

L'identification des fonctions cellulaires des produits des gènes

Chaque cellule humaine, à l'exception du sperme et des ovules, possède un ensemble complet de gènes. Cependant, une cellule sanguine diffère dans sa morphologie et sa physiologie d'une cellule hépatique. Comment donc expliquer ces différences si toutes les cellules ont le même matériel génétique ? La réponse est simple. **Tous les gènes ne sont pas transcrits et exprimés dans toutes les cellules. Seules les protéines nécessaires sont présentes dans une cellule à un moment donné de la vie de la cellule.**

Le protéome d'une cellule ou d'un tissu dépend donc du type de cellule et de son état actuel. En principe, l'ordre de base du gène (le génotype) doit donc être altéré par mutation pour une modification de l'expression du gène et les changements de phénotype qui en résultent. Mais il a été montré au cours des dernières décennies que des facteurs environnementaux peuvent influencer le phénotype en altérant l'expression des gènes sans adapter la séquence nucléotidique des gènes. Cette adaptation de l'expression des gènes est appelée **épigénétique** et joue un rôle fondamental dans l'activation et l'inactivation des gènes. La caractéristique de cette modification épigénétique est influencée par des facteurs environnementaux comme la nutrition et le stress.

La connaissance du génome et de ses gènes n'est pas suffisante pour expliquer le fonctionnement d'un gène, d'une cellule ou d'un organisme. Pour comprendre un système biologique complexe, il faut étudier la régulation et l'expression de ses gènes, la fonction des protéines exprimées, la présence quantitative de métabolites et les effets des anomalies génétiques sur le phénotype d'un organisme. Outre les connaissances sur les gènes, la fonction du produit génique doit également être connue. L'étude de cette complexité est souvent appelée **biologie des systèmes**, qui tente de comprendre des organismes biologiques complets et la dynamique derrière un système biologique dans son ensemble. Son objectif est d'obtenir une image intégrée de tous les processus de régulation à tous les niveaux, du génome au protéome et au métabolome, et du comportement d'une seule protéine à l'organite et à la biomécanique de l'organisme complet.

Les méthodes modernes d'analyse fonctionnelle des génomes (génomique fonctionnelle) sont appelées **transcriptomique**, **protéomique** et **métabolomique**. Il s'agit généralement de procédures à haut débit qui imposent de lourdes exigences en matière de gestion et d'analyse des données. Ces approches sont complétées par des analyses phénotypiques d'organismes et de cellules modèles *in vitro*, également dans un format à haut débit. Le phénotype décrit tous les phénotypes et son analyse est appelée phénotomique.

Transcriptomique

Malheureusement, les fonctions de nombreuses protéines basées uniquement sur les séquences nucléotidiques sont inconnues. Cependant, les informations concernant la régulation et l'expression des gènes peuvent donner un aperçu sur les fonctions des produits des gènes dans les cellules, les tissus et les organismes. Par exemple, comme un gène est exprimé exclusivement dans les cellules musculaires, on peut en déduire que le produit du gène est probablement important pour la physiologie de ce type de cellule. De nombreuses techniques existent pour analyser la régulation et l'expression des gènes, par exemple le Northern Blot, qui est une méthode de détection de l'ARNm dans des gels d'agarose utilisant l'hybridation d'acide nucléique, ou la réaction en chaîne par polymérase de transcriptase inverse (RT-PCR), une technique d'amplification de séquences nucléotidiques spécifiques dérivées de l'ARNm. Cependant, ces méthodes ne permettent que l'analyse simultanée de quelques gènes seulement et ne conviennent pas à l'analyse efficace de grandes quantités de données. Par conséquent, il est devenu nécessaire de développer des procédures à haut débit permettant une analyse plus rapide des transcriptomes entiers.

La quantification de l'ARNm par **les puces à ADN** (DNA microarrays) ou **SAGE** fournit des informations importantes sur les fonctions cellulaires potentielles des produits géniques. Cependant, la mesure de l'ARNm seule n'est pas suffisante pour décrire complètement et précisément des systèmes biologiques complexes. Les activités cellulaires comme les processus métaboliques sont médiées par les protéines du protéome et non par les gènes du génome ou l'ARNm du transcriptome. De manière analogue à la technologie des puces à ADN, des procédures à haut débit ont donc été développées pour l'analyse fonctionnelle parallèle des protéines, c'est-à-dire la protéomique.

La protéomique

La protéomique est classée en deux catégories : la protéomique classique ou quantitative et la protéomique fonctionnelle. La protéomique classique traite de l'identification et de la quantification des protéines dans les lysats cellulaires, tandis que l'objectif de la protéomique fonctionnelle est la détermination de la fonction des protéines.

Afin d'étudier l'ensemble des protéines exprimées par un organisme ou un type cellulaire donné, **l'électrophorèse bi-dimensionnelle** (2-DE) introduite par O'Farrell en 1975 est l'une des méthodes qui permet, à l'issue de deux migrations électrophorétiques successives, de visualiser un grand nombre de protéines de façon simultanée. Dans la première dimension, les protéines migrent dans un gradient de pH selon leur point isoélectrique. Cette étape, appelée isoélectrofocalisation (IEF), est suivie d'une seconde migration, généralement menée dans une direction perpendiculaire à la première, où les protéines sont séparées selon leurs poids moléculaires. L'avancée majeure en termes d'identification des protéines des gels 2D correspond à l'utilisation conjointe de la spectrométrie de masse et des banques de données. En s'adaptant à la biologie structurale, la **spectrométrie de masse** a permis de déterminer avec une sensibilité et une précision extrêmes la masse des molécules.

Métabolomique

L'ensemble de métabolites d'une cellule est appelé **métabolome**, et le domaine de recherche traitant du profilage métabolique est appelé **métabolomique**. La métabolomique est un domaine de recherche relativement nouveau, bien qu'en 1970 Robinson et Pauling aient déjà décrit des expériences pour identifier et quantifier les métabolites dans l'urine humaine. La base de données des métabolites humains [**hmdb**] contient tous les métabolites qui peuvent être trouvés dans le corps humain ou du moins devraient vraisemblablement se produire.

Phénomique

Le phénotype ou l'apparence physique est la somme de toutes les caractéristiques visibles extrinsèques d'un individu. Il fait référence à la fois aux propriétés morphologiques et physiologiques. Par conséquent, les propriétés visibles et mesurables d'un organisme ou d'une cellule qui sont basées sur les interactions du génotype avec l'environnement constituent le phénotype. Par cette définition, donc, la métabolomique, étant mesurable, est aussi une représentation d'un phénotype qui est basée sur les interactions du génotype avec l'environnement.

De nombreuses méthodes existent dans le contexte de la génomique fonctionnelle qui définissent la fonction des protéines en fonction des phénotypes. Ce domaine de recherche est également appelé phénomique s'il est mené dans un format à haut débit. Initialement, des cribles génétiques directs ont été utilisés dans lesquels les génomes ont été mutés de manière aléatoire, les phénotypes résultants enregistrés et les gènes responsables du phénotype modifié identifiés. Grâce à cette approche, plusieurs milliers de gènes ont été identifiés et caractérisés. L'arrivée de génomes entièrement séquencés a offert des approches alternatives pour effectuer des cribles génétiques pour ces gènes sans fonction attribuée. La stratégie qui relie un gène distinct à sa fonction s'appelle la génétique inverse.