



Université Batna 2

Faculté des sciences de la nature et de la vie

Département de biologie des organismes

**Cours de  
Génomique et protéomique  
fonctionnelle  
M2 BMC**

2022-2023

## Table des matières

1.Définition :.....	3
2.Les méthodes d'analyse du transcriptome :.....	3
2.1.RT-PCR ( <i>Retro- Transcriptase-Polymerase Chain Reaction</i> ) :.....	3
2.2. L'analyse en série de l'expression des gènes (SAGE ou <i>Serial Analysis of Gene Expression</i> ) :.....	4
2.3. Les puces à ADN : .....	6
2.5.Séquençage de l'ARN (RNA-seq) : .....	8
3.Intérêt des études de transcriptome :.....	9

### 1. Définition :

La connaissance du génome n'est qu'une étape dans la compréhension de la biologie d'un organisme : son phénotype est essentiellement déterminé par l'expression des gènes qu'il renferme. Il est alors indispensable de s'intéresser à cette expression, et l'analyse du **transcriptome** en est un aspect.

Le terme "**transcriptome**" désigne l'ensemble des transcrits (c'est à-dire des ARN messagers [ARNm] issus de la transcription des gènes) présents dans un type cellulaire à un moment du développement et/ou dans une condition physiologique donnée.

Ce terme a été proposé pour la première fois en 1997 par Velculescu et al. qui le définissent ainsi : "Contrairement au génome, qui est essentiellement statique, le transcriptome peut être modulé par des facteurs internes et externes [et] sert ainsi de lien dynamique entre le génome d'un organisme et ses caractéristiques physiques".

La **transcriptomique** est l'étude de l'ensemble des ARN messagers produits lors du processus de transcription d'un génome. Elle repose sur la quantification systématique de ces ARNm, ce qui permet d'avoir une indication relative du taux de transcription de différents gènes dans des conditions données, donc l'analyse du transcriptome offre la possibilité **d'annoter** les gènes par la mise en évidence de leurs **profils d'expression**. Plusieurs techniques permettent d'avoir accès à cette information, en particulier celle des puces à ADN, celle de la PCR quantitative ou encore celle du séquençage systématique d'ADN complémentaires.

### 2. Les méthodes d'analyse du transcriptome :

Plusieurs méthodes d'analyse du transcriptome sont couramment utilisées.

#### 2.1. RT-PCR (*Retro- Transcriptase-Polymerase Chain Reaction*) :

L'acronyme RT-PCR signifie Reverse Transcriptase PCR, soit une PCR après transcription inverse d'un acide ribonucléique (ARN) en ADN complémentaire (ADNc). La RT-PCR se déroule en deux phases. Dans un premier temps, l'ARN messager est copié en ADN complémentaire (ADNc). Tout d'abord, l'ARNm à étudier est repéré en utilisant une sonde oligonucléotidique spécifique (amorce 1 qui s'hybride à l'extrémité 3' du seul ARNm auquel on s'intéresse), puis la transcriptase inverse (ou rétrotranscriptase) permet la synthèse du brin complémentaire (sous une forme d'ADNc simple brin), une 2ème amorce oligonucléotidique spécifique (amorce 2) permettra la synthèse du 2ème brin par extension. Dans un second temps, l'ADN double brin synthétisé servira de matrice pour une réaction PCR classique (Figure 1).

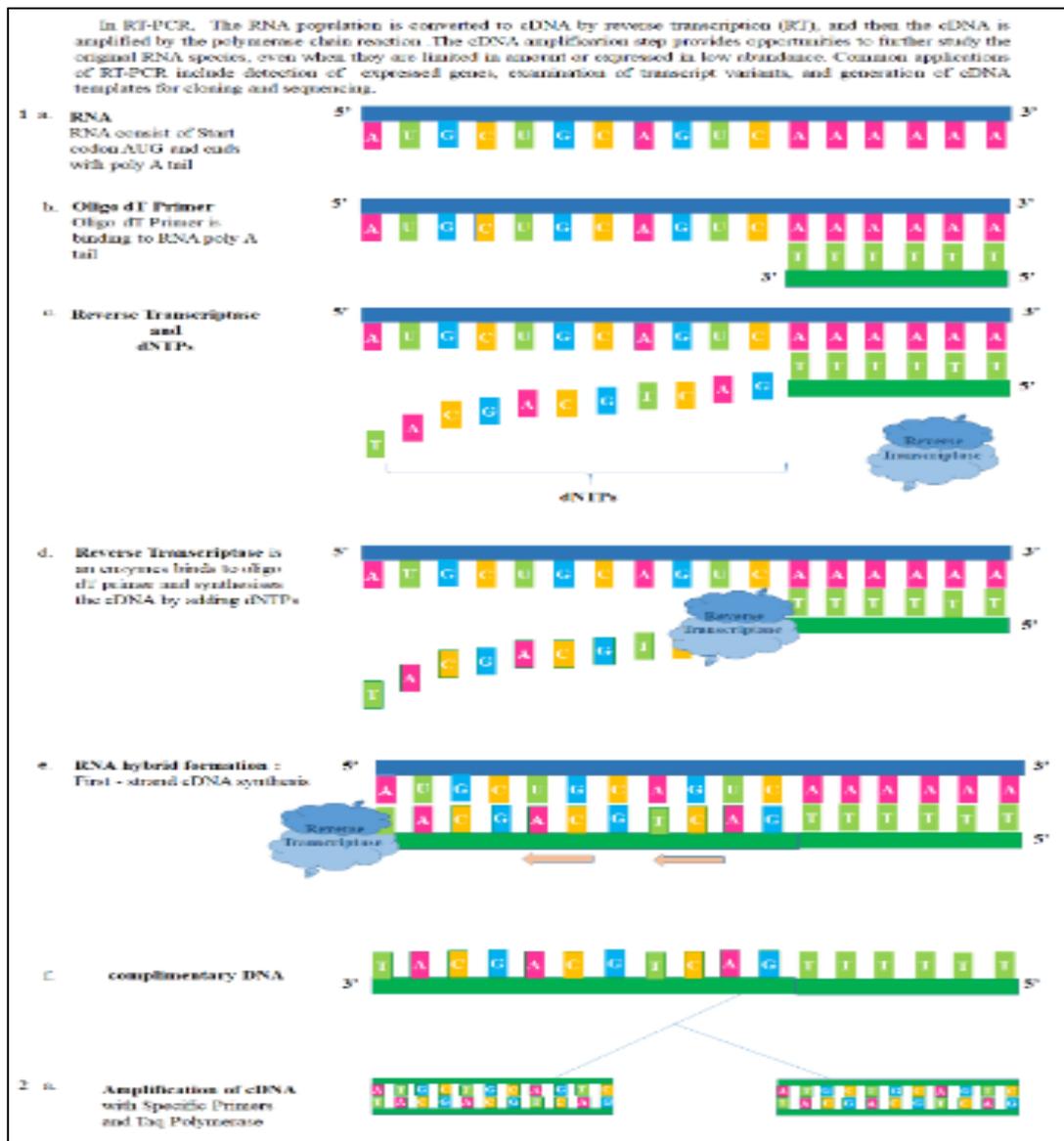


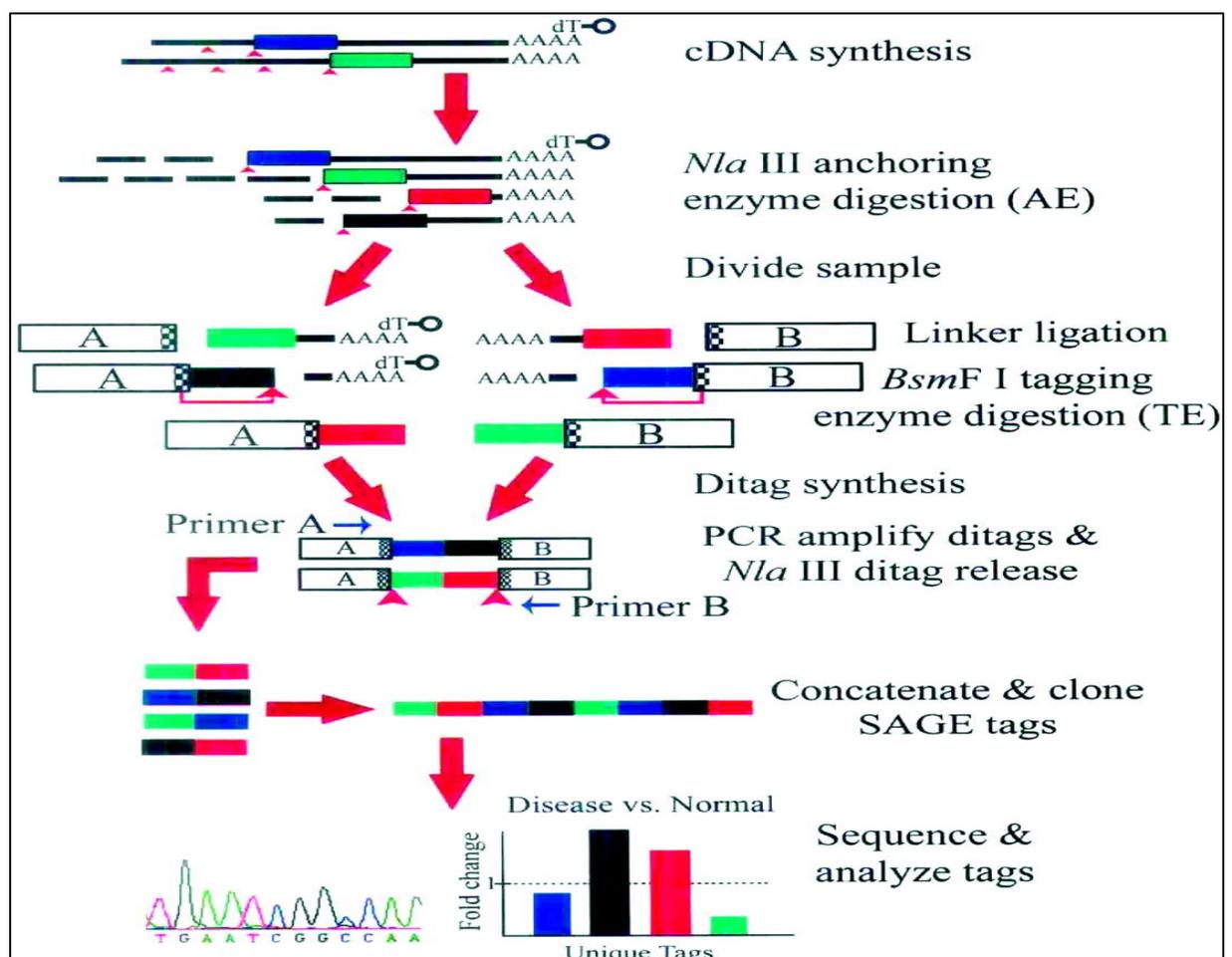
Figure1 : Les étapes de ta technique RT\_PCR

## 2.2. L'analyse en série de l'expression des gènes (SAGE ou *Serial Analysis of Gene Expression*) :

Le principe de base se fonde sur l'analyse **séquentielle** d'un **grand nombre de courts fragments d'ADNc**, dont chacun représente **la signature d'un gène**. La méthode SAGE est basée sur trois principes : Le premier consiste à isoler des séquences spécifiques (**étiquettes ou Tag**) de chaque ARNm, le second repose sur le fait que la condensation de l'information par la ligation de nombreuses **étiquettes** en une seule molécule d'ADN (**concatémère**), qui peut être facilement manipulé, permet une plus grande efficacité dans l'analyse des données. Le dernier admet que **la proportion relative** des différentes étiquettes dans la librairie est **le reflet de l'abondance des transcrits correspondant dans l'échantillon biologique**. En générant un grand nombre de séquences à partir d'un échantillon donné, il est possible de mesurer la

fréquence des différentes étiquettes et avoir une idée du **profil d'expression** d'un grand nombre de gènes.

Elle consiste à réaliser un inventaire des transcrits (ARNm présents dans un échantillon de cellules, tissus ou organes et convertis en ADNc) au moyen de courts fragments d'ADNc de 9 à 14 pb, nommés **étiquettes** ("tag"). Toutes les étiquettes d'un transcrit donné (jusqu'à une cinquantaine) sont ensuite assemblées sous forme de longue séquence appelée **concatémère**. Le séquençage de ces concatémères permet d'établir l'identité de chaque étiquette (**et donc du gène dont elle est issue**). L'analyse informatisée des séquences permet d'extraire les différentes étiquettes, de les classer et de les dénombrer. Enfin, l'interrogation des banques de données permet de les identifier. En générant ainsi un grand nombre de séquences à partir d'un échantillon donné, il est possible de mesurer la fréquence des différentes étiquettes et finalement de réaliser **une mesure quantitative du profil d'expression**. Chaque transcrit est ensuite identifié par homologie dans les bases de données et le niveau **de son expression est lié à la quantité relative de l'étiquette au sein du concaténat**.

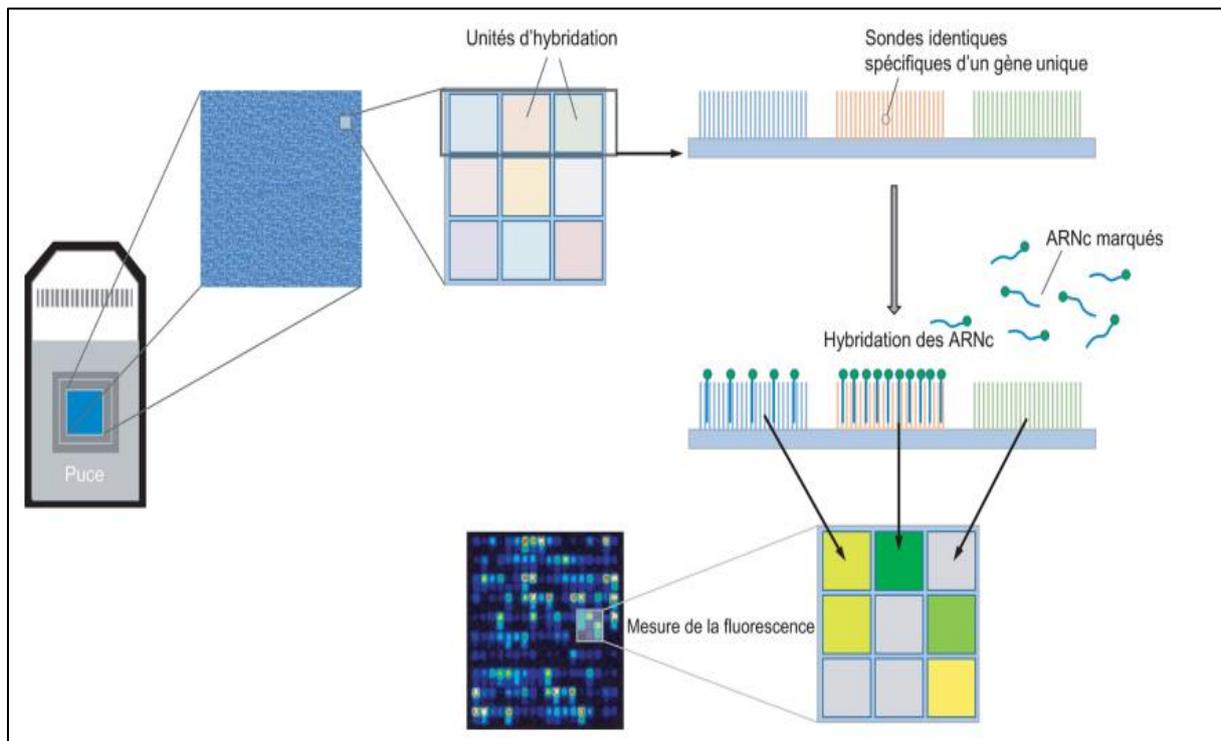


**Figure 2 :** Principe de différentes étapes de la technique SAGE.

### 2.3. Les puces à ADN :

La technique la plus couramment utilisée à l'heure actuelle est celle des puces à ADN ou *micro arrays*. Les puces à ADN, sont connues sous une multitude de noms différents tels que DNA arrays, microréseaux, micro-arrays, DNA chips, et par abus de langage biopuces. Leur développement date des années 1990. Malgré son coût important et la nécessité d'une plate-forme technique et bio-informatique, elle présente l'avantage de visualiser le niveau d'expression d'un grand nombre de gènes en une seule expérience et d'être disponible commercialement.

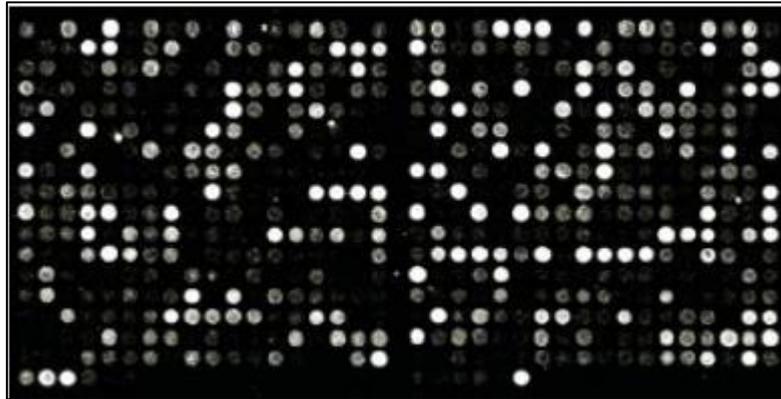
Les puces sont des matrices de verre sur lesquelles sont fixées de courtes séquences nucléotidiques appelées **sondes** (probes), organisées en unités d'hybridation. Chaque unité contient plusieurs millions d'exemplaires de la même sonde. Les ARN extraits d'une population cellulaire sont marqués à la biotine puis déposés sur la puce. Ils se fixent dans les unités d'hybridation contenant des sondes qui leur sont spécifiques. Les ARN sont ensuite révélés à l'aide d'un fluorochrome, et l'intensité de fluorescence émise par chaque unité est mesurée par un module de criblage. L'intensité de fluorescence émise dépend du nombre de molécules d'ARN spécifiquement appariées (Figure3).



**Figure 3 :** Puces transcriptomiques. Les ARN extraits d'une population cellulaire sont transcrits en ADNc par transcription inverse (les ADNc, plus stables, peuvent être conservés plus longtemps) puis les ADNc sont à nouveau transcrits en ARN complémentaires (ARNc) marqués à la biotine. Les ARNc sont déposés sur la puce et s'hybrident uniquement dans les régions contenant des sondes qui leur sont spécifiques. L'ARNc hybridé est coloré à l'aide d'une molécule fluorescente (streptavidine-phycoérythrine) qui se lie à la biotine. Les puces

sont scannées avec un laser. Un module de criblage mesure l'intensité de fluorescence émise par chaque unité d'hybridation. Chaque unité est spécifique d'un gène. Seules les unités contenant de l'ARN émettent de la fluorescence, et l'intensité de fluorescence de chaque unité est directement proportionnelle au nombre de molécules d'ARN présentes dans l'échantillon, c'est-à-dire au niveau d'expression de chaque gène.

La figure 4 est un exemple, chaque tache représente un gène, son intensité est proportionnelle à la quantité du transcrit. Il faut bien souligner que cette technique mesure l'expression d'un gène dans un tissu donné, au cours du développement ou dans des circonstances physiologiques ou pathologiques déterminées.



**Figure 4 :** Micro-alignements (« microarrays »). L'intensité des taches est proportionnelle à la quantité d'ARN messager (Livre Aide-mémoire Biologie et génétique moléculaires).

#### 2.4. La PCR en temps réel :

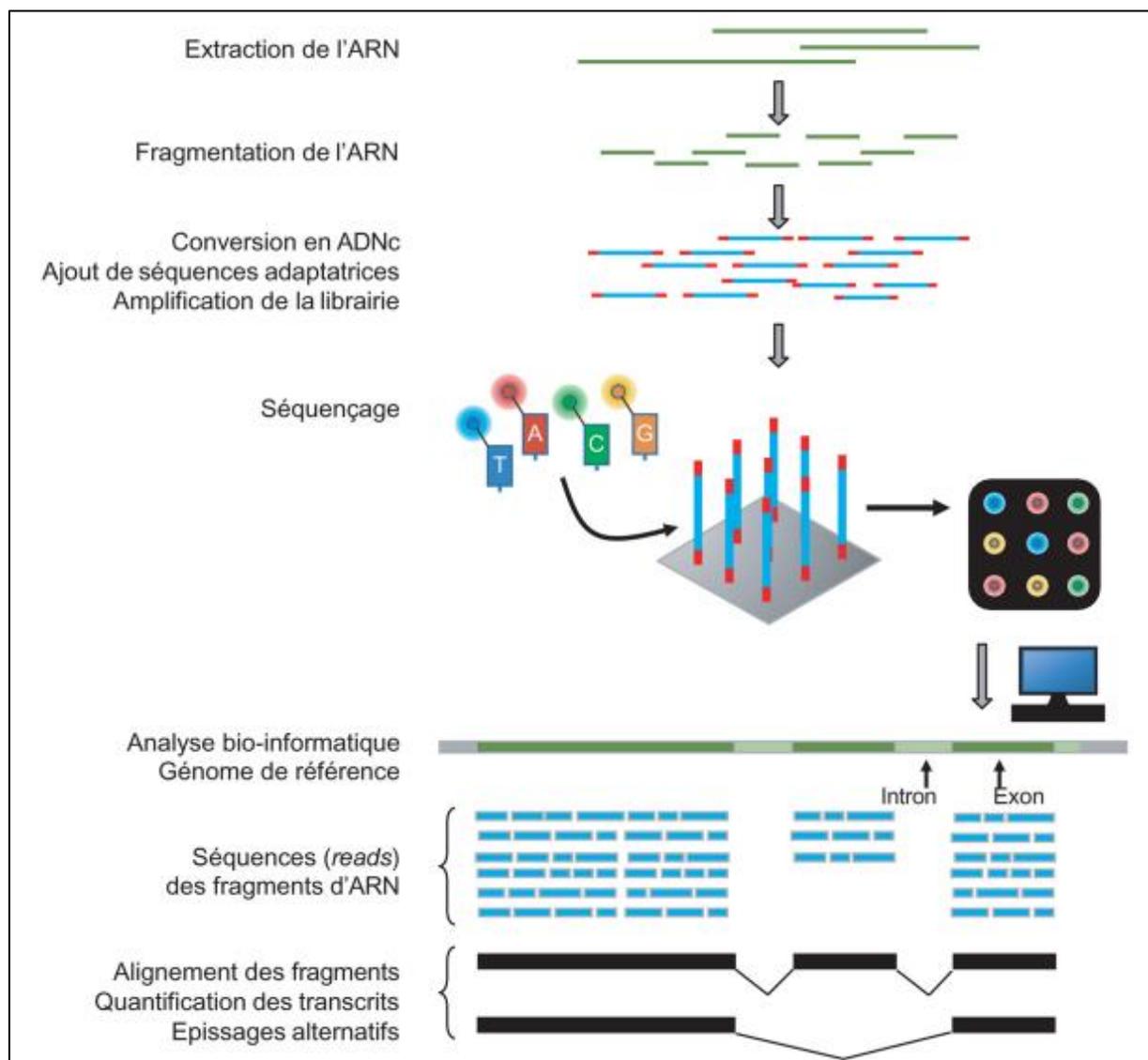
La quantification de l'expression des gènes par "RT-PCR" en temps réel est une alternative aux techniques de microarray et de SAGE qui présente des avantages en termes de sensibilité (moindre quantité d'ARN nécessaire, quantification de transcrits peu exprimés), précision (quantification du nombre de copies dès la phase exponentielle de l'amplification) et reproductibilité (meilleure standardisation des différentes étapes de la technique).

Elle combine l'amplification avec la détection et la quantification d'un signal fluorescent. La réaction d'amplification de la séquence ADN cible suit les mêmes étapes de la PCR classique dénaturation, hybridation, extension. Cette technique repose, non plus sur une détection en **point final** des produits de PCR formés, mais sur une analyse de la **cinétique** de la réaction PCR au moyen d'un système capable de détecter « en tube fermé » les produits de PCR formés après chaque cycle d'amplification. **Elle est basée sur une réaction enzymologique, la PCR et sur la mesure en continue de son produit.** À chaque cycle d'amplification, la quantité d'ADN total ou d'amplicon est mesurée grâce à un marqueur fluorescent.

## 2.5. Séquençage de l'ARN (RNA-seq) :

L'application des méthodes de séquençage à l'étude des ARN apporte une dimension nouvelle à l'étude du transcriptome. Alors que l'utilisation des puces repose sur la conception de sondes nucléotidiques, le RNA-seq permet **de déterminer la séquence des transcrits** sans connaissance préalable du gène d'intérêt. Le RNA-seq repose sur trois étapes (Figure 5) :

- l'extraction et le fractionnement des ARN, suivis de la constitution d'une librairie d'ADNc ;
- séquençage proprement dit ;
- analyse bio-informatique (alignement, assemblage et quantification) des millions de séquences générées.



**Figure 5 : Étapes du RNA-seq** 1. Préparation de la librairie d'ADNc : après purification, les ARN sont fragmentés en courtes séquences (les procédés de séquençage ne peuvent prendre en charge que de courts fragments, de l'ordre de 200 à 300 bases alors que les transcrits d'ARN peuvent contenir plusieurs milliers de bases). Les fragments d'ARN sont convertis en ADNc double brin (plus stable et plus faciles à amplifier et manipuler que l'ARN). Les fragments sont ligaturés de part et d'autre par des adaptateurs de séquençage qui

permettent à l'automate de reconnaître les fragments et de séquencer plusieurs échantillons en même temps. Les ADNc sont ensuite amplifiés. 2. Séquençage : la plateforme Illumina™ utilise des nucléotides marqués avec des fluorochromes différents. Les fluorochromes incorporés au brin synthétisé sont visualisés un par un à l'aide d'un lecteur optique. 3. Analyse bio-informatique des séquences. Les millions de séquences de fragments d'ARN sont réalignés et assemblés. Il est ensuite possible de quantifier les niveaux d'expression des transcrits, de détecter des altérations génomiques, de relever des variations de séquence non détectables à l'échelle génomique, telles que l'expression de transcrits alternatifs.

### 3. Intérêt des études de transcriptome :

Alors que l'information génomique est globalement identique dans toutes les cellules d'un individu, l'activation spatio-temporelle des différents gènes caractérise les différents organes et les cellules qui les constituent. Ainsi, le transcriptome (c'est-à-dire l'ensemble des molécules d'ARN résultant de l'activité des gènes exprimés) constitue une véritable **signature** et un **lien clé entre génome et phénotype**.

L'ARN obtenu dans le noyau est appelé ARN pré-messager ou précurseur. Après maturation avant son transfert vers le cytoplasme (*passage de membrane nucléaire*) ou l'ARNm mature sera traduit en protéines par les ribosomes. Sachant qu'à partir d'un même gène, plusieurs copies d'ARN pré-messager peuvent être produits à des niveaux différents en fonction de l'activité de la cellule. Le transcriptome reflétera donc **le niveau d'expression** de tous les gènes à un temps t pour une condition physiologique donnée. **La meilleure connaissance du niveau d'expression d'un gène dans différentes situations constitue une avancée vers la compréhension de sa fonction.**

D'une manière plus générale, la première annotation fonctionnelle du génome peut se faire par:

- Comparaison du transcriptome de différents types cellulaires, dans différentes conditions,
- L'analyse de l'ensemble du transcriptome d'une cellule à divers stades de son cycle cellulaire ou dans des conditions pathologiques.

Un autre intérêt de la transcriptomique est de pouvoir comparer les niveaux d'expression des transcrits entre des types cellulaires particuliers, entre des conditions différentes ou au cours du temps pendant un phénomène particulier. Que ces données proviennent de séquençage ou de puces à ADN elles conduisent à l'étude de matrice d'expression ou, pour chaque objet étudié (en général des gènes), on associe les intensités d'expression aux différentes conditions.

L'analyse globale du transcriptome par puces à ADN permet de définir un **“profil d'expression”** des gènes qui est une caractéristique intrinsèque de l'échantillon (type cellulaire, tissu) à un moment donné ou dans une condition physiologique déterminée (stade du

développement, pathologie ou réaction à une drogue par exemple). Il est alors possible de comparer les profils de plusieurs échantillons entre eux.

En cancérologie, par exemple, l'analyse simultanée du transcriptome de tumeurs peut aboutir à l'identification d'une liste de gènes **dont l'expression est spécifique d'un type tumoral donné** et/ou d'un stade de la tumorigenèse. On parle alors de **“signature moléculaire”**. **Cet outil est un moyen de classification, de diagnostic, et définit une “carte d'identité” de la tumeur.**

L'analyse comparative du transcriptome permet aussi de déterminer des gènes dont l'expression varie d'un échantillon à l'autre (**expression différentielle**), par exemple au cours du développement ou dans un cas pathologique, ou encore en réponse à une drogue. Ainsi, l'analyse du transcriptome de cellules souches neurales a mis en évidence, d'une part, des gènes responsables de la capacité de renouvellement des cellules souches et, d'autre part, des gènes responsables de la différenciation. Dans le cas de la chorée de Huntington, des études sur des modèles murins de la maladie ont permis de mettre en évidence des gènes dont l'expression varie dans le contexte de maladies à polyglutamines. Ces gènes apportent des informations sur la physiopathologie de ce type de maladies et peuvent s'avérer des cibles thérapeutiques intéressantes.