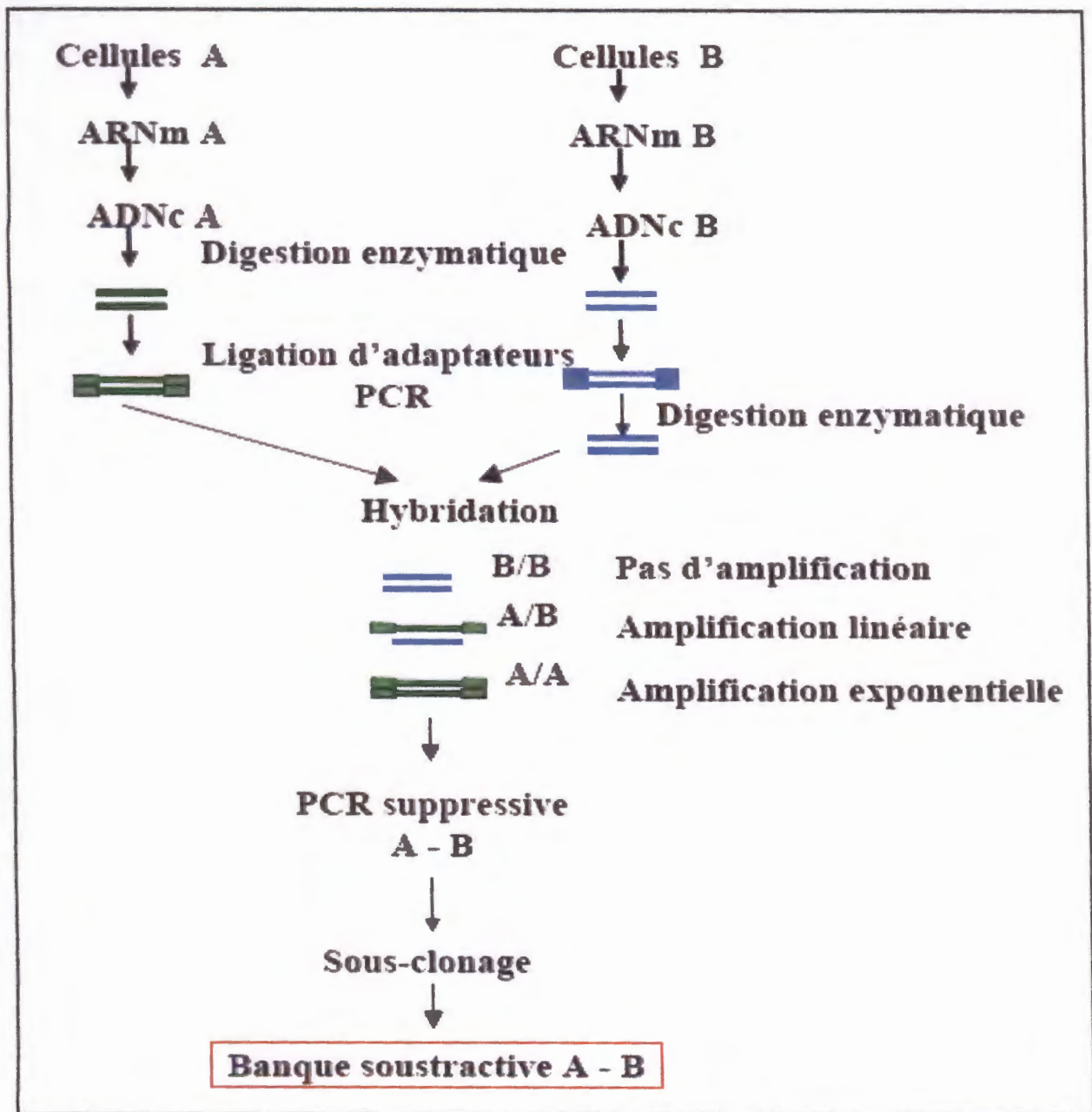


Figure 2. Principe de la technique de PCR soustractive (*Representational Difference Analysis* « RDA » et de *Suppressive Subtractive Hybridization* « SSH »).

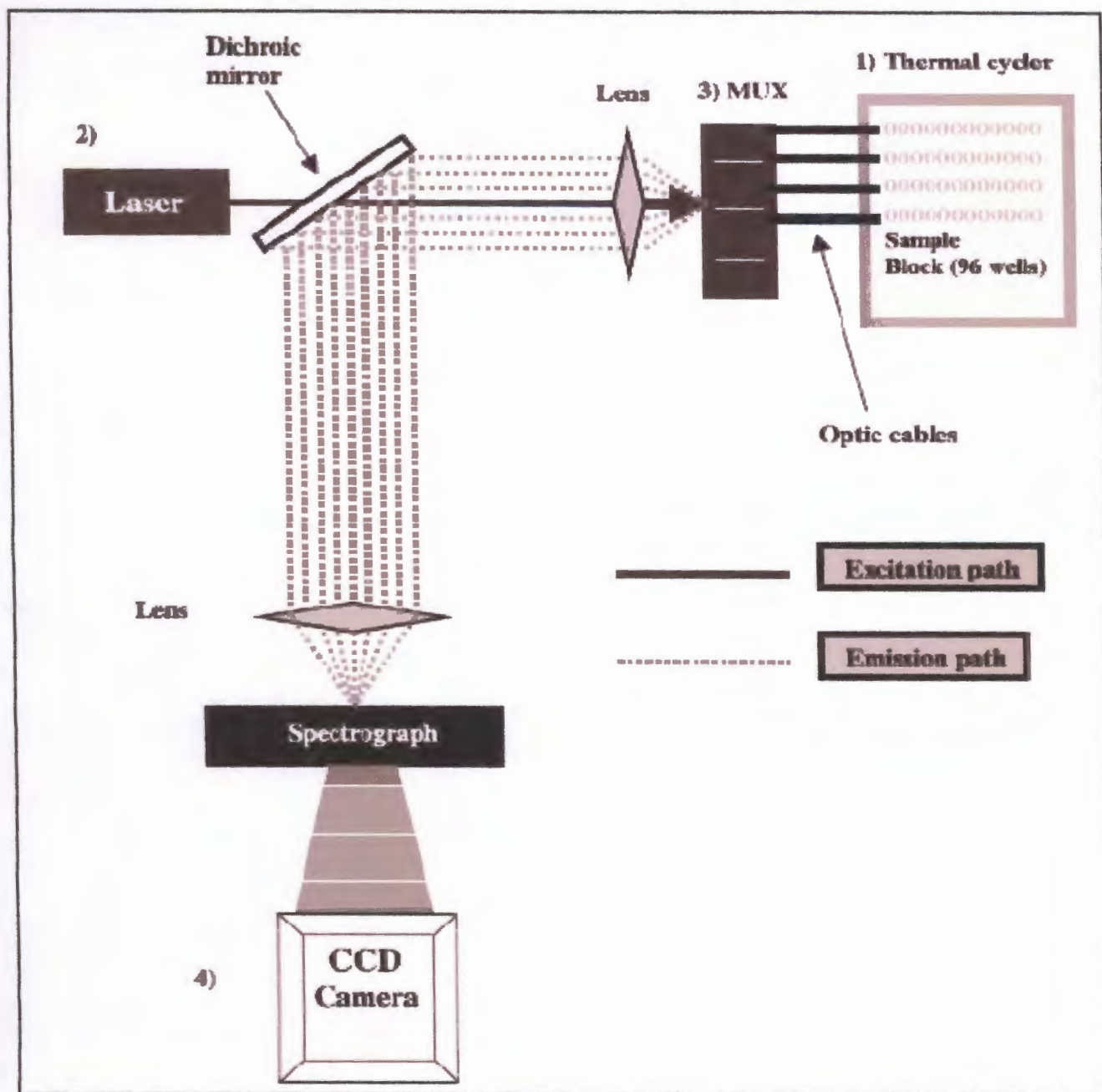


Figure 3. Principe de base d'un système PCR en temps réel (Yong-lee et Alexandra, 2002). Le système est composé : 1) d'un thermocycleur ; 2) d'un laser de fluorescence ; 3) d'un multiplexeur pour répartir l'excitation sur l'ensemble des échantillons ; 4) d'une caméra CCD pour mesurer la fluorescence émise.

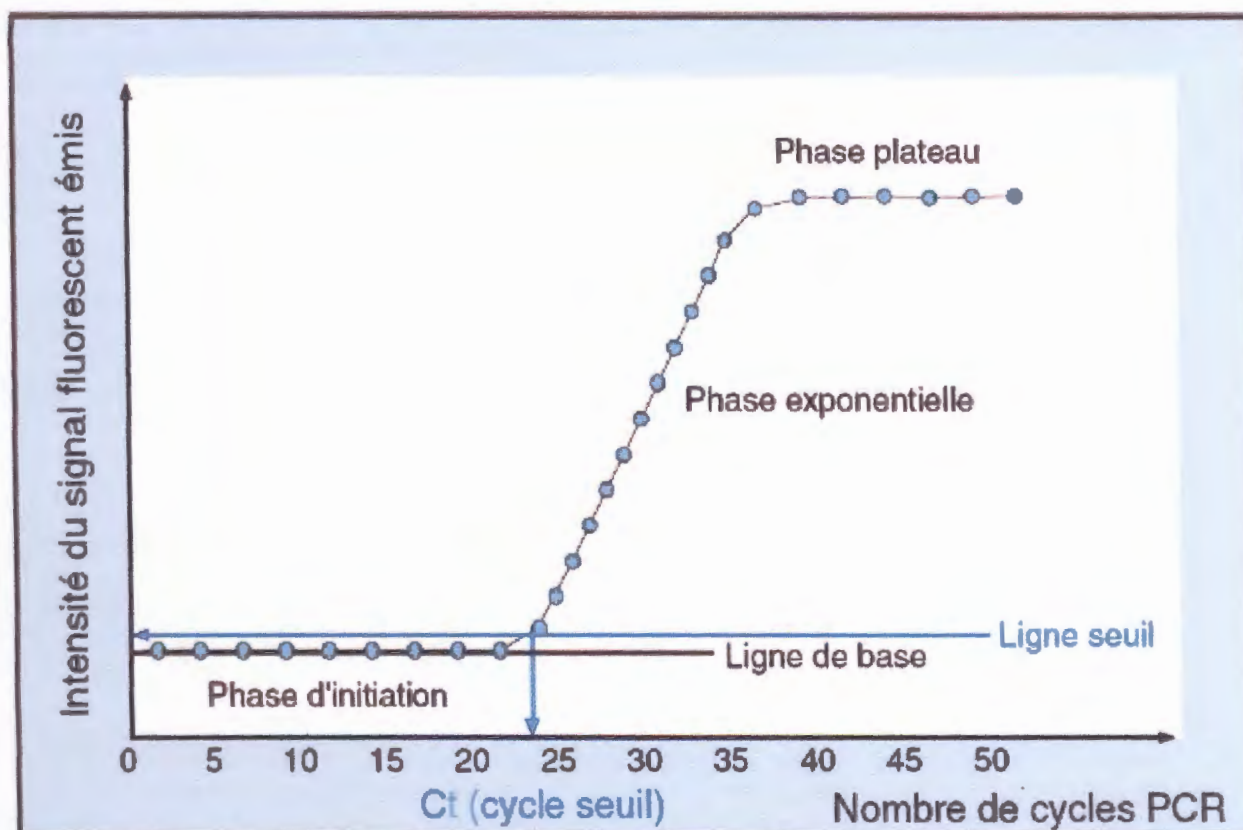


Figure 4. Profil d'une courbe PCR en temps réel. Le nombre de copies est en fonction du nombre de cycles. Plusieurs phases sont à distinguer : une phase au delà de laquelle aucun signal n'est détecté, une phase linéaire et une phase plateau où la PCR atteint un maximum (Tse et Capeau, 2003).

Une puce ADN, aujourd'hui communément appelée « DNA microarray » en anglais (de « array » = rang ordonné), est constituée de fragments d'ADN immobilisés sur un support solide selon une disposition ordonnée. Une puce comporte quelques centaines à plusieurs dizaines de milliers d'unités d'hybridation appelées « spots » (de l'anglais spot=tache), chacune étant constituée d'un dépôt de fragments d'ADN ou d'oligonucléotides correspondant à des sondes de séquences données.

II.3.1. Historique des puces ADN

Les premières puces ADN sont apparues en 1993, mais leur concept date de 1987 (Bellis et Casellas, 1997). Suite logique aux anciennes méthodes de *northern blotting* (Alwine et al., 1997) et d'expression différentielle (Liang et Pardee, 1992), la technologie des puces ADN est basée sur le principe d'hybridation développé par *Southern* (1974). Ce principe stipule que deux fragments d'acides nucléiques complémentaires peuvent s'associer et se dissocier de façon réversible sous l'action de la chaleur et de la concentration saline du milieu. Il s'agit simplement d'une miniaturisation du système classique de reverse dot blot (Lennon et Lehrach, 1991) qui a vu le jour grâce à une technologie pluridisciplinaire intégrant l'électronique (techniques de dépôt), la chimie (préparation des lames et greffes des sondes oligonucléotidiques ou synthèse *in situ*), l'analyse d'images (acquisition des données) et l'informatique (interprétation des données). Depuis leur apparition, les puces ADN suscitent un intérêt inversement proportionnel à leur taille, avec pour preuve l'explosion du nombre de publications qui leur sont dédiées depuis 2001.

Historiquement les *macroarrays*, les *microarrays* et les « véritables » puces ADN correspondent à trois méthodes différentes d'analyse (Lagoda et Reagad, 2000). Les *macroarrays* utilisaient des clones d'ADN complémentaire (ADNc) déposés 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 ADN associaient à chacun des gènes d'un organisme un ensemble d'oligonucléotides synthétisés *in situ*. La première de ces puces 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 (Hinfray, 1997). 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* (De Saizieu et al., 1998).

II.3.2. Principe des puces ADN

L'idée conceptuelle de la puce 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 (Ramsay, 1998 ; Rockett et Dix, 2000). 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 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 (Phimister, 1999) 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.

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. A chacune de ces sondes PM (*Perfect Match*) est associée une sonde MM (*Mis Match*) 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 (PM-MM) de chaque paire de sondes associées à ce gène (Chudin et al., 2002).

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 par transcription inverse au moyen de deux fluorochromes différents : Cyanine 5 (Cy5) et Cyanine 3 (Cy3). Elles sont ensuite assemblées pour former un mélange complexe et hybridées. Ce mélange pourra s'hybrider, dans des conditions de stringence particulières, avec les sondes (oligonucléotides ou clones d'ADNc purifiés et amplifiés) présentes sur la puce. Après hybridation, la lecture est réalisée par un scanner muni d'un microscope confocal, couplé à deux lasers. Ces lasers possèdent des longueurs d'ondes d'excitation spécifiques, correspondant à celles des deux marqueurs fluorescents. L'excitation et l'émission (amplifiée par des photomultiplicateurs) des fluorochromes permettent l'obtention de deux images (une pour chaque marqueur) en niveau de gris. Ces images sont ensuite converties en pseudo-couleur et fusionnées pour être analysées par un logiciel d'analyse d'images (Fig. 5). 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 (Duggan et al., 1999).

II.4. Techniques basées sur le séquençage d'ADN

De manière idéale, la meilleure vision que l'on pourrait avoir du transcriptome serait une banque d'ADNc complets correspondants aux ARNm présents dans les cellules

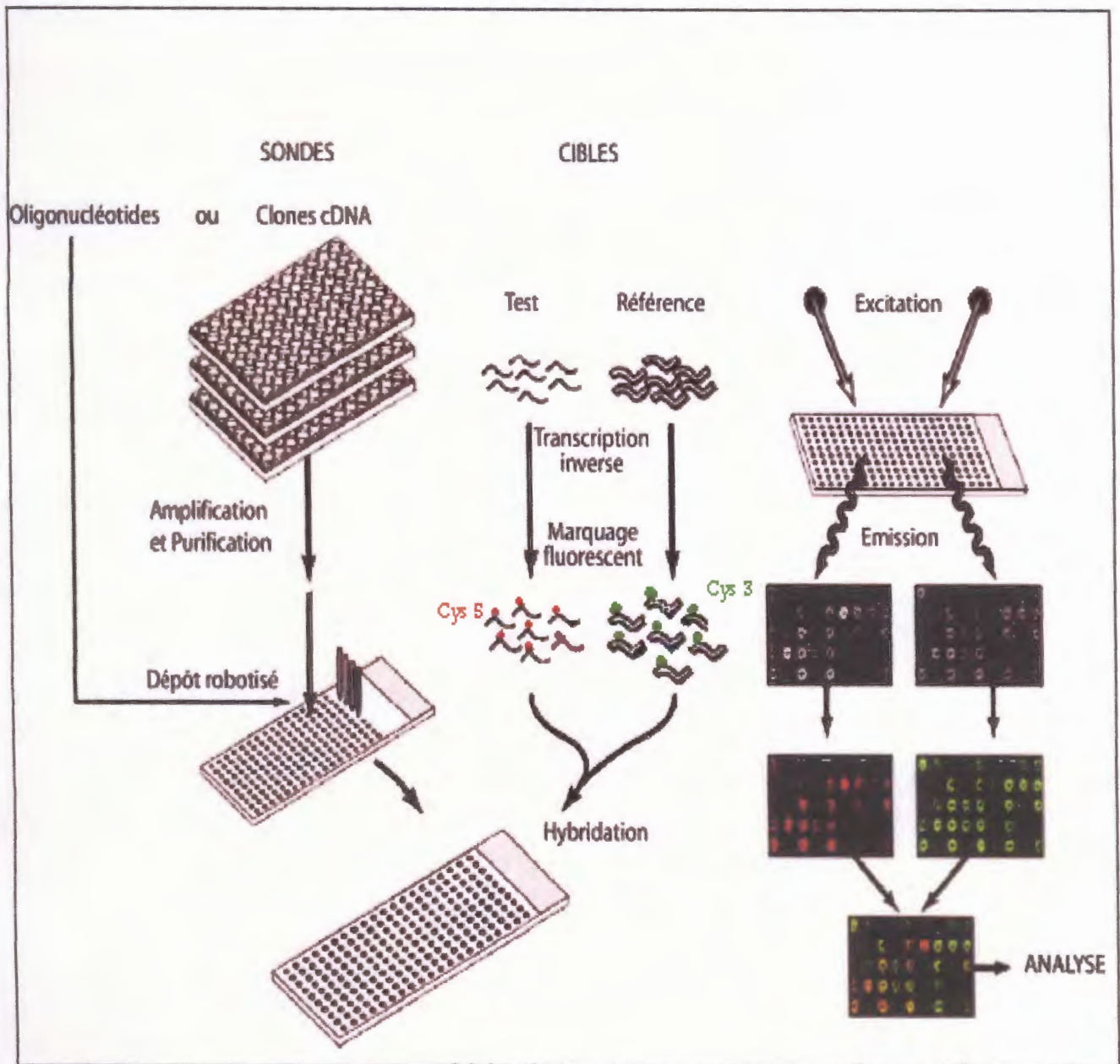


Figure 5. Etapes de l'analyse du transcriptome par la technologie de micro-alignements d'oligonucléotides (d'après Duggan et al., 1999). Les sondes sont déposées mécaniquement sur une lame de verre. Parallèlement, les cibles sont couplées à des marqueurs fluorescents. Par exemple, la cible test est marquée par une Cyanine 5 (Cy5) rouge et la cible de référence par une Cyanine 3 (Cy3) verte. Les cibles sont mélangées puis hybridées avec les sondes. La lecture est réalisée par un scanner couplé à deux lasers. Ces lasers possèdent des longueurs d'ondes d'excitation spécifiques, correspondant à celles des deux marqueurs fluorescents. L'excitation et l'émission des fluorochromes permettent l'obtention de deux images (une pour chaque marqueur) en niveau de gris. Ces images sont ensuite converties et analysées par un logiciel d'analyse d'images.

et les conditions étudiées. Ceci est bien sûr irréalisable dans la mesure où une cellule compte de l'ordre de 400000 ARNm (Hastie et Bishop, 1976). On ne peut donc pas séquencer l'ensemble de ces ARNm pour toutes les conditions que l'on étudie. Certains projets visent donc à séquencer un ensemble non redondant d'ADNc, de manière à ne pas séquencer plusieurs fois la même séquence. Si ces informations sont particulièrement intéressantes en terme de séquences, elles ne sont cependant ni exhaustives ni quantitatives. Afin de se rapprocher de l'exhaustivité et/ou pour permettre d'avoir des données quantitatives tout en limitant le coût de séquençage, d'autres méthodes permettent de ne séquencer qu'une partie des ADNc : il s'agit du séquençage d'EST (*Expressed Sequence Tags*), de la technique SAGE (*Serial Analysis of Gene Expression*) ou MPSS (*Massive Parallel Signature Sequencing*). Il est alors nécessaire d'utiliser des méthodes bioinformatiques pour retrouver ou reconstituer le transcrit auquel correspond ces fragments.

II.4.1. Micro-alignements d'ADNc

L'étude des ARNm passe par leur clonage sous forme d'ADN complémentaires (ADNc) et la constitution de banques d'ADNc, collections de clones d'ADNc représentatifs du transcriptome dans un type cellulaire, un tissu, voire un organisme entier ou plutôt un ensemble de tissus d'un organisme entier (Héroult et al., 2003).

II.4.1.1. Principe

La technologie des micro-alignements d'ADNc repose sur une méthode simple et bon marché, permettant de suivre simultanément le niveau relatif d'expression de milliers de gènes (Scherena et al., 1995). Des fragments d'ADNc amplifiés par PCR (EST) sont déposés à haute densité (10 à 50 dépôts par mm²) sur une lamelle de microscope, puis hybridés avec des échantillons d'ADNc marqués avec un fluorochrome ou un isotope radioactif. L'intensité du signal d'hybridation obtenu par un transcrit donné est supposée proportionnelle à la quantité de transcrit présent dans la population de l'ARN étudiée. Ces différences d'intensité reflèteront les variations de concentration pour un transcrit donné en fonction des traitements (fig. 6). On peut alors effectuer des analyses statistiques et bio-informatiques, qui seront à la base d'hypothèses à tester par des approches de biologie moléculaire classique (Brown et Botstein, 1999).

II.4.1.2. Les EST, étiquettes de séquences exprimées

La première analyse de l'expression d'un grand nombre de gènes a pris la forme de programmes de séquençage de transcrits à grande échelle. Des banques de transcrits ont été constituées à partir d'ARNm provenant de matériels biologiques variés et de diverses conditions physiologiques. La majorité de ces programmes se sont attachés à séquencer les extrémités de clones d'ADNc afin d'étiqueter un maximum de transcrits. Les séquences obtenues, appelées EST, sont de courtes séquences de 300 à 500 nucléotides, issues du séquençage unique de l'extrémité 5' ou 3' d'un ADNc. La qualité de ces séquences est relativement mauvaise car une EST peut comporter jusqu'à 5% d'erreurs et d'incertitudes. Cependant, elle est largement suffisante pour identifier avec certitude l'ADNc étiqueté. Ainsi, les EST fournissent un accès direct à la partie exprimée du génome. Cette approche a été largement utilisée lors de l'analyse de différents

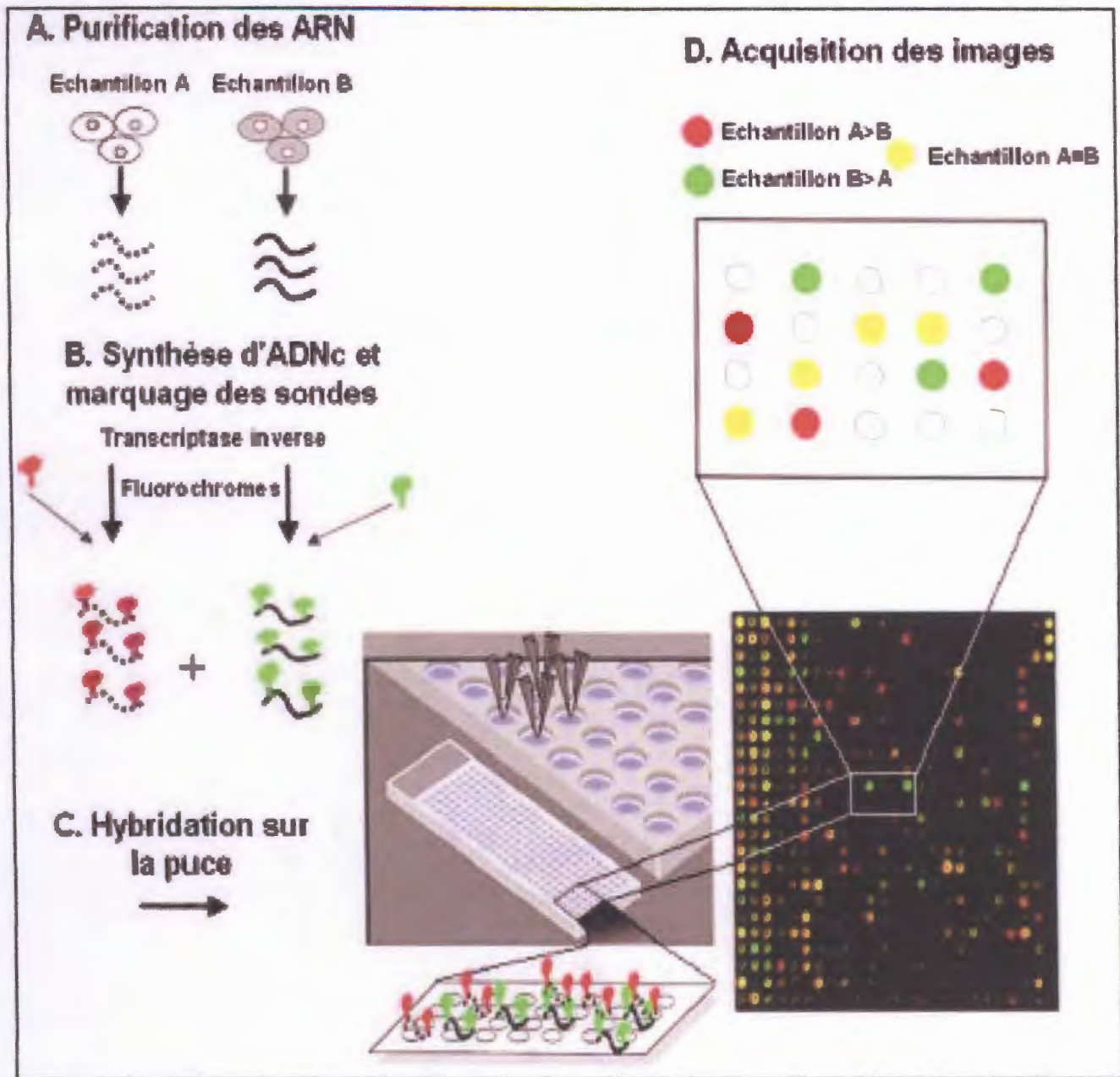


Figure 6: Principe de la technique de micro-alignements d'ADNc. Après extraction (A), les ARN de chaque échantillon sont rétrotranscrits en ADNc, marqués à l'aide d'un fluorochrome spécifique de chaque échantillon (B), puis mis en conditions d'hybridation sur la puce (C). La puce est alors scannée. La nature et l'intensité du signal, acquis et analysés par un logiciel spécialisé, permettent de comparer les abondances des différentes espèces moléculaires fixées sur la lame dans les échantillons A et B (D).

génomés comme celui du nématode *C. elegans* (Waterston et al., 1992), de la drosophile *D. melanogaster* (Rubin et al., 2000 ; Stapleton et al., 2002a ; Stapleton et al., 2002b) ou de l'Homme (Adams et al., 1991 ; Adams et al., 1992 ; Adams et al., 1993 ; Adams et al., 1995 ; Hillier et al., 1996).

Plus le nombre d'EST est important, plus l'estimation de l'expression des gènes est fiable. La qualité des analyses de l'expression du génome à partir des EST est donc directement liée au nombre de séquences (Audic et Claverie, 1997).

Le développement de nouvelles méthodes de sélection des ARNm entiers a permis de compléter les approches EST par des programmes de séquençage de transcrits pleine longueur, par exemple en utilisant la coiffe des ARNm comme point d'ancrage au lieu de la queue poly-A (Seki et al., 1998 ; Seki et al., 2002c ; Seki et al., 2002d ; Wu et al., 2002). S'ils apportent des données précieuses pour l'annotation des gènes (Castelli et al., 2004), ils ne fournissent pas plus d'indications quantitatives sur l'expression des gènes que les EST.

II.4.2. L'analyse en série de l'expression génique

L'analyse en série de l'expression génique, communément connue sous le nom de SAGE (*serial analysis of gene expression*), La technique SAGE, développée par Velculescu et al. (Velculescu et al., 1995), permet une analyse de la fréquence d'un ARNm parmi les milliers de messagers qui sont présents dans une cellule à un moment donné. Cette technique permet de caractériser un transcriptome de façon descriptive ou différentielle sans disposer d'informations de séquence au préalable, et permet la découverte de nouveaux gènes ainsi que leur localisation sur le génome.

II.4.2.1. Principe

La technique SAGE repose sur le fait que la séquence d'un tag est spécifique d'un ARNm donné. En effet, si on fait l'hypothèse que les nucléotides sont distribués aléatoirement le long de l'ARNm, une séquence de 10 pb peut permettre de distinguer $4^{10} = 1048576$ transcrits différents. Différentes publications estiment la proportion de transcrits humains ayant des tags différents (Stollberg et al., 2000 ; Unneberg et al., 2003) : ces estimations varient entre 93,4% et 98,5%. Cependant, on peut penser qu'il s'agit de surestimations, toutes les formes de transcrits humains n'étant vraisemblablement pas encore connues. La plupart des transcrits différents ayant le même tag correspondent à des transcrits épissés alternativement à partir d'un même gène, ou à des transcrits codant pour des protéines paralogues (Unneberg et al., 2003). Il a néanmoins été estimé que 38% des transcrits humains épissés alternativement à partir d'un même gène présentent des tags différents (Unneberg et al., 2003) : il y a donc un grand nombre d'épissages alternatifs que l'on peut mettre en évidence à l'aide de tags de 10 pb.

Cette méthode basée sur trois principes (fig. 7) : premièrement, une séquence étiquette (tag) de 10 pb isolée à partir d'une région définie de l'ARNm est suffisamment informative pour identifier son transcrit d'origine, deuxièmement, la concaténation de plusieurs tags en une seule molécule d'ADN linéaire permet une analyse plus efficace en diminuant et simplifiant les étapes de clonage et de séquençage, troisièmement,

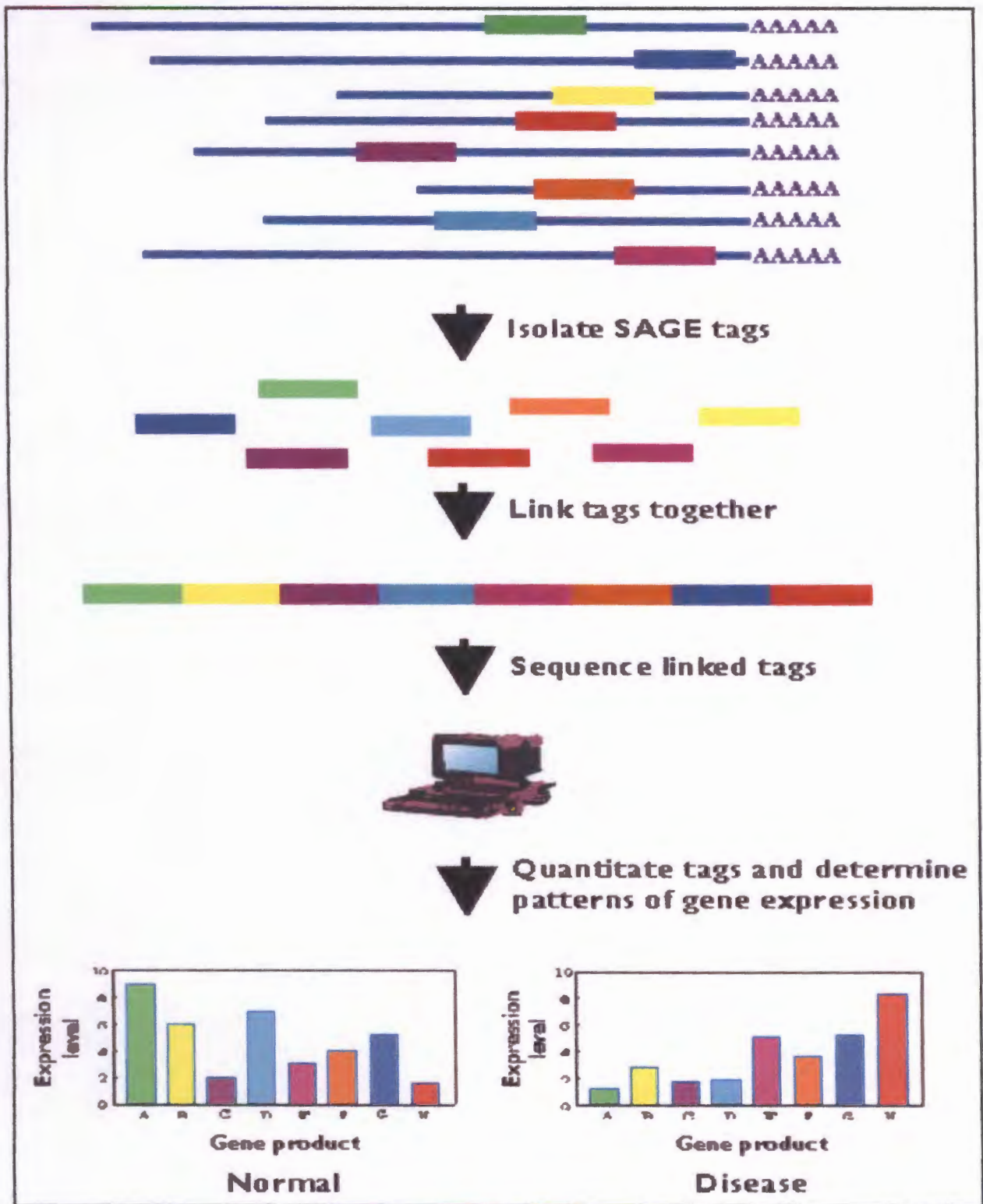


Figure 7. Principe de la technique SAGE. A partir de chaque ARNm est générée une étiquette, ou tag, qui possède une localisation précise sur le transcrit et une longueur suffisante pour en être spécifique. Les tags sont ensuite ligaturés les uns à la suite des autres pendant l'étape de concaténation. Après clonage, les concatémères sont séquencés, et les tags dénombrés. Pour chaque espèce d'ARNm, l'abondance de son tag spécifique est le reflet de son niveau d'expression dans l'échantillon dont il a été extrait.

l'abondance de chaque tag est proportionnelle au niveau d'expression de son transcrite dans la population cellulaire d'origine.

II.4.2.2. Développement de SAGE

Depuis la publication originale de la technique SAGE, différentes variantes de cette méthode ont été proposées, notamment pour permettre de l'appliquer à de petites quantités d'ARNm initiales: microSAGE (Datson et al., 1999), SAGE-lite (Peters et al., 1999), miniSAGE (Ye et al., 2000) et SADE (Virlon et al., 1999). D'autres variantes de la technique SAGE ont également été proposées, permettant d'obtenir des tags plus longs, afin d'augmenter leur spécificité : LongSAGE (Saha et al., 2002) permettant d'obtenir des tags de 17 pb et superSAGE (Matsumura et al., 2005) permettant d'obtenir des tags de 22 pb, cette dernière méthode a été mise au point pour l'étude du transcriptome de plusieurs espèces simultanément, par exemple pour l'étude des interactions hôte-pathogène. Une autre variante de la technique SAGE permet d'obtenir des tags situés en 5' des ARNm appelée CAGE « *Cap analysis gene expression* » (Shiraki et al., 2003), il est alors possible de connaître les sites d'initiation de la transcription des ARNm en plus de leur fréquence.

II.4.3. La technique de MPSS

Conceptuellement, le principe de la méthode MPSS (*Massively Parallel Signature Sequencing*) est très proche de celui de la technique SAGE. Il s'agit de dénombrer des étiquettes de 17 nucléotides obtenues par digestion enzymatique de l'extrémité 3' des ARNm (Brenner et al., 2000). Cependant, la méthode pour y parvenir est différente de celle utilisée dans la technique SAGE (fig. 8(a)): les ADNc obtenus à partir des ARNm sont attachés à des microbilles après digestion par l'enzyme DpnII (fig. 8(a)). Toutes les microbilles sont ensuite analysées simultanément, afin de déterminer le début de la séquence des ADNc liés à chacune des microbilles. La détermination de ces séquences est réalisée par l'ajout successifs d'adaptateurs ayant une séquence connue, auxquels des "décodeurs" couplés à des fluorochromes sont ensuite liés : la fluorescence de chaque bille permet de connaître quel adaptateur s'y est fixé, et donc la séquence de l'ADNc initial lié à la microbille. La fluorescence est enregistrée à chaque cycle. L'analyse finale des images obtenues permet donc de déterminer la séquence de chaque tag MPSS (figure 8(b)). Le MPSS présente donc les mêmes avantages que la technique Long SAGE (les tags générés ayant la même taille), mais également les inconvénients liés à l'utilisation d'une enzyme de restriction. Comparé à la technique SAGE, le principal avantage du MPSS réside dans la méthode employée pour déterminer la séquence des tags, qui permet d'analyser en parallèle un très grand nombre de tags. Ainsi, les bibliothèques obtenues comprennent généralement plusieurs millions de tags : par exemple plus de deux millions dans l'analyse du transcriptome de cellules souches embryonnaires par Wei et al. (2005) et plus de dix millions dans l'analyse du transcriptome de lignées cellulaires humaines par Jongeneel et al. (2003). La technique MPSS permet ainsi de connaître l'expression de gènes très peu exprimés que l'on ne détecte pas dans les bibliothèques SAGE les plus séquencées (parmi l'ensemble des bibliothèques SAGE disponibles dans la banque Gene Expression Omnibus, la plus grande est composée d'un peu plus de 400000 tags). Bien que théoriquement il soit possible de séquencer beaucoup plus de tags dans une bibliothèque

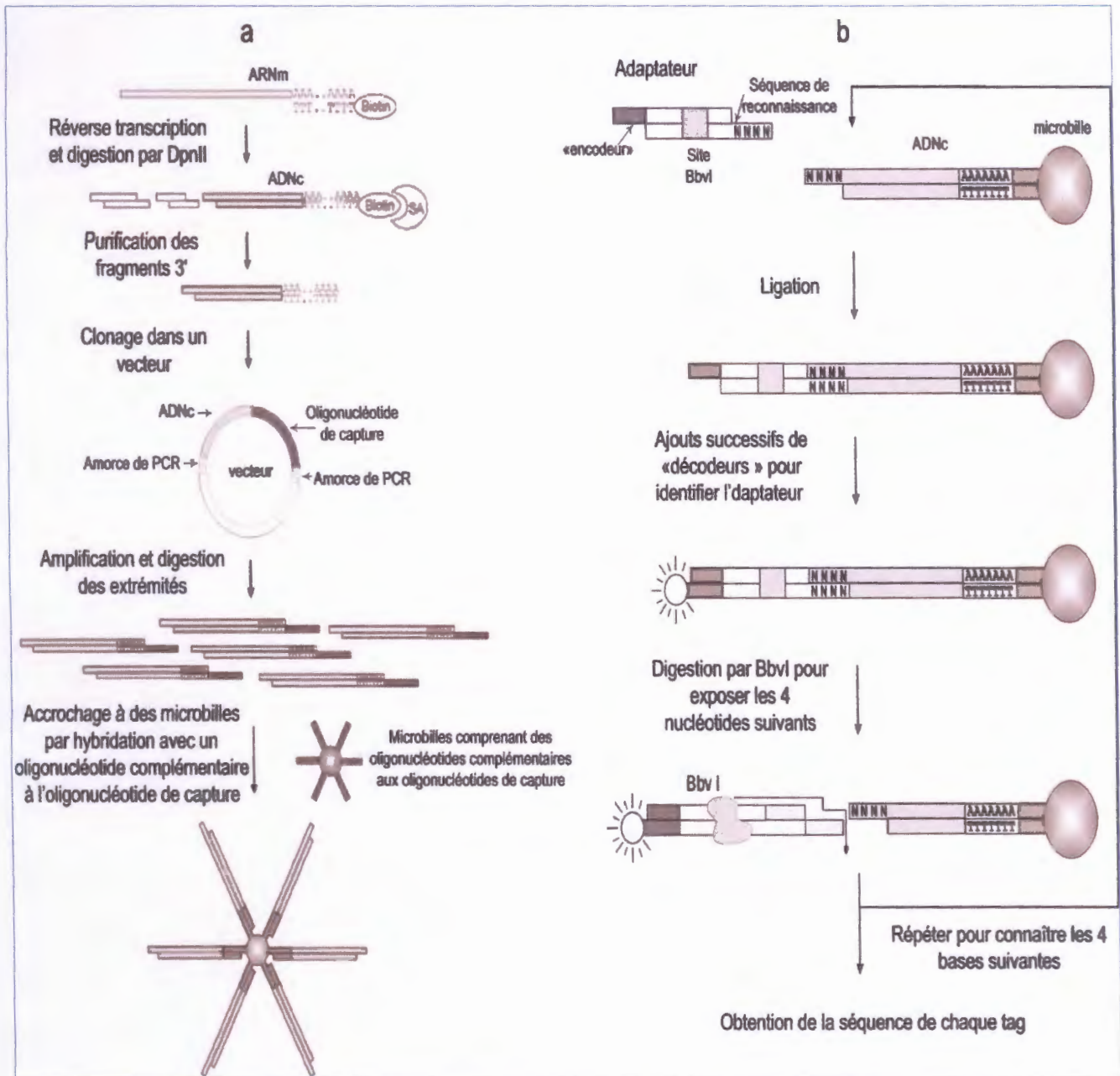


Figure 8. Principe de la technique MPSS. a)- Construction d'une librairie de fragments en 3' d'ADNc à la surface de microbilles. b)- Détermination du début de la séquence d'ADNc présente à la surface d'une microbille. Cette figure ne montre qu'une seule microbille, bien que ce processus soit réalisé simultanément pour toutes les microbilles porteuses de l'ensemble des ADNc initialement étudiés.

SAGE, ceci n'est pas réalisé en raison du coût élevé de séquençage. Cependant, la méthode MPSS a pour le moment été moins utilisée que la technique SAGE, vraisemblablement car elle est plus récente, mais également car la technique est lourde à mettre en œuvre (elle a pour le moment été utilisée par la société Solexa qui a breveté la technique).

CHAPITRE III

QUELQUES APPLICATIONS DE L'ANALYSE TRANSCRIPTOMIQUE

III.1. Introduction

Les champs d'application des outils de la transcriptomique sont nombreux et ils concernent toutes les sciences du vivant, en recherche fondamentale ou appliquée (médecine, pharmacologie, agroalimentaire). Les applications citées ici ne sont donc pas exhaustives.

III.2. Applications dans le domaine de la santé

La technologie de SAGE et celle de micro-alignements ont été très utilisées pour caractériser le transcriptome humain dans divers types cellulaires. On estime que le nombre de gènes exprimés dans n'importe quel type cellulaire est de l'ordre de 15 000 à 25 000, encore que ces chiffres, dérivés des marqueurs de SAGE, puissent surestimer le nombre de gènes uniques et sous-estimer le nombre de transcrits différent. On également caractérisé un transcriptome « minimum » qui consiste en au moins 1 000 gènes qui sont exprimés dans tous les types cellulaires. La plus grande diversité de gènes exprimés se trouve apparemment dans le cerveau ; cela reflète peut-être la complexité des types de neurones. A coté de l'assemblage d'atlas d'expression génique, de nombreuses études ont commencé pour caractériser des différences associées à d'autre maladie humaines.

III.2.1. Diagnostic des cancers

Des progrès technologiques récents (puces ADN et microplaques) permettent la mesure du niveau d'expression de milliers de gènes en parallèle, au cours du temps et pour de multiples conditions. Ces techniques, qui ont été initialement développées et testées sur des systèmes modèles tels que la levure ou des cultures cellulaires in vitro, commencent maintenant à être appliquées à l'analyse des cancers humains. Les résultats initiaux sont prometteurs, et l'analyse des profils d'expression à grande échelle va probablement devenir un outil pour l'identification des tumeurs, leur pronostic et le choix de leur traitement optimal. Il est donc important que les cliniciens se familiarisent avec les principes théoriques qui sous-tendent l'utilisation de ces nouvelles techniques.

III.2.1.1. Cancer du sein

Les études réalisées sur le cancer du sein ont été publiées à partir de 1999, dans un premier temps sur des lignées cellulaires cancéreuses mammaires, puis sur des échantillons tumoraux. Peu d'études se sont consacrées à la comparaison de lignées cellulaires épithéliales mammaires normales et cancéreuses. Celle-ci a plutôt été réalisée dans le but de repérer des gènes d'intérêt étudiés ensuite sur des tissus mammaires. Par exemple, (Nacht et al., 1999) ont sélectionné avec la technique SAGE les 68 gènes les plus différenciellement exprimés entre deux lignées cellulaires cancéreuses mammaires et des cellules épithéliales normales, afin de les déposer secondairement sur des puces ADN destinées à l'analyse de divers tissus mammaires. Parmi ces gènes, certains tels que RBB2, MUC1 et FGF1 ont déjà été impliqués dans l'oncogenèse mammaire, mais un grand nombre n'ont jamais été associés au cancer et constituent des données nouvelles.

D'autres études ont comparé le transcriptome de lignées cellulaires cancéreuses mammaires discordantes selon la présence de récepteurs aux œstrogènes (RE) dans le but d'identifier des gènes associés à ce statut. En effet, on sait que l'association RE-réponse à

l'hormonothérapie n'est pas parfaite : il est possible qu'il existe des gènes exprimés en association aux RE qui soient responsables de la réponse ou de la résistance à l'hormonothérapie. Leur connaissance permettrait d'améliorer notre compréhension du phénotype hormonosensible ou hormonorésistant du cancer du sein et pourrait déboucher sur l'identification de nouvelles cibles thérapeutiques. Certaines équipes ont combiné hybridation soustractive et puces ADN (Yang et al., 1999), d'autres ont appliqué d'emblée la technologie des puces sur un jeu de gènes ayant un rôle suspecté dans l'oncogenèse (Bertucci et al., 1999 ; Hoch et al., 1999). Ainsi, certains gènes non liés préalablement au statut hormonal du cancer du sein ont pu être repérés. Les lignées cellulaires ont également été analysées de manière expérimentale. L'équipe de Lander (Clarck et al., 2000) a publié un travail dont l'objectif était d'identifier des gènes intervenant dans le processus métastatique.

Les travaux sur tissus tumoraux ont le plus souvent utilisé une approche purement descriptive - l'approche expérimentale étant plus difficile à mettre en œuvre - avec un objectif principalement clinique, déboucher sur une meilleure classification des tumeurs en termes de pronostic, voire de diagnostic. En effet, le diagnostic actuel du cancer du sein, et des cancers en général, place un spectre continu de tumeurs de phénotypes différents dans des catégories définies. L'espoir est d'utiliser la masse des données fournies par les puces ADN pour identifier tous ces phénotypes et définir de nouvelles catégories diagnostiques et pronostiques sur le plan fondamental, deux travaux ont été récemment publiés. L'article de Sgori et al. (Sgori et al., 1999) présente l'utilisation combinée de deux techniques, la microdissection laser et les puces ADN, pour comparer, sur un jeu de 8 000 gènes, le transcriptome de populations pures de cellules épithéliales mammaires normales, cancéreuses invasives et cancéreuses métastatiques provenant d'un même patient. (Nacht et al., 1999) ont rapporté un travail assez similaire sur leur jeu de 68 gènes déposés sur des membranes hybridées avec des sondes préparées à partir de divers échantillons : 7 cancers du sein primaires, 7 métastases de cancer du sein et 4 cultures de cellules épithéliales mammaires normales. Ces deux études ont retrouvé des gènes impliqués dans le cancer du sein, comme ERBB2 et MUC1, et identifié des gènes jusqu'à présent non associés à l'oncogenèse mammaire. Certains avaient une fonction connue en conformité avec un rôle potentiel dans la cancérisation, alors que, pour d'autres, il s'agissait d'une relation sans rapport avec les données existantes.

Une approche plus expérimentale consiste à étudier *in vivo* l'effet de divers traitements anticancéreux sur les programmes moléculaires transcriptionnels des cellules tumorales. Dans le cas de leucémies aiguës avec blastes circulants, on peut comparer les profils d'expression des cellules cancéreuses prélevées dans le sang périphérique à différents moments durant le traitement. Dans le cas de tumeurs solides, cela est plus difficile, et la comparaison concernera le plus souvent l'échantillon tumoral avant et après radio- et/ou chimiothérapie. Dans leur étude sur le cancer du sein, Pérou et al. (Perou et al., 2000) ont ainsi comparé le profil d'expression de 20 paires d'échantillons de cancers du sein localement évolués traités par chimiothérapie première puis mastectomie. Après *clustering*, 15 de ces 20 paires voyaient les deux échantillons regroupés sur des branches terminales du dendrogramme. Chez 3 patientes ayant présenté une réponse complète à la chimiothérapie, les échantillons après traitement avaient des profils d'expression proches de celui des échantillons de tissu mammaire normal.

III.2.1.2. Cancer pancréatique

L'analyse du transcriptome du cancer du pancréas a permis de caractériser le profil d'expression génique associé au cancer du pancréas et d'identifier des gènes spécifiquement surexprimés par les cellules néoplasiques. Ces gènes sont des marqueurs candidats potentiels pour le diagnostic, le pronostic et le développement de thérapies ciblées du cancer du pancréas. Un des principaux espoirs des résultats du transcriptome est la découverte d'un groupe de marqueurs diagnostiques précoces qui permettraient une meilleure surveillance chez les patients à risque élevé de développer un cancer du pancréas.

L'analyse du transcriptome des tumeurs, principalement par la technique de puce ADN, a permis d'identifier des profils d'expression génique spécifiques de certains types tumoraux et de mettre en évidence des associations entre des profils d'expression et des données cliniques (Bertucci et al., 2002 ; Yang et al., 1999). Le transcriptome du cancer du pancréas a été étudié selon deux approches méthodologiques : analyse sériée de l'expression de gènes ou SAGE et puce ADN. Dix prélèvements de tissu d'origine pancréatique humain ont été analysés par SAGE: 6 lignées cellulaires d'adénocarcinomes canaux du pancréas (AsPC1, Capan-1, Capan-2, Hs766T, Panc-1 et PL45), 2 lignées issues de cellules épithéliales normales de canaux pancréatiques transformées par la télomérase ou les protéines E6 et E7 du papillomavirus humain (H126 et HX) et 2 adénocarcinomes canaux infiltrants primitifs (Panc-91-16113 et Panc-96-6252) (Ryu et al., 2002). Après filtrage des données, 86 gènes ont été identifiés comme différentiellement exprimés dans les prélèvements tumoraux comparés aux prélèvements de pancréas normaux (49 surexprimés et 37 sous-exprimés).

Parallèlement à l'analyse par SAGE, les puces ADN ont été utilisées pour l'analyse du transcriptome du cancer du pancréas. Une première étude portant sur 11 tissus pancréatiques normaux, 14 carcinomes infiltrants canaux et 8 lignées cellulaires de carcinome, étudiés sur puces Affymetrix U95, a permis d'établir une liste de 97 gènes surexprimés au moins 5 fois plus dans les prélèvements tumoraux que dans les prélèvements normaux (Iacobuzio-Donahue et al., 2002). Une deuxième étude a porté sur un nombre à peu près identique d'échantillons (5 tissus normaux, 17 carcinomes primitifs et 14 lignées cellulaires) avec les puces du laboratoire de P.O. Brown de l'université de Stanford (Iacobuzio-Donahue et al., 2003). Après analyse à l'aide de l'algorithme SAM (*significance analysis of microarray*), 149 gènes ont été identifiés, surexprimés au moins 3 fois plus dans les prélèvements tumoraux que dans les prélèvements normaux. Parmi ces gènes, 46 sont communs avec la précédente étude ou avec les résultats de l'analyse par SAGE. Parmi les 103 gènes nouvellement mis en évidence, 40 ont déjà été retrouvés impliqués avec d'autres types tumoraux et 63 n'ont jamais été rapportés jusqu'à présent dans aucune tumeur.

Ces études permettent l'identification de nombreux gènes marqueurs candidats du cancer du pancréas. Ces gènes peuvent être classés selon leur fonction principale dans le métabolisme cellulaire : jonction cellulaire (claudines, connexines, annexines), cytosquelette (kératines, fascine, pleckstrine), contrôle de la réplication et/ou de la transcription de l'ADN (topoisomérase II, ARNTL2, RUNX1), réparation de l'ADN (XRCC4), régulation du cycle cellulaire (CDC2). Un groupe de gènes préférentiellement surexprimés par les cellules du stroma tumoral a également été identifié (Ryu et al.,

2001). En comparant les profils d'expression génique obtenus par l'analyse des carcinomes primitifs et des lignées cellulaires de carcinome, les gènes préférentiellement exprimés par le stroma tumoral étaient présents dans les cancers primitifs et absents dans les lignées cellulaires. Parmi ces gènes figurent les gènes agissant sur la matrice extracellulaire (gènes de collagène et gènes à activité protéasique), les gènes associés à l'angiogenèse (hevin, IGFBP7, SPARC), les gènes liés à la réponse immune (chaînes légères et lourdes des immunoglobulines) et d'autres gènes de fonction peu connue dans l'oncogenèse comme les gènes du métabolisme lipidique apolipoprotéines).

III.2.2. Détection d'activité biologique d'extraits végétaux à visée cosmétique ou dermatologique

Des fragments de gènes, ADN amplifiés par PCR, sont greffés sur des puces et dénaturés afin qu'ils se trouvent sous forme simple brin sur la puce et puissent s'accrocher au brin complémentaire issu de la cible testée. Des cellules de peau, kératinocytes humains normaux, mélanocytes, fibroblastes, sont cultivées *in vitro* avec et sans extrait végétal pendant un temps donné. Puis les cellules sont lysées et les ARN messagers récupérés. Le principe de dosage est basé sur le pourcentage de fixation des ADNc provenant d'un gène donné issu des cellules cultivées sur la séquence complémentaire du gène fixé sur la puce. Selon le pourcentage de fixation par rapport au contrôle (cellules incubées avec le solvant ou sans traitement), on pourra déterminer si on a activation ou inhibition de l'expression d'un ou plusieurs gènes-cibles. Les résultats doivent être interprétés par des méthodes statistiques complexes et dans le contexte complexe des voies métaboliques. L'analyse bio-informatique, à partir de la littérature scientifique, identifie des groupes de gènes stimulant et répresseur et leur hiérarchie. Elle regroupe les gènes qui jouent sur des motifs protéiques communs, qui sont impliqués dans des mêmes voies métaboliques, localisés de façon proche sur même chromosome ou qui sont corrélés à une même pathologie.

On peut ensuite essayer de regrouper des gènes ayant le même profil d'expression sur plusieurs expériences ayant un rapport biologique. Ce regroupement peut se faire de proche en proche et rassemble les profils les plus similaires. On peut également faire appel à des techniques plus complexes, comme l'analyse en composante principale ACP ou les réseaux neuronaux. Plusieurs expériences sont nécessaires pour faire une analyse correcte.

Au final, on représente en général le « clustering hiérarchique » sous la forme d'une matrice où chaque colonne correspond à une expérience et chaque ligne correspond à un gène. On représente les ratios grâce à une échelle de couleur, par exemple du vert pour les gènes réprimés au rouge pour les gènes induits (Hegde et al., 2000 ; Schena et al., 1995).

III.3. Applications dans le domaine de la recherche

III.3.1. Etude du transcriptome en neurobiologie

La technique des puces ADN a été abondamment utilisée pour explorer le cerveau dans un contexte pathologique. Plusieurs études se sont intéressées aux changements intervenant dans le système nerveux central au cours du vieillissement et/ou de la neurodégénérescence en utilisant des tissus *post-mortem* de patients atteints de la maladie

d'Alzheimer (Blalock et al., 2004 ; Ginsberg et al., 2000 ; Yao et al., 2003), ou encore des modèles animaux (Blalock et al., 2003 ; Dickey et al., 2003 ; Mirnics et al., 2003). L'étude de l'hippocampe de 22 patients souffrant de cette maladie a notamment permis de mettre en évidence une activation des gènes impliqués dans les processus biologiques de différenciation et prolifération, dont des suppresseurs de tumeurs, des facteurs de croissance des oligodendrocytes et des modulateurs de la protéine kinase A (Blalock et al., 2004).

Il n'existe qu'un nombre limité d'études transcriptomiques de la maladie de Huntington (MH) sur des cerveaux humains (Hodges et al., 2006). De nombreuses études ont exploré l'expression des gènes dans des modèles de maladie de Huntington (MH), et ont ainsi mis en évidence un nombre considérable de gènes altérés dans leur expression. Ainsi, Luthi-Carter et al., ont montré que 63% (184/293) des gènes exprimés dans le striatum et le cervelet qui sont régulés par l'expression de huntingtine mutante tronquée, montraient aussi des niveaux d'expression altérés dans un modèle murin d'atrophie dentato-rubro-pallidoluysienne, une autre condition neurodégénérative provoquée par une expansion de glutamine dans le gène de l'atrophine1 (Luthi-Carter et al., 2002b). Cette découverte corrobore l'hypothèse de l'existence de mécanismes communs entre les maladies à expansion de glutamine. Au moins huit conditions de ce genre affectent l'homme, et il serait intéressant de les comparer entre elles de la même façon. Par ailleurs, selon Chan *et al.*, l'expression d'une huntingtine mutante entière a beaucoup moins d'effet sur l'expression des gènes que celle d'une protéine tronquée, ce qui suggère que le contexte de la protéine serait un aspect important de la neuropathologie de la maladie (Chan et al., 2002). Enfin, l'étude récente chez la souris transgénique R6/1, de la régulation de 54 gènes préférentiellement exprimés dans le striatum a montré la baisse de l'expression de 81% de ces gènes, ce qui tend à démontrer l'altération de fonctions spécifiques du striatum lors du processus pathologique (Desplats et al., 2006).

L'application des puces ADN à l'étude des profils d'expression géniques dans des tissus de patients atteints de la maladie de Parkinson (MP) devrait mettre en évidence des niveaux d'expression altérés pour les gènes liés à la vulnérabilité spécifique des neurones dopaminergiques dans cette maladie. Récemment, une analyse par puces ADN à haute densité de l'expression des gènes dans la substance noire compacte (SNc) de patients parkinsoniens a révélé la régulation de 137 gènes, dont 68 réprimés et 69 induits (Grunblatt et al., 2004). Les gènes réprimés codent des protéines impliquées dans la transduction du signal, la transmission dopaminergique, le transport des ions et le métabolisme énergétique. Un de ces gènes, SKP1, code une protéine impliquée dans la formation de l'E3 ubiquitine ligase et de plusieurs sous-unités du protéasome. La baisse d'activité de la voie ubiquitine-protéasome pourrait expliquer l'accumulation anormale de protéines comme l' α -synucléine qui est le principal composant des corps de Lewy. Cette étude a également mis en évidence l'activation de gènes impliqués dans des processus biologiques comme l'inflammation et le stress. Une autre étude par puces ADN a comparé les signatures *post-mortem* de SNc de patients atteints de la MP, de paralysie supranucléaire progressive (PSP), de démence frontotemporale et parkinsonisme (DFTP) et de sujet sains (Hauser et al., 2005). Ce travail a identifié 142 gènes différentiellement exprimés entre les patients atteints de la MP et les sujets sains, et 96 gènes pour les cas de PSP et DFTP comparés aux échantillons témoins. Les gènes codant les protéines de choc thermique HSPA1A et HSPA1B sont induits dans les trois pathologies, ce qui suggère une réponse commune de ces maladies pour tempérer les effets toxiques des protéines

mal repliées. De plus, l'expression du gène PARK5 (codant l'ubiquitine carboxyterminale-hydrolase L1 et responsable d'une forme génétique de la MP) est divisée par deux chez les patients parkinsoniens, apportant un argument supplémentaire en faveur du rôle joué par le déficit de la voie ubiquitine-protéasome dans la MP. D'autres études ont utilisé des modèles de la MP pour explorer le rôle potentiel du stress oxydant, de l'inflammation et de mécanismes inconnus dans la neurodégénérescence. Grünblatt et al., ont notamment étudié le modèle MPTP (1-méthyl-4-phényl-1,2,3,6-tetrahydropyridine) de la MP par la technologie des puces ADN. L'expression de 51 gènes impliqués dans l'inflammation, le stress oxydant, les voies du glutamate et des facteurs neurotrophiques, ainsi que dans des processus inconnus, était altérée par l'intoxication (Grünblatt et al., 2001). Cette étude a identifié une série de « cascades » d'expression de gènes associée à l'administration de la neurotoxine, indiquant de nouvelles cibles thérapeutiques potentielles.

III.3.2. Etude transcriptomique chez le *Leishmania*

Bien qu'il n'y ait pas, chez *Leishmania*, de régulation de l'expression des gènes au niveau transcriptionnel, plusieurs messagers semblent s'accumuler à des stades de développement spécifiques. Les mécanismes impliqués sont peu connus mais pourraient faire intervenir une modulation de la stabilité des ARNm, soit par régulation de la polyadénylation soit par interaction avec des facteurs spécifiques (Clayton, 2002). Des techniques de criblage différentiel, ont permis d'isoler une vingtaine de gènes exprimés spécifiquement à l'un des stades de différenciation du parasite. Les analyses récentes de criblage différentiel, réalisées sur des ADNc de promastigotes et d'amastigotes hybridés sur des filtres à haute densité, ont montré que quelques gènes sont différentiellement exprimés entre ces deux stades parmi lesquels des gènes codant pour une famille de protéines membranaires appelée les amastines (Lamontagne et Papadopoulou, 1999 ; Wu et al., 2000). D'autres travaux utilisant des technologies apparentées ont également révélé l'expression de gènes spécifiques de ces stades de différenciation (Charest et al., 1996 ; Bellatin et al., 2002). Une technique qui gagne en popularité pour mesurer l'expression différentielle des gènes est celle des puces ADN. Cette technique permet de déposer sur des lames de verre des milliers de sondes qui peuvent ensuite être hybridées simultanément avec des ADNc marqués par des nucléotides fluorescents provenant, par exemple, de deux stades différents de différenciation (Fig. 9). Ce type de technique a déjà montré son efficacité pour de nombreux micro-organismes (Schoolink, 2002) et pour quelques parasites comme l'agent de la malaria (Rathod et al., 2002) ou *T. brucei* (Diehl et al., 2002). Bien que le génome de *Leishmania* ne soit pas connu dans son intégralité, des puces ont été réalisées soit avec des fragments d'ADNc, soit avec des fragments aléatoires génomiques de 1 kb (Almeida et al., 2002 ; Beverly et al., 2002) et quelques leçons peuvent d'ores et déjà être tirées de leur utilisation. Tout d'abord, il a été possible de confirmer l'expression différentielle qui avait été prédite pour certains gènes, ce qui a validé la technique. Cependant, l'expression de seulement 1 à 2% des gènes s'est trouvée modulée par un facteur de 2 ou plus entre la forme promastigote et métacyclique (Beverly et al., 2002). Ce pourcentage est en deçà des changements observés entre la forme logarithmique et stationnaire par exemple chez *E. coli* ou chez *S. cerevisiae* (environ 5%) (Talaat et al., 2002). Ceci peut s'expliquer par le mode particulier de régulation génétique de *Leishmania* qui utiliserait préférentiellement une régulation de type traductionnel. Nous avons effectivement identifié dans la région 3' non traduite du messenger de l'amastine une séquence qui gouverne l'efficacité de la traduction et ce

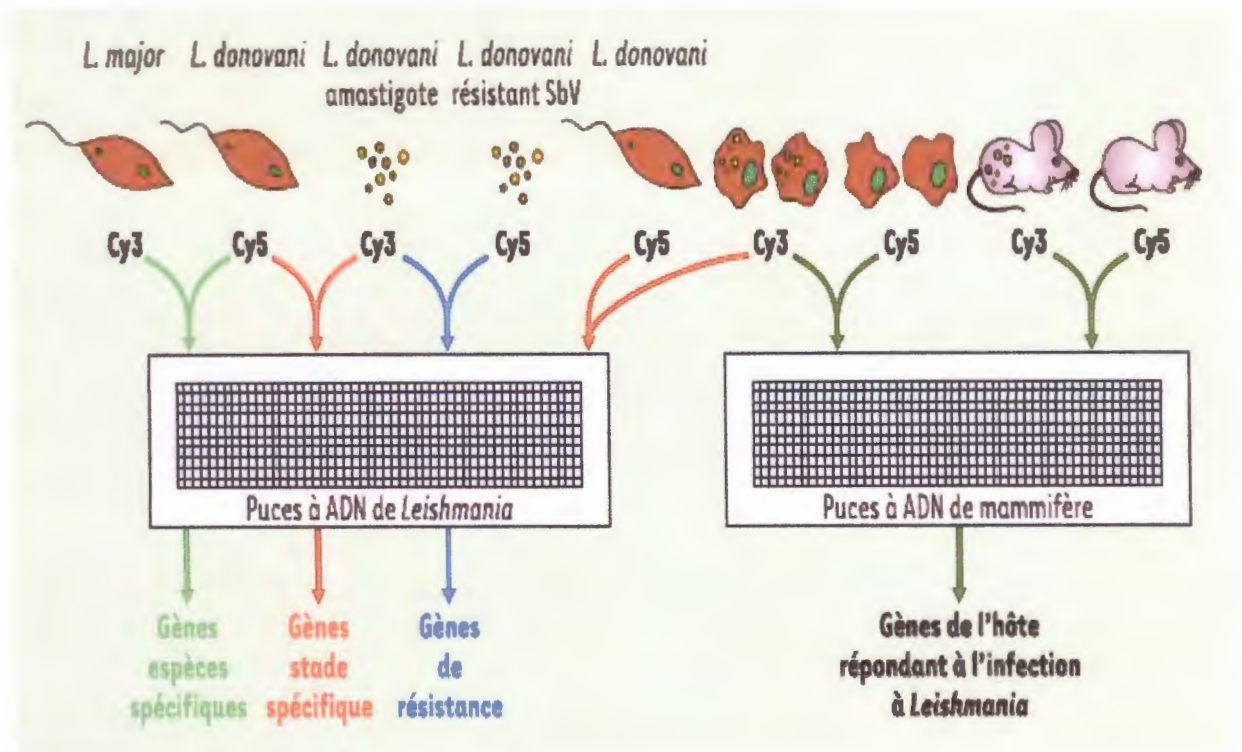


Figure 9. Études transcriptomiques chez le *Leishmania*. On peut utiliser des puces à ADN par exemple pour trouver des gènes propres à une espèce de *Leishmania*, en marquant les ADN de deux espèces différentes telles que *L. major* et *L. donovani* avec deux fluorophores différents comme Cy3 et Cy5 et en réalisant une hybridation compétitive sur lame. De la même façon, il est possible de détecter des gènes exprimés spécifiquement à un stade de vie du parasite (promastigote ou amastigote), ou différemment exprimés chez les parasites sensibles ou résistants. En marquant avec des fluorophores différents des ARN provenant de macrophages naïfs ou infectés par le parasite, ou provenant d'animaux infectés ou non infectés, il sera possible d'observer la réponse de l'hôte de façon globale lorsqu'il interagit avec le parasite.

spécifiquement au stade amastigote (Boucher et al., 2002). Une analyse *in silico* des 34 Mb de séquences de *Leishmania* a révélé que cette séquence est très répandue et dans chaque cas étudié, contrôle la traduction des gènes spécifiquement au stade amastigote (Boucher et al., 2002). Les mécanismes impliqués dans ce nouveau mode de régulation de l'expression des gènes sont actuellement à l'étude (pour revue : Marc Ouellette et al., 2003).

III.3.3. Mise en évidence de gènes différentiellement exprimés

Les premières analyses du transcriptome avec les puces ADN ont établi des catalogues de gènes exprimés dans des cellules ou des tissus d'intérêt. Ces études ont participé à la validation de la technologie par l'obtention de résultats conformes à ceux obtenus avec d'autres approches. Ces catalogues ont été comparés à partir des échantillons biologiques dans le but de détecter des gènes différentiellement exprimés. Par exemple, la comparaison du transcriptome du muscle avec celui d'autres organes a permis de mettre en évidence des gènes d'expression spécifiquement musculaire (Pietu et al., 1996), voire d'expression modulée dans des pathologies musculaires (Tkatchenko et al., 2000 ; DeRisi et al., 1996) ont comparé le profil d'expression d'une lignée cellulaire cancéreuse de mélanome avec le profil de la même lignée dont le phénotype cancéreux avait été annulé par l'introduction du chromosome 6 et ont pu ainsi identifier plusieurs gènes potentiellement responsables du caractère tumoral. De même, l'utilisation d'animaux transgéniques ou ayant subi une inactivation génique a permis de caractériser des gènes jouant un rôle fondamental dans l'ontogenèse thymique chez la souris (Wurzel et al., 2000). La limite majeure de ce type d'approche différentielle qui détecte des gènes d'expression modulée est qu'elle ne renseigne pas sur les raisons de cette modulation. En particulier, de nombreux gènes peuvent être modulés dans leur expression pour des raisons autres que celles espérées (fond génétique, vitesse de croissance...). Ainsi, l'analyse de plusieurs lignées cellulaires prélevées à différents temps de culture et soumises à différents traitements de stimulation et/ou d'inhibition apportera plus d'informations biologiques que la simple comparaison de deux lignées « au repos ».

III.3.4. Etablissement d'un profil caractéristique d'un état biologique donné

Une alternative à l'approche différentielle consiste en l'établissement de profils d'expression génique de différents échantillons caractéristiques d'un état biologique donné, un phénotype tumoral par exemple (Alizadeh et al., 2000 ; Bertucci et al., 2000). L'analyse comparative des profils aura pour but une classification des échantillons basée uniquement sur le transcriptome, avec l'espoir d'identifier de nouveaux sous-types non repérés par les facteurs histocliniques usuels et pertinents en termes diagnostiques et/ou pronostiques. Il existe plusieurs systèmes de classification, mais l'outil le plus utilisé jusqu'à présent est le logiciel de *hierarchical clustering*, ou classification hiérarchique, développé par Eisen et al. À Stanford : le programme *Cluster* (Eisen et al., 1998). Il permet de regrouper les échantillons les plus similaires entre eux sur la base de leur profil d'expression des gènes analysés et regroupe également les gènes corrélés entre eux sur la base de leurs niveaux d'expression dans les échantillons étudiés. Le degré de similarité des échantillons ou des gènes entre eux est mesuré par le coefficient de corrélation de Pearson calculé sur la base des niveaux d'expression génique. Ces distances de corrélation permettent le regroupement des profils les plus similaires entre eux, sous la forme d'un arbre de distance de similarité, un dendrogramme, et les longueurs des

branches reliant les échantillons reflètent le degré de similarité entre ces échantillons. Un arbre du même type peut être tracé au niveau des gènes, en mesurant les corrélations d'expression des gènes à travers les différents échantillons. Une représentation graphique sous forme d'un tableau avec une échelle de couleur est ensuite couplée à cette classification et permet la visualisation de l'ensemble des résultats. Dans un tableau, les lignes correspondent aux gènes, les colonnes aux échantillons, et les niveaux de couleur reflètent de façon quantitative et qualitative les niveaux d'expression génique. Cette représentation permet une vision globale de l'information contenue dans la masse de données générées par la technique, mettant rapidement en évidence des groupes de gènes distinguant différentes classes d'échantillons. La comparaison de ces profils avec ceux de modèles simplifiés (lignées cellulaires « au repos » ou stimulées ou inhibées par divers produits) permet d'envisager une explication biologique à ces profils (Alizadeh et al., 2000).

III.3.5. Caractérisation des voies de régulation

Une autre application de l'étude du transcriptome concerne la caractérisation des voies de régulation (Wen et al., 1998 ; Roberts et al., 2000). En mesurant les niveaux de transcription génique dans un organisme dans différentes conditions, à différents stades du développement et dans différents tissus, on peut construire un profil d'expression génique qui caractérise la dynamique de chaque gène dans le génome. Par exemple, on va s'intéresser à des cinétiques de réponse à des effecteurs permettant de mettre en évidence des vagues successives de gènes activés ou réprimés sur la voie métabolique de la molécule ciblée, mais aussi sur d'autres voies coréglées (Liu et al., 2000 ; Walker et Rigley, 2000). L'utilisation de mutants d'inactivation génique ou de mutants bloquant la signalisation permet également de caractériser la succession des gènes au sein d'une cascade métabolique (Fambrough et al., 1999). Ce type d'analyse permet, non seulement d'établir une vision plus claire de l'activité cellulaire dans diverses conditions, mais aussi de déceler différents points potentiels d'intervention thérapeutique au sein de la cellule. D'autre part, la mise en évidence, grâce aux puces ADN, de gènes présentant des mécanismes communs de régulation est en mesure d'améliorer notre compréhension des signaux de régulation au niveau de l'ADN. En effet, il a été suggéré que des gènes dont le profil d'expression était similaire (gènes coexprimés) pouvaient avoir quelque chose en commun dans leur mécanisme de régulation, c'est-à-dire être coréglées, et que cette corégulation se faisait par l'intermédiaire de séquences cibles identiques pour des facteurs de transcription situées dans leur promoteur. Cela a été réalisé récemment dans le cadre de l'analyse des profils d'expression de 6220 ORF de la levure mesurés à 15 intervalles de temps différents durant le cycle mitotique (Tavazoie et al., 1999). Les auteurs ont identifié 30 groupes ou *clusters* de gènes coréglés avec, le plus souvent, des gènes de fonction similaire regroupés à l'intérieur d'un *cluster*. Puis ils ont cherché à identifier en amont des ORF des motifs d'ADN communs aux membres de chaque *cluster* et contribuant à leur corégulation. En utilisant un algorithme d'alignement de séquences, ils ont ainsi pu identifier 17 motifs dans 12 *clusters*, correspondant à 17 sites de fixation de facteurs de transcription dont certains étaient connus - permettant de valider l'approche - et dont d'autres constituaient des nouveaux motifs. Ce type d'approche, combinant puces ADN, algorithme de *clustering* et recherche systématique de motifs consensus d'ADN dans des promoteurs potentiels, peut donc permettre la détermination de réseaux de régulation transcriptionnelle dans des organismes séquencés (Brazma et Vilo, 2000).

III.3.6. Etude du transcriptome de cellules souches embryonnaires (CSE) et hématopoïétiques (CSH)

Les techniques d'analyse du transcriptome utilisées sont également différentes. Les puces ADN constituent de loin la technique la plus utilisée, une sélection représentative des publications analysant le transcriptome de CSE à l'aide de cette technique. (Sato et al., 2003 ; Rao, 2004 ; Bhattacharya et al., 2004 ; Abeyta et al., 2004 ; Zeng et al., 2004 et Bhattacharya et al., 2005). Le transcriptome des CSE a également été analysé en utilisant des EST des CSE humaines (Brandenberger et al., 2004) et des CSE murines (Sharov et al., 2003), avec la technique SAGE pour les CSE humaines (Richards et al., 2004) et pour les CSE murines (Anisimov et al., 2002a) et avec la technique MPSS pour les CSE humaines et murines (Wei et al., 2005). Ces techniques sont certes différentes, mais également complémentaires. Par exemple, le nombre d'EST séquencées (entre 10000 et 30000 par librairie pour les deux publications citées) est petit devant le nombre de tags SAGE (entre 70000 et 140000 par librairies pour les deux publications citées) ou MPSS (environ 2000000 par librairie dans la publication citée). L'image du transcriptome obtenue par le séquençage de ces EST est donc plus partielle que celle donnée par les librairies SAGE ou MPSS. Par contre, ces données de séquences d'EST ont été publiées, elles ont donc permis d'identifier des tags SAGE et MPSS qui n'auraient pas été identifiés dans le cas contraire. Certaines de ces séquences d'EST ont également été déposées sur des puces ADN (c'est le cas par exemple d'une partie des données d'EST obtenues dans l'étude de Sharov et al., (2003). Par ailleurs, les études par puces ADN n'ont permis d'étudier que l'expression des gènes dont les transcrits correspondants ont une sonde sur la puce : on ne peut donc pas étudier l'expression d'un certain nombre de gènes, comme par exemple le gène *Rex1* pour lequel aucune sonde n'est présente sur la puce Affymetrix U-133A utilisée par Sato et al. (2003). Au contraire, les techniques basées sur le séquençage (EST, SAGE et MPSS) ont permis de mettre en évidence l'expression de nombreux gènes pour lesquels les transcrits ne sont pas encore caractérisés. Cependant, un petit nombre de transcrits ne peuvent être observés par SAGE ou MPSS, ou ne peuvent pas être identifiés : on peut par contre étudier leur expression en utilisant des puces ADN. C'est par exemple le cas de l'ARNm humain *Nanog2*, dont le tag MPSS se trouve dans une région répétée, ne pouvant donc pas être identifié de manière unique.

Le transcriptome des CSH foetales murines, le transcriptome de CSH adultes murines a également été analysé (Park et al., 2002). Dans cette étude, les auteurs ont d'abord réalisé une première librairie d'ADNc à partir d'une population de CSH, et une seconde à partir d'une population contenant des progéniteurs multipotents dérivés de CSH. Les clones correspondants ont été transférés sur des filtres nylon, sur lesquels ont été hybridés des ADNc issus de cellules matures de thymus, de rate et de moelle osseuse. Seuls les clones qui n'étaient pas exprimés dans les différentes cellules matures ont ensuite été sélectionnés, afin de réaliser une puce ADN sur lame de verre. Sur cette puce, des ADNc issus de CSH et de progéniteurs multipotents dérivés de CSH ont ensuite été co-hybridés, ce qui a permis aux auteurs de mettre en évidence des gènes différentiellement exprimés entre des CSH et des progéniteurs multipotents, mais non exprimés dans des cellules matures.

D'autres études ont également permis d'étudier l'expression des gènes dans les CSH par des approches soustractives, dont notamment l'étude de Terskikh (Terskikh et al. (2001). En comparant les résultats de ces analyses du transcriptome des CSH, nous pouvons mettre en évidence des similarités : ces différentes approches mettent toutes en évidence une surexpression de gènes codant pour des protéines impliquées dans les voies de signalisation Wnt, TGF-, Sonic hedgehog et Notch dans des CSH. Cette observation a conduit différents auteurs à chercher si ces voies étaient impliquées dans la régulation de l'autorenouvellement des CSH.

Si ces études ont permis de mieux connaître les gènes exprimés dans les CSH, elles utilisent une approche soustractive pour enrichir les transcrits étudiés en transcrits spécifiquement exprimés dans les CSH, ce qui mène à une vision partielle du transcriptome de ces cellules. L'approche choisie par Akashi et al. (2003) permet quant à elle d'avoir une connaissance plus large du transcriptome des CSH (mais néanmoins restreinte aux gènes pour lesquels une sonde correspondante est présente sur la puce utilisée) : les auteurs ont utilisé des puces Affymetrix, pour étudier le transcriptome de populations de CSH, de progéniteurs multipotents, de progéniteurs communs lymphoïdes et de progéniteurs communs myéloïdes. Ils ont ainsi mis en évidence qu'un grand nombre de gènes que l'on pensait spécifiques des lignages myéloïdes et qu'un petit nombre de gènes que l'on pensait spécifiques des lignages lymphoïdes sont également exprimés dans les cellules souches, mais à un plus faible niveau. Les progéniteurs multipotents analysés expriment également un grand nombre de gènes des lignages myéloïdes, mais aussi lymphoïdes. Par contre, l'expression des gènes des lignages lymphoïdes (respectivement myéloïdes) n'est plus détectée dans les progéniteurs communs myéloïdes (respectivement lymphoïdes). On pourrait penser que ces résultats sont dus à des populations cellulaires initiales hétérogènes. Cependant, Miyamoto et al. (2002) ont abouti aux mêmes conclusions pour quelques gènes, en réalisant des PCR en temps réel à partir d'une seule cellule. Ces résultats suggèrent donc que les gènes exprimés par les cellules des différents lignages seraient déjà exprimés préalablement à l'engagement dans un lignage de différenciation. Ainsi l'engagement dans un lignage serait accompagné à la fois d'une augmentation de l'expression des gènes caractérisant ce lignage et d'une répression des gènes caractérisant les lignages alternatifs.

III.4. Applications dans le domaine agroalimentaire

III.4.1. Approche transcriptomique de l'engraissement chez le poulet de chair

L'augmentation du poids de tissu adipeux abdominal chez le poulet de chair est fortement corrélée à la sélection sur la vitesse de croissance. Cet engraissement excessif entraînant des pertes économiques à différents niveaux de la production, il est l'une des préoccupations de l'INRA. Le laboratoire de Génétique Animale a pour objectif, entre autres, de mettre en évidence des gènes impliqués dans la variabilité de l'état d'engraissement des poulets. Le modèle animal utilisé correspond à des lignées de poulets obtenues par sélection divergente sur ce caractère lignées Grasse et Maigre, (Leclercq et al. 1980). Parmi les approches mises en œuvre au laboratoire, un dispositif de Différentiel Display (Liang et al, 1992) a permis d'isoler, à partir d'ARN hépatiques de poulets, plusieurs centaines de fragments d'ADNc, correspondant à des gènes susceptibles de présenter une différence d'expression entre les poulets gras et maigres (Carré et al, 2001). La comparaison des séquences de ces ESTs (*Expressed Sequence*

Tags) avec les banques de données (Genbank, EMBL) a montré que ces gènes codaient des protéines impliquées dans plusieurs métabolismes, dont le métabolisme énergétique. Néanmoins, de nombreuses séquences n'ont pas pu être identifiées à ce jour.

Cette approche de display différentiel s'intègre dans l'analyse globale du transcriptome du poulet, permettant, à terme, la mise en évidence de gènes impliqués dans la variabilité de l'état d'engraissement du poulet de chair.

III.4.2. Analyse du transcriptome de microtubercules de pomme de terre

Une approche cDNA-AFLP (*complementary DesoxyriboNucleicAcid - Amplified Fragment Length Polymorphism*) a été développée afin d'étudier l'évolution du transcriptome de microtubercules de pomme de terre au cours de leur cycle de développement (principalement lors des étapes de tubérisation, de dormance et de germination) (Bachem et al., 1996 ; Trindade et al., 2004).

Des banques d'*Expressed Sequence Tag* (EST) construites à partir de stolons, de microtubercules en développement, de tubercules dormants et de germes en croissance ont été comparées par Ronning et al. (2003) et Flinn et al. (2005). Des études utilisant des puces ADN complémentaire ont finalement été réalisées par Kloosterman et al. (2005) afin d'évaluer l'expression génique au cours du développement des tubercules. Ces études présentent des perspectives techniques intéressantes, mais n'ont pas ciblé de façon quantitative le processus de vieillissement lors du stockage à basse température. Leurs résultats restent donc à l'heure actuelle d'un intérêt limité pour la mesure de l'âge physiologique des tubercules semences de pomme de terre.

III.5. Applications dans le domaine de l'environnement

Dans le domaine environnemental, la diversité spécifique est très importante. De plus, dans le cas de certains micro-organismes, beaucoup d'espèces sont difficilement cultivables. Les techniques moléculaires sont alors particulièrement bien adaptées. L'environnement contient un très grand nombre d'espèces inconnues (plus de 90% pour les bactéries). Les applications dans ce domaine sont très nombreuses et diverses. On distingue deux grands types d'applications: soit une approche populationnelle afin de caractériser un milieu (O'brien et al., 2005 ; Poretsky et al., 2005 ; Wu et al., 2002) soit on recherche des organismes spécifiques dans un échantillon. C'est cette dernière approche qui sera le thème central de projet : « Aquachip » pour l'identification de bactéries pathogènes des eaux.

III.5.1. Identification des bactéries pathogènes des eaux : le projet Aquachip

Les bactéries pathogènes des milieux aquatiques sont un problème majeur de santé publique (Straub et Chandler, 2003). Une surveillance des eaux constitue un enjeu de santé essentiel pour prévenir des maladies provoquées par les bactéries pathogènes véhiculées par les eaux, telles que des espèces de salmonelles, des pneumonies, de legionella et de vibrio choléra. Le respect des réglementations oblige les fournisseurs d'eau potable à une surveillance régulière.

Actuellement, les analyses bactériennes se fondent sur des méthodes de culture, c'est à dire le comptage de colonies sur boîte de Pétri. Le comptage sur culture et le «

most probable number (NPM) », sont les analyses menées dans le cadre des procédures standards employées par les laboratoires de toute l'Union Européenne.

Les mesures prises pour le contrôle de la qualité des eaux préconisent les analyses sur des organismes indicateurs d'une contamination fécale : les coliformes totaux, les coliformes fécaux (Rompré et al., 2002), *E. coli* et les streptocoques fécaux. Ces organismes ne sont pas pathogènes mais leur présence dans un échantillon peut traduire la présence d'autres espèces qui elles, sont dangereuses. Ce type restreint d'analyse est dû au fait qu'il est difficile de cultiver toutes les bactéries pathogènes.

Des expérimentations récentes mettent en évidence que la stratégie de contrôle de la présence de ces indicateurs bactériens par culture présente plusieurs défauts : ces indicateurs bactériens ne reflètent pas le risque réel pour l'homme puisque seules des bactéries indicatrices de contamination fécale et non les pathogènes eux-même sont détectées, la culture et l'obtention des résultats prennent un temps important (au moins trois à quatre jours), les modifications de l'environnement chez certaines bactéries peut induire une perte de leur faculté à pousser sur les milieux habituellement utilisés (Blackburn et Mccarthy, 2000 ; Ward et al., 1990), la détection de certains pathogènes spécifiques peut être longue (jusqu'à 15 jours).

Le développement de nouvelles méthodes d'analyses efficaces et rapides est alors un élément primordial pour la prévention des infections liées à l'eau. Le projet « Aquachip » est un projet européen créé pour le développement et la validation d'une technologie de puce ADN pour l'évaluation de la qualité bactériologique des eaux de baignade et de consommation. L'objectif du projet est de développer des méthodes moléculaires pour identifier et mesurer les bactéries pathogènes directement dans n'importe quel échantillon d'eau donné et d'évaluer leur état d'activité.

L'utilisation de techniques de détection moléculaire a été retenue car elles permettent de résoudre les problèmes évoqués précédemment pour les analyses par cultures: les bactéries pathogènes sont directement identifiées sans l'aide des indicateurs, le temps d'analyse est considérablement réduit (de 4 jours à moins d'une journée), l'activité bactérienne peut être quantifiée.

L'utilisation des techniques de PCR semble être une bonne solution pour détecter les bactéries pathogènes (Kingombe et al., 1999 ; Malorny et Hoorfar, 2005 ; Paton et Paton, 2005). Les bactéries pathogènes présentent des risques majeurs pour l'homme, même si elles sont très largement sous-représentées au sein d'un échantillon. Il faut donc passer par une phase d'amplification du signal pour arriver à les détecter et donc calculer en plus du rapporteur pour la puce ADN un couple d'amorces pour l'amplification par PCR.

La technologie quantitative de PCR (Fey et al., 2004) est combinée à la technologie de puce ADN pour rechercher des séquences spécifiques de gène et pour évaluer leur quantité dans des échantillons d'eau. Des séquences d'ARN (ARNr 16S principalement) servent à identifier l'espèce ou le genre. La présence et la quantité de ces ARN servent aussi d'indicateurs de l'activité et la pathogénicité bactérienne afin d'évaluer les risques infectieux. La finalité est de créer une nouvelle technologie pour la détection des bactéries pathogènes des eaux et de valider son application de masse.

DISCUSSION
ET
CONCLUSION

Comme nous l'avons précisé précédemment, la transcriptomique est une approche qui s'appuie sur différentes techniques que se soit classiques ou nouvelles, possèdent des avantages très larges et des applications dans plusieurs domaines, mais aussi il y a des inconvénients. Le progrès de cette approche dépend du développement de ces techniques pour mieux étudier l'expression des gènes.

La génomique est à la base des recherches en biotechnologie. C'est l'étude exhaustive des génomes c'est-à-dire de l'ensemble des gènes, de leur cartographie, de leur séquence et de leur rôle fonctionnel. La génomique permet d'identifier les gènes liés à certaines maladies. En complément, la post-génomique va plus loin dans la compréhension du fonctionnement de la cellule. Elle englobe notamment l'étude des ARN transcrits ou transcriptomiques, l'étude des protéines exprimées par un génome ou protéomique.

La technologie de PCR en temps réel est aujourd'hui l'outil de choix qui a permis de simplifier et de quantifier le nombre de molécules ou de copies d'un gène ou d'un messager d'intérêt dans un échantillon donné. Ses performances peuvent se résumer en quatre mots : spécificité, sensibilité, bonne reproductibilité et linéarité sur une gamme dynamique comprise entre 10 et 10^8 copies. Il faut ajouter à ses nombreuses qualités, la rapidité de la technique sans nécessité d'une étape post-PCR. La méthode est automatisable, et par conséquent parfaitement adaptée à une analyse de routine. Par ailleurs, la réaction s'effectuant en tube fermé, le risque de contamination est faible.

Si le déroulement de la réaction de PCR est relativement bien connu, celui de la réaction de rétrotranscription reste largement mystérieux faute de modélisation. La qualité des résultats obtenus par RT-PCR dépend de nombreuses étapes difficilement contrôlables et se déroulant en aveugle car il n'existe pas à l'heure actuelle de techniques permettant de connaître l'efficacité de la réaction. De ce fait, la mise au point reste largement empirique. L'étape de PCR étant mieux connue et plus contrôlable que la réaction de RT, la transcription inverse des ARN en ADNc reste donc l'étape limitante de la réaction de RT-PCR.

L'utilisation accrue de techniques quantitatives, grâce en particulier au développement récent de la PCR en temps réel, a permis de prendre conscience de ces difficultés. En fournissant des résultats plus précis qu'une simple révélation en point final, elle a fait apparaître la reproductibilité limitée de la technique de RT-PCR. Il est devenu évident que la qualité de l'ARN obtenu est très différente d'un échantillon à l'autre, surtout lorsque l'on s'adresse à des échantillons biologiques. Il est clair également que l'efficacité de la réaction de RT est extrêmement variable et dépend étroitement de la qualité des réactifs utilisés. L'évaluation de la qualité de l'ARN et de l'étape de RT par analyse d'un transcrit référent endogène adéquat est donc primordiale, même lorsque l'on ne cherche pas à obtenir de résultat quantifié, afin de limiter les résultats faussement négatifs. Il est également nécessaire d'entreprendre un important travail de standardisation afin d'améliorer la fiabilité des résultats. Cette standardisation ne pourra probablement se faire qu'au prix de l'automatisation des différentes étapes, de l'extraction de l'ARN jusqu'à la révélation de la PCR. Elle doit s'accompagner du développement de contrôles de qualité permettant à chacun de s'évaluer et de contrôler régulièrement ses résultats.

Il faut également souligner que le soin apporté aux étapes situées en amont de la RT-PCR, telles que les modalités de prélèvement, sa conservation pendant l'acheminement au laboratoire, vont en grande partie conditionner la qualité des résultats obtenus (Bastard et al., 2002). En effet, l'ARN est un matériau biologique fragile, mais également dynamique et des modifications de la quantité ou même de la nature des transcrits peuvent se produire *ex vivo* (Bolduc et al., 2001 ; Ars et al., 2000). Ne pas tenir compte de ces aspects expose à des artefacts ou même des erreurs diagnostiques (Gayther et al., 1997 ; Birrel et al., 2001), quelle que soit la qualité technique de la réaction de RT-PCR.

D'autres techniques basées sur la PCR ont également été développées, mais elles ont peu été utilisées (Green et al., 2001). Si les méthodes basées sur la PCR ont été particulièrement utilisées au début des années 1990, elles sont actuellement moins utilisées, au profit de méthodes quantitatives et se rapprochant de plus près d'une analyse exhaustive du transcriptome. Ces dernières méthodes sont cependant plus lourdes à mettre en œuvre et plus onéreuses que les techniques précédentes, ce qui explique pourquoi les premières techniques sont encore parfois utilisées aujourd'hui.

Les puces ADN ont été qualifiées de méthodes "closes" d'analyse du transcriptome, dans la mesure où l'espace des gènes analysés est fini, limité à ceux pour lesquels une sonde correspondante est disponible sur la puce (Green et al., 2001). Au contraire, d'autres méthodes sont dites "ouvertes" car elles n'utilisent pas de connaissance a priori sur les transcrits, elles permettent donc la découverte de nouveaux gènes exprimés, et la réalisation d'analyses du transcriptome chez des espèces pour lesquelles la connaissance des gènes exprimés est encore limitée.

Si l'utilité des puces ADN pour l'analyse du transcriptome dans un large spectre d'applications n'est plus à prouver étant donné la quantité et la qualité de publications à ce sujet, il est toutefois important de connaître les sources de biais ou de variabilité pouvant affecter la mesure de l'expression des gènes. Au niveau de la puce, il peut y avoir des irrégularités spatiales, et dans le cas des sondes déposées par un robot, il a été montré que les différentes aiguilles de dépôt pouvaient induire des différences dans les résultats obtenus (Smyth et Speed, 2003). De manière générale, pour certaines puces, il a été montré que la position relative des sondes sur la puce influe sur les résultats obtenus. En effet, les gènes correspondant aux sondes les plus proches sur la puce sont plus souvent co-exprimés qu'attendu (Kluger et al., 2003). Il est donc particulièrement important de répartir aléatoirement les sondes sur la puce, et de mesurer l'expression de chaque gène en utilisant plusieurs sondes différentes présentes à différents endroits de la puce. La localisation sur la puce et éventuellement l'aiguille de dépôt utilisée peuvent également être pris en compte dans l'étape de normalisation des données (Smyth et Speed, 2003).

De nombreuses puces ont été développées avec des jeux de sondes spécifiques pour étudier le transcriptome d'organismes divers. Actuellement, des chercheurs essaient 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 (Van Dam et Quake, 2002). 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 (Roth et al., 2004).

Au niveau des cibles, la réaction de marquage est particulièrement sensible à différents paramètres, pouvant conduire à un marquage différentiel de différents ARNm (Hoen et al., 2003). L'hybridation doit ensuite être réalisée simultanément pour l'ensemble des sondes. Si elles sont choisies de manière à présenter les paramètres thermodynamiques les plus proches, elles ont néanmoins une composition différente, qui influence la stabilité des duplex sondes/cibles spécifiques ou aspécifiques. C'est pour cela que les puces ne peuvent être utilisées que pour une mesure relative de l'expression des gènes entre différentes conditions analysées avec le même type de puce. Dans le cas où deux populations de cibles différentes sont hybridées sur la même puce, des fluorochromes différents sont utilisés, pouvant induire des différences artefactuelles : notamment des différences d'incorporation lors du marquage, des stabilités différentes de ces fluorochromes, et des différences entre la lecture de ces deux intensités de fluorescence par le scanner utilisé (Leung et Cavalieri, 2003). En général, deux puces différentes sont donc réalisées pour comparer les deux mêmes conditions, en inversant les fluorochromes utilisés pour chaque population d'ARNm. La comparaison entre les résultats obtenus avec ces deux puces permet d'estimer l'effet dû au fluorochrome. Ce biais doit ensuite être pris en compte dans les estimations des rapports d'expression. Des biais ont également été mis en évidence lors de la lecture des intensités de fluorescence : on observe souvent une saturation du signal pour les sondes correspondant à des gènes fortement exprimés (Leung et Cavalieri, 2003). Mais la lecture du signal est surtout problématique pour les gènes les moins exprimés, dans la mesure où le transcriptome comprend un très grand nombre de gènes peu exprimés : pour ces gènes, il est particulièrement difficile de distinguer le signal du bruit de fond optique sur la puce.

Une force majeure de la méthode SAGE est la possibilité d'analyser de façon simultanée et quantitative un grand nombre de gènes et même, en théorie, l'intégralité du transcriptome. La sensibilité et la fiabilité des résultats obtenus par SAGE ont été montrées de nombreuses fois, notamment par des méthodes indépendantes d'analyse d'expression comme le criblage d'une banque d'ADNc (Velculescu et al., 1995), le northern blot (Chrast et al., 2000 ; Velculescu et al., 1997), la RT-PCR quantitative (Hashimoto et al., 2000), l'hybridation *in situ* (Datson et al., 2001; Feldker et al., 2003) et les puces ADN (Nacht et al., 1999). Un des autres atouts de SAGE est que les niveaux d'expression des transcrits sont représentés par un nombre absolu de tags ce qui facilite le partage et la comparaison des données entre différentes expériences et différents groupes de recherche, si l'on s'assure que les données sont issues du même organisme et que le nombre total de tags est pris en considération. A cette fin, les nombreuses données SAGE générées ces dix dernières années ont été collectées et déposées sur une base de données publique.

Bien que SAGE soit une technique puissante d'analyse de l'expression des gènes, elle possède certaines limites (Stollberg et al., 2000). Premièrement, la construction de banques SAGE et le séquençage des concatémères sont coûteux en temps et en argent. Ensuite, bien que la profondeur d'analyse soit théoriquement infinie, l'étude de certains tissus complexes semble limitée aux transcrits d'abondance moyenne à haute (Datson et

al., 2001 ; Feldker et al., 2003). Pour détecter des transcrits faiblement abondants, il faudrait séquencer environ trois « équivalents transcriptome », c'est-à-dire près d'un million de tags (Clarke et Carbon, 1976). De plus, certains transcrits ne possédant pas le site de coupure de l'enzyme d'ancrage échapperont à l'analyse. Enfin, les contraintes liées au caractère fastidieux de la construction des banques autant qu'à un accès en général limité au séquençage automatique bon marché ont limité l'utilisation de SAGE. Cependant, l'émergence récente de nouvelles méthodes de séquençage de l'ADN à très haut débit capable de déchiffrer jusqu'à 25 millions de bases en quelques heures (Margulies et al., 2005) pourrait être à l'origine d'un regain d'intérêt pour SAGE.

Nous avons donc vu qu'il existe différentes méthodes pour analyser le transcriptome. Afin d'en réaliser une analyse quantitative, il est possible d'utiliser des puces à ADN, ou des méthodes basées sur le séquençage d'étiquettes de séquences transcrites (SAGE et MPSS). Si les méthodes basées sur le séquençage n'utilisent pas de connaissance a priori sur les gènes exprimés, ce n'est pas le cas pour les puces ADN, où l'analyse est limitée aux gènes pour lesquels une sonde correspondante est présente sur la puce. Les méthodes basées sur le séquençage présentent donc l'avantage de pouvoir mettre en évidence de nouvelles séquences transcrites. Par contre, elles sont plus chères à mettre en œuvre. La question de l'utilisation de puces ADN ou de techniques basées sur le séquençage est souvent sujette à débats, des arguments pouvant être avancés en faveur de chacune des techniques selon la problématique biologique posée et les moyens dont on dispose pour y répondre. Cependant, ces méthodes ne sont pas toujours mutuellement exclusives, elles peuvent également être complémentaires. Il est par exemple possible d'utiliser initialement une méthode basée sur le séquençage afin d'avoir une vision suffisamment large et sans a priori sur le transcriptome. On peut alors sélectionner un certain nombre de gènes qui semblent intéressants par rapport à la question biologique posée, afin de déposer des sondes correspondantes sur une puce, et d'étudier l'expression de ces gènes dans un plus grand nombre de situations biologiques différentes (par exemple pour évaluer la variabilité de l'expression de ces gènes en fonction des échantillons biologiques ou au cours du temps). On peut donc se demander avec quelle fiabilité ces méthodes reflètent l'expression des gènes. Pour cela, on peut s'intéresser d'une part à la reproductibilité de ces méthodes d'analyse (qui ne révélera certes pas la présence d'éventuels biais systématiques) et d'autre part à la comparaison entre les estimations de l'expression des gènes fournies par ces différentes méthodes.

Les méthodes d'analyse parallèle de transcriptome ont effectué une réelle percée en 1995. Dans un premier temps, les discussions concernant ces nouvelles approches portaient essentiellement sur des aspects techniques : sensibilité, reproductibilité, coût, flexibilité. Depuis, leur utilité a été établie et les biologistes, les mathématiciens et les informaticiens se penchent désormais sur l'éventail des applications possibles et inventent ou actualisent des méthodes d'analyse adaptées au volume et à la complexité des résultats obtenus (Eisen et al., 1998 ; Bassett et al., 1999). Ce champ d'investigation qui s'ouvre à peine a pour but de mettre en évidence les liens qui révèlent les structures sous-jacentes des systèmes étudiés. A l'avenir, les données d'expression pourront être exploitées en conjonction avec l'ensemble des informations biologiques disponibles -dont beaucoup proviendront de programmes de recherche en génomique- afin de mieux formuler les hypothèses à tester expérimentalement (Vidal, 2001).

Le développement de nouveaux outils biologiques a permis d'envisager une étude à grande échelle du fonctionnement cellulaire. Ce projet révolutionnaire et ambitieux repose sur l'étude de la cellule à deux niveaux : d'une part, l'analyse du génome dans sa globalité, réalisée par des techniques de séquençage à haut débit et, d'autre part, l'étude de l'expression de ces gènes soit par analyse de l'expression des ARN messagers (transcriptome), soit par analyse de l'expression des protéines correspondant à ces ARN messagers (protéome). Les informations fournies par ces études suscitent de grands espoirs dans les domaines de l'identification de nouvelles cibles thérapeutiques et de la découverte de nouveaux médicaments. Les analyses à échelle génomique et postgénomique sont complémentaires. En effet, les phénomènes de régulation de l'expression des gènes au niveau du promoteur et les modifications post-transcriptionnelles et post-traductionnelles sont à l'origine de modulations des voies de signalisation cellulaire. En couplant les données issues de l'étude des gènes et de leurs produits (ARN messagers, protéines), il est possible d'appréhender de façon globale le fonctionnement cellulaire.

Les avantages des méthodes d'étude des ARNm sont qu'elles sont en général moins difficiles, plus sensibles et impliquent des approches à plus haut débit que les méthodes d'étude des protéines. Par ailleurs, la mesure des ARNm apporte des informations directes sur l'état d'activation des gènes. De plus, la pertinence des approches transcriptomiques pour l'exploration fonctionnelle est démontrée, notamment par des études établissant une relation entre le transcriptome d'un tissu ou d'un organe et sa physiologie (Son et al., 2005 ; Zhang et al., 2004). Enfin, seule l'étude des transcriptomes permet de prendre en compte la population des ARN non-codants dont l'importance fonctionnelle est de plus en plus évidente (Kapranov et al., 2007).

Une des applications de l'analyse des profils d'expression est l'amélioration du diagnostic et pronostic clinique. Les études du transcriptome humain réalisées sur les cancers (Golub et al., 1999 ; Sorlie et al., 2001 ; van de Vijver et al., 2002) et les hémopathies (Alizadeh et al., 2000) ont apporté des résultats essentiels à la compréhension de ces pathologies. Des profils d'expressions, caractéristiques de certaines tumeurs, ont permis d'améliorer les classifications cliniques, parfois insuffisantes (Sorlie et al., 2001; Bertucci et al., 2004). Grâce aux puces ADN, il est désormais possible de distinguer différents types d'hémopathies aux pronostics de survie différents (Alizadeh et al., 2000). De la même manière, les portraits moléculaires de patientes atteintes du cancer du sein ont mis en évidence 70 marqueurs pronostiques (gènes) jugés plus pertinents que les paramètres clinicobiologiques existants (van de Vijver et al., 2002). Plus récemment, l'utilisation des puces ADN pour étudier les maladies cardio-vasculaires apporte des résultats encourageants (Steenman et al., 2003; Napoli et al., 2003; Liew et Dzau, 2004). Les profils moléculaires de patients en insuffisance cardiaque semblent pouvoir affiner les classifications cliniques (Liew et Dzau, 2004; Steenman et al., 2005). Ces études ont également permis d'identifier de nouveaux facteurs génétiques et environnementaux impliqués dans les problèmes d'insuffisance cardiaque comme l'hypertrophie, l'infarctus ou l'ischémie. Parmi les facteurs environnementaux, ces travaux citent entre autres les thérapies et l'alcool. Les facteurs génétiques ont des origines héréditaires ou sont les conséquences d'autres dysfonctionnements (maladie des artères coronaires, dysfonctionnement valvulaire, hypertension ...). Finalement, ces nouveaux marqueurs offrent de nouvelles pistes pour le

diagnostic et le pronostic de ces pathologies ainsi que pour la recherche de nouvelles cibles thérapeutiques (Liew et Dzau, 2004).

Le fonctionnement d'une cellule implique de nombreux mécanismes complexes que les biologistes cherchent à appréhender depuis de longues années. Les évolutions technologiques récentes ont rendu possible l'accès simultané à la connaissance de la concentration d'un grand nombre de transcrits ou de polypeptides, définis respectivement comme le transcriptome et le protéome. La connaissance de ces deux entités doit permettre à terme d'étudier précisément la manière dont est exprimé le génome et de répondre à de très nombreuses problématiques. En effet, cela permet de trouver des cibles thérapeutiques, d'identifier des marqueurs pour un diagnostic rapide, d'étudier les processus biologiques comme ceux impliqués dans la sporulation de la levure (Chu et al., 1998) ou les mécanismes de différenciation dans les tumeurs humaines (Perou et al., 2000), de déterminer l'effet de la mutation d'un gène régulateur, d'analyser sous quelles conditions environnementales, un gène est activé et comment il évolue.

Les questions auxquelles les études de transcriptome apportent des réponses sont très variées. On peut citer la variété suivante des thématiques concernant les plantes : la recherche de gènes dont l'expression varie au cours du développement ou en réponse à une modification de l'environnement et étude des mécanismes biologiques affectés (Girke et al., 2000 ; Schaffer et al., 2001 ; Wang et al., 2000), la recherche de gènes dont les profils d'expression sont proches et soumis aux mêmes circuits de régulation (Cho et al., 1998 ; Spellman et al., 1998), l'association de profils d'expression à des classes de mutants ou des types de cancers (Alizadeh et al., 2000 ; Ross et al., 2000) et construction d'indices de diagnostic médical (Golub et al., 2000), l'étude des fonctions de gènes orphelins ou de l'effet de molécules pharmacologiques s'appuyant sur de vastes bases de données d'expression (Hughes et al., 2000).

A la fin, la transcriptomique, comme la génomique, ouvre de nouveaux horizons dans la biologie. De nouveaux domaines, verront sans doute le jour grâce aux données accumulées sur l'expression des gènes. Il est cependant difficile d'en déterminer leur étendue. Il est néanmoins raisonnable de reconnaître que ces données d'expression de l'ensemble des gènes sont une nouvelle étape dans la compréhension du fonctionnement du système dynamique qu'est la cellule. La vision ouverte par le transcriptome est beaucoup plus dynamique et fonctionnelle que l'est la vision issue de la génomique. Au fur et à mesure des progrès techniques, on se rapproche ainsi des réseaux fonctionnels de l'organisme. Certaines personnes pensent que la prochaine étape sera la connaissance de l'ensemble des protéines avec leurs différentes formes présentes au sein d'une cellule à un temps donné. Ces protéines représentent l'ensemble de la machinerie cellulaire, base des différents métabolismes présents. Cependant, le protéome ne représente pas l'ultime étape de la connaissance en biologie. Il serait plus juste de dire que le transcriptome et le protéome sont complémentaires. En effet, les protéines sont stockées au sein de la cellule, ce que ne sont pas les ARNm. Les protéines donnent donc l'état cellulaire à un moment donné tandis que le transcriptome renseigne sur la dynamique du système et des réponses biologiques aux différents facteurs extérieurs.

RÉFÉRENCES
BIBLIOGRAPHIQUES

Abeyta MJ, Clark AT, Rodriguez RT, Bodnar MS, Pera RA et Firpo MT. (2004) Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet*; 13(6): 601-608.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferreira S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM et Venter JC. (2000) the genome sequence of *Drosophila melanogaster*. *Science*; 287: 2185-2195.

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C et Venter JC. (1992) Sequence identification of 2,375 human brain genes. *Nature*; 355: 632-634.

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*; 252: 1651-1656.

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*; 377: 3-174.

Adams MD, Soares MB, Kerlavage AR, Fields C et Venter JC. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat Genet*; 4: 373-380.

Akashi K, He X, Chen J, Iwasaki H, Niu C, Steenhard B, Zhang J, Haug J et Li L. (2003) Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood*; 101(2): 383-389.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldriche JC, Sabet H, Tron T, Xu X, Powell JL, Xong L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chon Wc, Greiner Tc, Weisenburger DD, Armitage Jo, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstien D, Brown PO et Staudt LM. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*; 403: 503-511.

- Almeida R, Norrish A, Levick M, Vetrie D, Freeman T, Vilo J, Ivens A, Lange U, Stober C, McCann S et Blackwell JM. (2002) From genomes to vaccines : Leishmania as a model. *Philos Trans R Soc Lond B Biol Sci*; 357: 5-11.
- Alwine JC, Kemp DJ et Stark GR. (1997) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the USA*; 74(12): 5350-5354.
- Anisimov SV, Tarasov KV, Stern MD, Lakatta EG et Boheler KR. (2002a) A quantitative and validated SAGE transcriptome reference for adult mouse heart. *Genomics*; 80(2): 213-222.
- Ars E, Serra E, de la Luna S, Estivill X et Lázaro C. (2000) Cold shock induces the insertion of a cryptic exon in the neurofibromatosis type 1 (NF1) mRNA. *Nucl Ac Res*; 28: 1307-1312.
- Audic S et Claverie JM. (1997) The significance of digital gene expression profiles. *Genome Res*; 7: 986-995.
- Bachem CW, Vander Hoven RS, De Bruijn SM, Vrengdenhil D, Zabeau M et Visser RG. (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J*; 9(5): 745-753.
- Bassett DE, Eisen MB et Boguski MS. (1999) Gene expression informatics - it's all in your mine. *Nature Genet*; 21: 51-55.
- Bastard JP, Chambert S, Ceppia F. (2002) Les méthodes d'extraction et de purification des ARN. *Ann Biol Clin*; 60: 513-523.
- Bellatin JA, Murray AS, Zhao M et McMaster WR. (2002) *Leishmania mexicana*: identification of genes that are preferentially expressed in amastigotes. *Exp Parasitol*; 100: 44-53.
- Bellis M et Casellas P. (1997) La puce à ADN : un multi-réacteur de paillasse. *Médecine/Sciences*; 13: 1317-1324.
- Bertucci F, Finetti P, Rougemont J, Charaf-Jauffret E, Nasser V, Loriod B, Camerlo J, Tagett R, Tarpin C, Houvenaeghel G, Nquven C, Maraninchi D, Jacquemier J, Houlgatte R, Birnbaum D et Viens P. (2004) Gene expression profiling for molecular characterization of inflammatory breast cancer and prediction of response to chemotherapy. *Cancer Res*; 64: 8558- 8565.
- Bertucci F, Houlgatte R, Benziane A, Granjeaud S, Adelaide J, Tagett R, Loriod B, Jacquemier J, Viens P, Jordon B, Birnbaum D et Nguyen C. (2000) Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet*; 9: 2981-2991.
- Bertucci F, Salas S, Nasser V, Houlgatte R, Birnbaum D et Xerri L. (2002) Profils d'expression génique par puces à ADN dans les lymphomes malins : intérêt pronostique. *Bull Cancer*; 89: 661-665.
- Bertucci F, Van Hulst S, Bernard K, Loriod B, Granjeaud S, Tagett R, Starkey M, Nguyen C, Jordon B et Birnbaum D. (1999) Expression scanning of an array of growth control genes in human tumor cell lines. *Oncogene*; 18: 3905-3912.
- Beverley SM, Akopyants NS, Goyard S, Matlib RS, Gordon JL, Browstein BH, Stormo GD, Bukanova EN, Hott CT, Li F, MacMillian S, Muo JN, Schwertman LA, Smeds MR et Wang Y. (2002) Putting the *Leishmania* genome to work: functional genomics by transposon trapping and expression profiling. *Philos Trans R Soc Lond B Biol Sci*; 357: 47-53.

Bhattacharya B, Cai J, Luo Y, Miura T, Mejido J, Brimble SN, Zeng X, Schulz TC, Rao MS et Puri RK. (2005) Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies. *BMC Dev Biol*; 5(1): 22.

Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, Joshi BH, Ginis I, Thies RS, Amit M, Lyons I, Condie BG, Itskovitz-Eldor J, Rao MS et Puri RK. (2004) Gene expression in human embryonic stem cell lines : unique molecular signature. *Blood*; 103(8): 2956-2964.

Birrel GW, Ramsay JR, Tung JJ et Lavin MF. (2001) Exon skipping at the ATM gene in normal individuals: the effect of blood sample storage on RT-PCR analysis. *Hum Mut*; 17: 75-76.

Blackburn CW et Mccarthy JD. (2000) Modifications to methods for the enumeration and detection of injured *Escherichia coli* O157:H7 in foods. *International Journal Of Food Microbiology*; 1-3: 285-290.

Blalock EM, Chen KC, Sharrow K, Herman JP, Porter NM, Foster TC et Landfield PW. (2003) Gene microarrays in hippocampal aging: statistical profiling identifies novel processes correlated with cognitive impairment. *J Neurosci*; 23: 3807-3819.

Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR et Landfield PW. (2004) Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses; *Proc Natl Acad Sci U S A*; 101: 2173- 2178.

Bolduc L, Labrecque B, Cordeau M, Blanchette M et Chabot B. (2001) Dimethylsulfoxyde affects the selection of splice sites. *J Biol Chem*; 276: 17597-17602.

Bolufer P, Sanz GF, Barragan E, Sanz MA, Cervera J, Lerma E, Senent L, Moreno J et Planelles MD. (2000) Rapid quantitative of BCRABL transcripts in chronic myeloid leukemia patients by real-time reverse transcriptase polymerase-chain reaction using fluorescently labeled probes. *Haematologia*; 85: 1248-1254.

Boucher N, Wu Y, Dumas C, Dube M, Sereno D, Breton M et Pajadopoulos B. (2002) A common mechanism of stage regulated gene expression in *Leishmania* mediated by a conserved 3'- untranslated region element. *J Biol Chem*; 277: 19511-19520.

Brandenberger R, Wei H, Zhang S, Lei S, Murage J, Fisk GJ, Li Y, Xu C, Fang R, Guegler K, Rao MS, Mandalam R, Lebkowski J et Stanton LW. (2004) Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol*; 22(6):707-716

Brazma A et Vilo J. (2000). Gene expression data analysis. *FEBS Lett*; 480: 17-24.

Brent R. (2000) Genomic biology. *Cell*; 100(1): 169-183.

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J et Corcoran K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*; 18(6): 630-634.

Brown PO et Botstein D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genetics* ;21: 33-37.

Carré W, Diot C, Fillon V, Crooijmans RP, Lagarrigue S, Morrison M, Vignal A, Groenen MA et Douaire M. (2001) *Anim. Genet*; 32(5): 289-297.

- Castelli V, Aury J, M, Jaillon O, Wincker P, Clepet C, Menard M, Cruaud C, Quetier F, Scarpelli C, Schachter V, Temple G, Caboche M, Weissenbach J et Salanoubat M. (2004) Whole genome sequence comparisons and "full-length" cDNA sequences: a combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res*; 14: 406-413.
- Chan EY, Luthi-Carter R, Strand A, Solano SM, Hanson SA, DeJohn MM, Kooperberg C, Chase KO, DiFiglia M, Young AB, Leavitt BR, Cha JH, Aronin N, Hayden MR et Olson J M. (2002) Increased huntingtin protein length reduces the number of polyglutamine-induced gene expression changes in mouse models of Huntington's disease. *Hum Mol Genet*; 11: 1939-1951.
- Charest H, Zhang WW et Matlashewski G. (1996) The developmental expression of *Leishmania donovani* A2 amastigote-specific genes is post-transcriptionally mediated and involves elements located in the 3'- untranslated region. *J Biol Chem*; 271: 17081-17090.
- Chrast R, Scott HS, Pappasavvas MP, Rossier C, Antonarakis ES, Barras C, Davisson MT, Schmidt C, Estivill X, Dierssen M, Pritchard M, et Antonarakis SE. (2000). The mouse brain transcriptome by SAGE: differences in gene expression between P30 brains of the partial trisomy 16 mouse model of Down syndrome (Ts65Dn) and normals. *Genome Res*; 10: 2006-2021.
- Chen T, He HL et Church GM. (1999) Modeling gene expression with differential equations. *Pac Symp Biocomput*; 29-40.
- Chen Y, Dougherty ER et Bittner ML. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J Biomed Opt*; 2: 364-374.
- Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockart DJ et Davis RW. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell*; 2: 65-73.
- Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO et Herskowitz I. (1998) The transcriptional program of sporulation in budding yeast. *Science*; 282: 699-705.
- Chudin E, Walker R, Kosaka A, Wu SX, Rabert D, Chang T et Kreder D. (2002) Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. *Genome Biology*; 3(1): research0005.
- Clark EA, Golub TR, Lander ES et Hynes RO. (2000) Genomic analysis of metastasis reveals an essential role for RhoC. *Nature*; 406: 532-535.
- Clarke L et Carbon J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E. coli* genome. *Cell* 9: 91-99.
- Clayton CE. (2002) Life without transcriptional control? From fly to man and back again. *EMBO J*; 21: 1881-1888.
- Datson NA, van der Perk-de Jong J, van den Berg MP, de Kloet ER et Vreugdenhil E. (1999) MicroSAGE : a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res*; 27: 1300-1307.
- Datson NA, van der Perk J, de Kloet ER et Vreugdenhil E. (2001). Expression profile of 30,000 genes in rat hippocampus using SAGE. *Hippocampus* ; 11: 430-444.

- Debouck C et Goodfellow PN. (1999) DNA microarrays in drug discovery and development. *Nature Genet*; 21: 48-50.
- DeRisi JL, Iyer VR. et Brown PO. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*; 278: 680-686.
- DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA et Trent JM. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet*; 14: 457-460.
- DeSaizieu A, Certa U, Warrington J, Gray C, Keck W et Mous J. (1998) Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nature biotechnology*; 16(1): 45-48.
- Desplats PA, Kass KE, Gilmartin T, Stanwood GD, Woodward EL, Head SR, Sutcliffe JG et Thomas EA. (2006). Selective deficits in the expression of striatal-enriched mRNAs in Huntington's disease. *J Neurochem*; 96: 743-757.
- D'haeseleer P, Liang S et Somogyi R. (1999) Gene expression data analysis and modeling. *Pac Symp Biocomput*; Tutorial: 2-34.
- Dickey CA, Loring JF, Montgomery J, Gordon MN, Eastman PS et Morgan D. (2003) Selectively reduced expression of synaptic plasticity-related genes in amyloid precursor protein + presenilin-1 transgenic mice. *J Neurosci*; 23: 5219-5226.
- Diehl S, Diehl F, El-Sayed N, Clayton C et Hoheisel J. (2002) Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA-microarray. *Mol Biochem Parasitol*; 123: 115.
- Duggan DJ, Bittner M, Chen Y, Meltzer P et Trent JM. (1999) Expression profiling using cDNA microarrays. *Nat. Genetics*; 21: 10-14.
- Eisen MB, Spellman PT, Brown PO et Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*; 95: 14863-14868.
- Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y et al. (1998) Data management and analysis for gene expression arrays. *Nature Genet*; 20: 19-23.
- Fambrough D, McClure K, Kazlauskas A et Lander ES. (1999) Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell*; 97: 727-741.
- Feldker DE, Datson NA, Veenema AH, Meulmeester E, de Kloet ER et Vreugdenhil E. (2003). Serial analysis of gene expression predicts structural differences in hippocampus of long attack latency and short attack latency mice. *Eur J Neurosci*; 17: 379-387.
- Fey A, Eichler S, Flavier S, Christen R, Höfle MG et Guzmán CA. (2004) Establishment of a real-time PCR-based approach for accurate quantification of bacterial RNA targets in water, using *Salmonella* as a model organism. *Applied And Environmental Microbiology*; 6: 3618-3623.
- Flinn B, Rothwell C, Griffiths R, Lägue M, DeKoeper D, Sardana R, Audy P, Goyer C, Li XQ, Wang-Pruski G et Regan S. (2005) Potato expressed sequence tag generation and analysis using standard and unique cDNA libraries. *Plant Mol. Biol*; 59(3): 407-433.

- Gayther SA, Barski P, Batley SJ, Li L, de Foy KA, Cohen SN, Pander BA et Caldas C. (1997) Aberrant splicing of the TSG101 and FHIT genes occurs frequently in multiple malignancies and in normal tissues and mimics alterations previously described in tumors. *Oncogene*; 15: 2119-2126.
- Gibson UEM, Heid CA et Williams PM. (1996) A novel method for real-time RT-PCR. *Genome Res*; 6: 995-1001.
- Ginsberg SD, Hemby SE, Lee VM, Eberwine JH et Trojanowski JQ. (2000) Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann Neurol* 48; 77-87.
- Girke T, Todd J, Ruuska S, White J, Benning C et Ohlrogge J. (2000) Microarray analysis of developing Arabidopsis seeds. *Plant Physiol*; 124: 1570-1581.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD et Lander ES. (2000) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*; 286: 531-537.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri Ma et Bloomfield CD. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*; 286: 531-537.
- Green CD, Simons JF, Taillon BE et Lewin DA. (2001) Open systems : panoramic views of gene expression. *J Immunol Methods*; 250(1-2): 67-79.
- Grunblatt E, Mandel S, Jacob-Hirsch J, Zeligson S, Amariglio N, Rechavi G, Li J, Ravid R, Roggendorf W, Riederer P et Youdim MB. (2004) Gene expression profiling of parkinsonian substantia nigra pars compacta; alterations in ubiquitin-proteasome, heat shock protein, iron and oxidative stress regulated proteins, cell adhesion/cellular matrix and vesicle trafficking genes. *J Neural Transm*; 111: 1543-1573.
- Grunblatt E, Mandel S, Maor G et Youdim MB. (2001) Gene expression analysis in N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine mice model of Parkinson's disease using cDNA microarray: effect of R-α-methylphenyl-N-methylpyridinium. *J Neurochem*; 78: 1-12.
- Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov KA, Vagner LL, Ermolaeva OD, Lukyanov SA et Sverdlov ED. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction : cloning of Jurkat cell transcripts induced by phytohemagglutinin and phorbol 12-myristate 13-acetate. *Anal Biochem*; 240(1): 90-97.
- Hashimoto SI, Suzuki T, Nagai S, Yamashita T, Toyoda N et Matsushima K. (2000). Identification of genes specifically expressed in human activated and mature dendritic cells through serial analysis of gene expression, *Blood* 96: 2206-2214.
- Hastie ND et Bishop JO. (1976) The expression of three abundance classes of messenger RNA in mouse tissues. *Cell*; 9: 761-774.
- Hauser MA, Li YJ, Xu H, Nouredine MA, Shao YS, Gullans SR, Scherzer CR, Jensen RV, McLaurin AC, Gibson JR, Scott BL, Jewett RM, Stenger JE, Chmechel DE, Hulette CM et Vance JM. (2005) Expression profiling of substantia nigra in Parkinson disease, progressive supranuclear palsy, and frontotemporal dementia with parkinsonism. *Arch Neurol*; 62: 917-921.
- Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N et Quackenbush J. (2000) A concise guide to cDNA microarray analysis. *Biotechniques*; 3: 548-554.

Hérault F, Le Meuth-Metzinger V, Désert C, Retout E, Piumi F, Govoroun M, Duclos M J et Douaire M. (2003) Cinquièmes Journées de la Recherche Avicole, Tours, 26 et 27 mars.

Higuchi R, Fockler C, Dollinger G et Watson R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*; 11: 1026-1030.

Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chisoe S, Dietrich N, DuBuque T, Favello A, Gish W, Hawkins M, Hultman M, Kucaba T, Lacy M, Le M, Le N, Mardis E, Moore B, Morris M, Parsons J, Prange C, Rifkin L, Rohlfing T, Schellenberg K, Lennon G Bento Soares M, Tan F, Thierry-Meg J, Trevaskis E, Underwood K, Wohldman P, Waterston R, Wilson R et Marra M. (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res*; 6: 807-828.

Hilsenbeck SG, Friedrichs WE, Schiff R, O'Connell P, Hansen RK, Osborne CK et Fuqua SA. (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J Natl Cancer Inst*; 91: 453-459.

Hinfray J. (1997) Les puces à ADN. *Biofutur*; 166: 1-15.

Hoch RV, Thompson DA, Baker RJ et Weigel RJ. (1999) GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int J Cancer*; 84: 122-128.

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D, Hollingsworth ZR, Collin F, Synek B, Holmans PA, Young AB, Wexler NS, Delorenzi M, Kooperberg C, Augood SJ, Faull RL, Olson JM, Jones L et Luthi-Carter R. (2006) Regional and cellular gene expression changes in human Huntington's disease brain; *Hum Mol Genet* 15: 965-977.

Hoen PAC, de Kort F, van Ommen GJB et den Dunnen JT. (2003) Fluorescent labelling of cRNA for microarray applications. *Nucleic Acids Res*; 31(5): e20.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Keyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraborty K, Simon J, Bard M et Friend SH. (2000) Functional discovery via a compendium of expression profiles. *Cell*; 95: 717-728.

Iacobuzio-Donahue CA, Maitra A, Olsen M, Lowe AW, van Heek NT, Rosty C, Walter K, Sato N, Parker A, Ashfaq R, Jaffee E, Ryu B, Jones J, Eshleman JR, Yeo CJ, Cameron JL, Kem SE, Hruban RH, Brown PO et Goggins M. (2003) Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. *Am J Pathol*; 162: 1151-1162.

Iacobuzio-Donahue CA, Maitra A, Shen-Ong GL, van Heek T, Ashfaq R, Meyer R, Walter K, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kem SE, Goggins M et Hruban RH. (2002) Discovery of novel tumor markers of pancreatic cancer using global gene expression technology. *Am J Pathol*; 160: 1239-1249.

Jeong KS, Ahn J et Khodursky AB. (2004) Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol*; 5(11): R86.

Jongeneel CV, Iseli C, Stevenson BJ, Riggins GJ, Lal A, Mackay A, Harris RA, O'Hare MJ, Neville AM, Simpson AJ et Strausberg RL. (2003) Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. *Proc Natl Acad Sci USA*, 100(8) :4702-5.

Jordan B. (2000) Jusqu'ou iront les puces. *Med Sci (Paris)*; 16: 950-953.

- Kafatos FC, Jones CW et Efstratiadis A. (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res*; 7(6): 1541-1552.
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H et Gingeras TR. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*; 316: 1484-1488.
- Kingombe CI, Huys G, Tonolla M, Albert MJ, Swings J, Peduzzi R et Jemmi T. (1999) PCR detection, characterization, and distribution of virulence genes in *Aeromonas* spp. *Applied And Environmental Microbiology*; 12: 5293-5302.
- Kloosterman B, Vorst O, Hall RD, Visser RG et Bachem CW. (2005) Tuber on a chip: differential gene expression during potato tuber development. *Plant Biotechnol. J*; 3(5): 505-519.
- Kluger Y, Yu H, Qian J et Gerstein M. (2003) Relationship between gene co-expression and probe localization on microarray slides. *BMC Genomics*; 4(1): 49.
- Kooperberg C, Augood SJ, Faull RL, Olson JM, Jones L et Luthi-Carter R. (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet*; 15: 965-977.
- Lagoda P et Regad F. (2000) Les puces à ADN outils pour une nouvelle vision de la diversité et des ressources génétiques. *Cahiers Agriculture*; 9: 329-340.
- Lamontagne J et Papadopoulou B. (1999) Developmental regulation of spliced leader RNA gene in *Leishmania donovani* amastigotes is mediated by specific polyadenylation. *J Biol Chem*; 274: 6602-6609.
- Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO et Davis RW. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A*; 94(24): 13057-13062.
- Leclercq B, Blum JC et Boyer JP. (1980). *Poult. Sci* ; 21 : 107-113.
- Lennon G et Lehrach H. (1991) hybridization analyses of arrayed cDNA libraries. *Trends in Genetics*; 7(10): 314-317.
- Leung YF et Cavalieri D. (2003) Fundamentals of cDNA microarray data analysis. *Trends in Genetics*; 19 (11): 649-659.
- Liang P et Pardee AB. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*; 257: 967-971.
- Liew CC et Dzau VJ. (2004) Molecular genetics and genomics of heart failure. *Nat.Rev.Genet*; 5: 811-825.
- Lipshutz RJ, Fodor S, Gingeras T et Lockhart D. (1999) High-density synthetic oligonucleotide arrays. *Nat. Genetics*; 21: 20-24.
- Lisitsyn N, Lisitsyn N et Wigler M. (1993) Cloning the differences between two complex genomes. *Science*; 259(5097): 946-951.

- Liu TX, Zhang JW, Tao J, Zhang RB, Zhang QH, Zhao CJ, Tong JH, Lanotte M, Waxman S et Chen SJ. (2000) Gene expression networks underlying retinoic acid-induced differentiation of acute promyelocytic leukemia cells. *Blood*; 96: 1496-1504.
- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H et Brown EL. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat.Biotechnol*; 14: 1675-1680.
- Lockhart DJ et Winzeler EA. (2000) Genomics, gene expression and DNA arrays. *Nature*; 405: 827-836.
- Luthi-Carter R, Strand AD, Hanson SA, Kooperberg C, Schilling G, La Spada AR, Merry D E, Young AB, Ross CA, Borchelt DR et Olson J. M. (2002b) Polyglutamine and transcription : gene expression changes shared by DRPLA and Huntington's disease mouse models reveal context-independent effects. *Hum Mol Genet*; 11: 1927-1937.
- Malorny B et Hoorfar J. (2005) Toward standardization of diagnostic PCR testing of fecal samples: lessons from the detection of salmonellae in pigs. *Journal Of Clinical Microbiology*; 7: 3033-3037.
- Mantripragada KK, Buckley PG, de Stahl TD et Dumanski JP. (2004) Genomic microarrays in the spotlight. *Trends Genet*; 20: 87-94.
- Marc Ouellette, Martin Olivier, Sachiko Sato et Barbara Papadopoulou. (2003) Le parasite *Leishmania* à l'ère de la post-génomique. *M/S : médecine sciences*; 19 (10): 900-909.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu , Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF et Rothberg JM. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*; 437: 376-380.
- Matsumura H, Ito A, Saitoh H, Winter P, Kahl G, Reuter M, Kruger DH et Terauchi R. (2005) SuperSAGE. *Cell Microbiol*; 7(1): 11-18.
- Mirnic ZK, Mirnic K, Terrano D, Lewis DA, Sisodia SS et Schor NF. (2003) DNA microarray profiling of developing PS1-deficient mouse brain reveals complex and coregulated expression changes. *Mol Psychiatry*; 8: 863-878.
- Miyamoto T, Iwasaki H, Reizis B, Ye M, Graf T, Weissman IL et Akashi K. (2002) Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev Cell*; 3(1): 137-147.
- Nacht M, Ferguson AT, Zhang W, Petroziello JM, Cook BP, Gao YH, Maguire S, Riley D, Coppola G, Landes GM, Madden SL et Sukumar S. (1999) Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res*; 59: 5464-5470.
- O'brien HE, Parrent JL, Jackson JA, Moncalvo JM et Vilgalys R. (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied And Environmental Microbiology*; 9: 5544-5550.

Park IK, He Y, Lin F, Laerum OD, Tian Q, Bumgarner R, Klug CA, Li K, Kuhr C, Doyle MJ, Xie T, Schummer M, Sun Y, Goldsmith A, Clarke MF, Weissman IL, Hood L et Li L. (2002) Differential gene expression profiling of adult murine hematopoietic stem cells. *Blood*; 99(2): 488-498.

Paton AW et Paton JC. (2005) Multiplex PCR for direct detection of Shiga toxigenic *Escherichia coli* strains producing the novel subtilase cytotoxin. *Journal Of Clinical Microbiology*; 6: 2944-2947.

Perou C, Sorlie T, Eisen M, Van De Rijn M, Jeffrey S, Rees C, Pollack J, Ross D, Johnsen H, Akslen L, Fluge O, Pergamenschikov A, Williams C, Zhu S, Lonning P, Borresen-Dale A, Brown P et Botstein D. (2000) Molecular portraits of human breast tumours. *Nature*; 406 (6797): 747-752.

Phimister B. (1999) Going global. *Nature Genetics*; 21: 1.

Peters DG, Kassam AB, Yonas H, O'Hare EH, Ferrell RE et Brufsky AM. (1999) Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite. *Nucleic Acids Res*; 27(24): e39.

Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Sampson R, Houlgatte R, Soularue P et Auffray C. (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res*; 6: 492-503.

Poretzky RS, Bano N, Buchan A, Lecleir G, Kleikemper J, Pickering M, Pate WM, Moran MA et Hollibaugh JT. (2005) Analysis of microbial gene transcripts in environmental samples. *Applied And Environmental Microbiology*; 7: 4121-4126.

Ramsay G. (1998) DNA chips: state-of-the-art. *Nature Biotechnology*; 16 (1): 40-44.

Rao M. (2004) Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol*; 275(2): 269-286.

Rathod PK, Ganesan K, Hayward RE, Bozdech Z et DeRisi JL. (2002) DNA microarrays for malaria. *Trends Parasitol*; 18: 39-45.

Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, Kramer M et Woychik R. (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Brief Funct Genomic Proteomic*; 1(1): 95-104.

Richards M, Tan SP, Tan JH, Chan WK et Bongso A. (2004) The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells*; 22(1) :51-64.

Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Teyers M, Boone C et Friend SH. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science*; 287: 873-880.

Rockett J et Dix D. (2000) DNA arrays: technology, options and toxicological applications. *Xenobiotica*; 30 (2): 155-177.

Rompré A, Servais P, Baudart J, De-Roubin MR et Laurent P. (2002) Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. *J Microbiol Methods*; 1: 31-54.

Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, Van Der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML,

- Baker BJ et Buell CR. (2003) Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol*; 131(2): 419-429.
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D et Brown PO. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet*; 24: 227-235.
- Roth M, Feng L, McConnell K, Schaffer P, Guerra C, Affourtit J, Piper K, Guccione L, Hariharan J, Ford M, Powell S, Krishnaswamay H, Lane J, Intrieri G, Merkel J, Perbost C, Valerio A, Zolla B, Graham C, Hnath J, Guccione L, Michaelson C, Wang R, Ying B, Halling C, Parman CE, Raha D, Orr B, Jedrzkiewicz B, Liao J, Tevelev A, Mattessich MJ, Kranz DM, Lacey M, Kaufman JC, Kim J, Latimer DR et Lizardi PM. (2004) Expression profiling using a hexamer-based universal microarray. *Nature Biotechnology*; 22(4): 418-426.
- Rubin GM, Hong L, Brokstein P, Evans-Holm M, Frise E, Stapleton M et Harvey DA. (2000) A *Drosophila* complementary DNA resource. *Science*; 287: 2222-2224.
- Ryu B, Jones J, Blades NJ, Parmigiani G, Hollingsworth MA, Hruban RH, Kern SE. (2002) Relationships and differentially expressed genes among pancreatic cancers examined by large-scale serial analysis of gene expression. *Cancer Res*; 62: 819-826.
- Ryu B, Jones J, Hollingsworth MA, Hruban RH et Kern SE. (2001) Invasionspecific genes in malignancy : serial analysis of gene expression comparisons of primary and passaged cancers. *Cancer Res*; 61: 1833-1838.
- Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW et Velculescu VE. (2002) Using the transcriptome to annotate the genome. *Nat. Biotechnol*; 20(5):508-512.
- Sato N, Sanjuan IM, Heke M, Uchida M, Naef F et Brivanlou AH. (2003) Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*; 260(2): 404-413.
- Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M et Wisman E. (2001) Microarray Analysis of Diurnal and Circadian-Regulated Genes in *Arabidopsis*. *Plant Cell*; 13: 113-123.
- Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E et Davis RW. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol*; 16: 301-306.
- Schena M, Shalon D, Davis RW et Brown PO. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*; 270: 467-470.
- Schena M, Shalon D, Davis RW et Brown PO. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*; 270: 368-371.
- Schoolnik GK. (2002) Functional and comparative genomics of pathogenic bacteria. *Curr Opin Microbiol*; 5: 20-26.
- Seki M, Carninci P, Nishiyama Y, Hayashizaki Y et Shinozaki K. (1998) High-efficiency cloning of *Arabidopsis* full-length cDNA by biotinylated CAP trapper. *Plant J*; 15: 707-720.
- Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K,

- Shinagawa A et Shinozaki K. (2002c) Functional annotation of a full-length Arabidopsis cDNA collection. *Science*; 296: 141-145.
- Seki M, Satou M, Sakurai T et Shinozaki K. (2002d) RIKEN Arabidopsis full-length cDNA database. *Trends Plant Sci*; 7: 562-563.
- Sgori DC, Teng S, Robinson G, LeVangie R, Hudson JR et Elkahloun AG. (1999) *In vivo* gene expression profile analysis of human breast cancer progression. *Cancer Res*; 59: 5656-5661.
- Sharov AA, Piao Y, Matoba R, Dudekula DB, Qian Y, VanBuren V, Falco G, Martin PR, Stagg CA, Bassey UC, Wang Y, Carter MG, Hamatani T, Aiba K, Akutsu H, Sharova L, Tanaka TS, Kimber WL, Yoshikawa T, Jaradat SA, Pantano S, Nagaraja R, Boheler KR, Taub D, Hodes RJ, Longo DL, Schlessinger D, Keller J, Klotz E, Kelsoe G, Umezawa A, Vescovi AL, Rossant J, Kunath T, Hogan BL, Curci A, D'Urso M, Kelso J, Hide W et Ko MS. (2003) Transcriptome analysis of mouse stem cells and early embryos. *PLoS Biol*; 1(3): E74.
- Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P et Hayashizaki Y. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*; 100 (26): 15776-15781.
- Sidiropoulos M, Chang A, Jung K et Diamandis EP. (2001) Expression and regulation of prostate androgen regulated transcript-1 (PART-1) and identification of differential expression in prostatic cancer. *Br J Cancer*; 85: 393-397.
- Smyth GK et Speed T. (2003) Normalization of cDNA microarray data. *Methods*; 31(4): 265-273.
- Son CG, Bilke S, Davis S, Greer BT, Wei JS, Whiteford CC, Chen QR, Cenacchi N et Khan J. (2005) Database of mRNA gene expression profiles of multiple human organs. *Genome Res*; 15: 443-450.
- Sorlie T, Perou CM, Tibshirani R et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc.Natl.Acad.Sci.U.S.A.*; 98: 10869- 10874.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown, PO, Botstein D et Futcher B, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D et Futcher B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol.Biol.Cell*; 9: 3273-3297.
- Stapleton M, Carlson J, Brokstein P, Yu C, Champe M, George R, Guarin H, Kronmiller B, Pacleb J, Park S, Wan K, Rubin GM et Celniker SE. (2002a) A *Drosophila* full-length cDNA resource. *Genome Biol*; 3: research0080.0081-0080.0088.
- Stapleton M, Liao G, Brokstein P, Hong L, Carninci P, Shiraki T, Hayashizaki Y, Champe M, Pacleb J, Wan K, Yu C, Carlson J, George R, Celniker S et Rubin GM. (2002b) The *Drosophila* gene collection : identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res*; 12: 1294-1300.
- Steenman M, Chen YW, Le Cunff M, Lamirault G, Varro A, Hoffman E et Leger JJ. (2003) Transcriptomal analysis of failing and nonfailing human hearts. *Physiol Genomics*; 12: 97-112.
- Steenman M, Lamirault G, Le Meur N, Le Cunff M, Escande D et Leger JJ. (2005) Distinct molecular portraits of human failing hearts identified by dedicated cDNA microarrays. *Eur.J.Heart Fail*; 7: 157-165.

- Stollberg J, Urschitz J, Urban Z et Boyd CD. (2000) A quantitative evaluation of SAGE. *Genome Res*; 10(8): 1241-1248.
- Straub TM et Chandler DP. (2003) Towards a unified system for detecting waterborne pathogens. *J Microbiol Methods*; 2: 185-197.
- Strausberg RL, Dahl CA et Klausner RD. (1997) New opportunities for uncovering the molecular basis of cancer. *Nature Genet*; 16: 415-416.
- Talaat AM, Howard ST, Hale W 4th, Lyons R, Garner H et Johnston SA. (2002) Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res*; 30: E104.
- Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES et Golub TR. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA*; 96: 2907-2912.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ et Church GM. (1999) Systematic determination of genetic network architecture. *Nature Genet*; 22: 281-285.
- Terskikh AV, Easterday MC, Li L, Hood L, Kornblum HI, Geschwind DH et Weissman IL. (2001) From hematopoiesis to neurogenesis: evidence of overlapping genetic programs. *Proc. Natl. Acad. Sci. USA*; 98(14): 7934-7939.
- Thieffry D et Thomas R. (1998) Qualitative analysis of gene networks. *Pac Symp Biocomput*; 77-88.
- Tkatchenko AV, Le Cam G, Leger JJ et Dechesne CA. (2000) Large-scale analysis of differential gene expression in the hindlimb muscles and diaphragm of mdx mouse. *Biochim Biophys Acta*; 1500: 17-30.
- Trindade LM, Horvath BM, Van Berloo R et Visser RGF. (2004) Analysis of genes differentially expressed during potato tuber life cycle and isolation of their promoter regions. *Plant Sci*; 166(2): 423-433.
- Tse C et Capeau J. (2003) Quantification des acides nucléiques par PCR quantitative en temps réel. *Annale de Biologie Clinique*; 61: 279-293.
- Unneberg P, Wennborg A et Larsson M. (2003) Transcript identification by analysis of short sequence tags influence of tag length, restriction site and transcript database. *Nucleic Acids Res*; 31(8): 2217-2226.
- Van Dam R et Quake S. (2002) Gene expression analysis with universal n-mer arrays. *Genome Research*; 12 (1): 145-152.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, Vandervelde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH et Bernards S. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N.Engl.J.Med*; 347: 1999- 2009.
- Velculescu VE, Vogelstein B et Kinzler KW. (2000) Analysing uncharted transcriptomes with SAGE. *Trends Genet*; 16: 423-425.
- Velculescu VE, Zhang L, Vogelstein B et Kinzler KW. (1995) Serial analysis of gene expression. *Science*; 270: 484- 487.

- Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett Jr DE, Heiter P, Vogelstein B et Kinzler KW. (1997) Characterization of the yeast transcriptome. *Cell*; 88: 243-251
- Vidal M. (2001) A biological atlas of functional maps. *Cell*; 104: 333-339.
- Vingron M. (2001) Bioinformatics needs to adopt statistical thinking. *Bioinformatics*; 17(5): 389-390.
- Virlon B, Cheval L, Buhler JM, Billon E, Doucet A et Elalouf JM. (1999) Serial microanalysis of renal transcriptomes. *Proc. Natl. Acad. Sci. USA*; 96(26): 15286-15291.
- Walker J et Rigley K. (2000) Gene expression profiling in human peripheral blood mononuclear cells using high-density filter-based cDNA microarrays. *J Immunol Methods*; 239: 167-179.
- Wang RC, Guegler K, LaBrie ST et Crawford NM. (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell*; 12: 1491-1509.
- Ward DM, Weller R et Bateson MM. (1990) 16S rRNA sequences reveal uncultured inhabitants of a well-studied thermal community. *Fems Microbiology Reviews*; 2-3: 105-115.
- Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N. (1992) A survey of expressed genes in *Caenorhabditis elegans*. *Nat Genet*; 1: 114-123.
- Wei CL, Miura T, Robson P, Lim SK, Xu XQ, Lee MY, Gupta S, Stanton L, Luo Y, Schmitt J, Thies S, Wang W, Khrebtukova I, Zhou D, Liu ET, Ruan YJ, Rao M et Lim B. (2005) Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells*; 23(2): 166-185.
- Welsh J et McClelland M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res*; 18(24): 7213-7218.
- Wen X, Fuhrman S, Michaels GS, Carr DB, Smith S, Barker JL et Somogyi R. (1998) Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci USA*; 95: 334-339.
- Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T et Sasaki T. (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell*; 14: 525-535.
- Wu Y, El Fakhry Y, Sereno D, Tamar S et Papadopoulou B. (2000) A new developmentally regulated gene family in *Leishmania* amastigotes encoding a homolog of amastin surface proteins. *Mol Biochem Parasitol*; 110: 345-357.
- Wu Z, Blomquist G, Westermarck SO et Wang XR. (2002) Application of PCR and probe hybridization techniques in detection of airborne fungal spores in environmental samples. *J Environ Monit*; 5: 673-678.
- Wurbel MA, Philippe JM, Nguyen C, Victorero G, Freeman T, Wooding P, Miazek A, Mattei MG, Malissen M, Jordan BR, Malissen B, Carrier A et Naquet P. (2000) The chemokine TECK is expressed by thymic and intestinal epithelial cells and attracts double- and single-positive thymocytes expressing the TECK receptor CCR9. *Eur J Immunol*; 30: 262-271.

- Yang GP, Ross DT, Kuang WW, Brown PO et Weigel RJ. (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res*; 27: 1517-1523.
- Yao PJ, Zhu M, Pyun EI, Brooks AI, Therianos S, Meyers VE, et Coleman PD. (2003). Defects in expression of genes related to synaptic vesicle trafficking in frontal cortex of Alzheimer's disease. *Neurobiol Dis*; 12: 97-109.
- Ye SQ, Zhang LQ, Zheng F, Virgil D et Kwitrovich PO. (2000) miniSAGE : gene expression profiling using serial analysis of gene expression from 1 microg total RNA. *Anal Biochem*; 287(1): 144-152.
- Yong-lee Ong et Alexandra Irvine. (2002) Quantitative real-time PCR : a critique of method and practical considerations. *Hematology*; 7(1): 59-67.
- Zeng X, Miura T, Luo Y, Bhattacharya B, Condie B, Chen J, Ginis I, Lyons I, Mejido J, Puri RK, Rao MS et Freed WJ. (2004) Properties of pluripotent human embryonic stem cells BG01 and BG02. *Stem Cells*; 22(3): 292-312.
- Zhang W, Morris QD, Chang R, Shai O, Bakowski MA, Mitsakakis N, Mohammad N, Robinson MD, Zirngibl R, Somogyi E, Laurin N, Eftekharpour E, Sat E, Grigull J, Pan Q, Peng WT, Krogan N, Greenblatt J, Fehlings M, van der Kooy D, Aubin J, Bruneau BG, Rossant J, Blencowe BJ, Frey BJ et Hughes TR. (2004) The functional landscape of mouse gene expression. *J Biol*; 3: 21.

Titre : La transcriptomique : une approche systématique pour l'étude de l'expression des gènes

Résumé

Depuis une vingtaine d'années, l'étude systématique de l'expression des gènes est devenue une réalité grâce à certain nombre d'innovations techniques telles que la production d'oligonucléotides de synthèse, les sondes fluorescentes, l'automatisation du séquençage d'ADN et l'amplification de l'ADN par la PCR. En effet, ce genre d'approches permet de compléter les résultats d'une démarche unitaire en biologie. Des progrès très importants dans certaines disciplines telles que la Biologie Moléculaire, la Chimie Combinatoire et la Bioinformatique ont permis le développement de la transcriptomique qui s'appuie sur différentes techniques, on peut citer, Display différentiel, PCR soustractive, PCR en temps réel, puces ADN, SAGE et MPSS. Ces techniques portent plusieurs applications notamment dans le domaine de la santé, la recherche, le domaine vétérinaire et environnemental.

Grâce à l'approche transcriptomique, il est possible d'identifier et analyser simultanément un nombre considérable des gènes, il est néanmoins raisonnable de reconnaître que les données d'expression de l'ensemble des gènes sont une nouvelle étape dans la compréhension du fonctionnement du système dynamique qu'est la cellule.

Summary

Over the past twenty years, the systematic study of gene expression has become a reality due to a number of technical innovations such as the production of synthetic oligonucleotides, the fluorescent probes, automated sequencing of DNA and DNA amplification by the PCR. Indeed, these approaches can supplement the results of a unit in biology. Very significant progress in some disciplines such as molecular biology, combinatorial chemistry and bioinformatics have enabled the development of transcriptomic based on different techniques, for example, differential Display, subtractive PCR, real-time PCR, DNA chips, SAGE and MPSS. These techniques have many applications especially in the field of health, research and veterinary and environmental. With the transcriptomic approach, it is possible to identify and analyze a considerable number of genes, it is reasonable to recognize that the expression of all genes are a further step in understanding the functioning of dynamic system that is the cell.

المخلص

أصبحت دراسة منهجية التعبير الجيني منذ السنوات العشرين الماضية حقيقة واقعة بفضل عدد من الابتكارات التقنية، مثل إنتاج متعديرات النيوكليوتيدات، الفلوروكروومات، أوتوماتيكية سلسلة الحمض النووي منقوص الأوكسجين و تضخيمه بواسطة تقنية PCR. والواقع أن دراسة التعبير الجيني سمحت بتكملة نتائج أسلوب أحادي في البيولوجيا. تقدم كبير للغاية في بعض التخصصات مثل البيولوجيا الجزيئية، الكيمياء التوافقية والإعلام الآلي الحيوي، أدى إلى تطوير la transcriptomique التي تعتمد على تقنيات مختلفة، على سبيل المثال، Display différentiel، PCR soustractive، PCR en temps réel، puces ADN، SAGE و MPSS. هذه التقنيات لديها تطبيقات عديدة، خاصة في مجال الصحة، البحث، البيطرية والبيئية.

بفضل la transcriptomique أصبح من الممكن تحديد و تحليل عدد كبير من الجينات في آن واحد، و بذلك فبته من المعقول أن نعرف بأن معطيات التعبير الجيني هي مرحلة جديدة لفهم الوظائف الديناميكية للخلية.