

Du génome au « transcriptome » : un saut de puce !

L'irruption du couple Internet-banques de données a bouleversé en un temps très court la vie de nombreux laboratoires. Les chercheurs dont l'objet d'étude est la levure sont atteints de plein fouet par ce changement majeur. Depuis un peu plus d'un an, ayant à leur disposition l'ensemble de la séquence du génome de leur organisme d'étude, le séquençage par lui-même n'est plus pour eux une activité de découverte mais souvent de contrôle subsidiaire. Cela a bouleversé de nombreux projets, réduisant des années-projets à des mois, des mois à des jours. Cette remise en cause brutale de la programmation de la démarche expérimentale n'est pas facile car l'activité doit alors se reporter sur d'autres approches. Une de celles-ci, grande consommatrice de temps, est l'étude de l'expression des gènes et de sa régulation. Après quelques précurseurs, l'équipe de Patrick Brown (Stanford, CA, USA) vient de démontrer que, dans ce domaine, la fracture avec le passé sera encore plus dure [1].

En effet, jusqu'à récemment, la méthode expérimentale la plus couramment utilisée pour évaluer l'expression d'un gène consistait à suivre l'expression de ce gène par hybridation d'une sonde spécifique d'ADN sur l'ensemble des d'ARN extraits dans une condition donnée (technique dite de *northern blot*). Cette méthode, généralement réservée à l'étude d'un faible nombre de gènes, a été appliquée à l'échelle d'un chromosome de levure [2-4] mais n'est pas envisageable en routine à l'échelle d'un génome, fût-il aussi petit que celui de la levure. L'utilisation de fusions entre le gène étudié et un gène rapporteur (*lacZ* ou, plus récemment, le gène de la GFP *green fluorescent protein*) a également contribué de façon importante à l'étude de l'expression individuelle des gènes. Des méthodes plus globales telles que l'électrophorèse bidimensionnelle des protéines [5] ou les

banques de gènes fusionnés* [6, 7] se sont avérées être des outils s'appliquant à un plus grand nombre de gènes. Elles restent cependant laborieuses et éloignées d'une mesure de la transcription. Enfin, il existe des méthodes prédictives [8], telles que la recherche de séquences régulatrices dans les régions promotrices des gènes. Ce type de méthode a l'avantage d'être global mais l'inconvénient majeur de ne pas être expérimental. Les résultats ainsi obtenus ne relèvent pas de la preuve scientifique mais de l'hypothèse de travail.

Dans ce contexte, l'article de DeRisi *et al.* sur l'étude de l'expression génique à l'échelle génomique chez la levure *Saccharomyces cerevisiae*, est une percée technologique importante [1]. La technique utilisée est celle des *DNA microarrays* (parfois appelées *DNA chips* ou puces ADN [9]) qui permet l'hybridation d'acides nucléiques marqués sur une matrice portant des milliers de *spots* d'ADN précisément positionnés et donc repérables.

En résumé, l'ADN de 6 116 cadres ouverts de lecture, identifiés lors du séquençage systématique du génome de levure, a été amplifié individuellement par la technique de réaction de polymérisation en chaîne (PCR) à l'aide d'amorces commercialement disponibles (Research Genetics). Chaque produit de PCR a été ensuite déposé individuellement à un endroit précis sur une lamelle de microscopie recouverte de poly-L-lysine. Cette plaque de verre de 18 mm sur 18 mm forme donc une matrice portant les quelque 6 000 cadres ouverts de lecture du génome de la levure. L'ADN une fois fixé est dénaturé et est prêt à être hybridé avec une sonde (un acide nucléique marqué). Les auteurs utili-

sent des ADN complémentaires (ADNc) marqués par des fluorophores comme sonde. Le principe est le suivant: les ARN polyadénylés de levure sont extraits à partir de cellules dans deux conditions physiologiques différentes (conditions A et B). A partir de ces ARN, une synthèse d'ADNc est effectuée par transcription inverse en présence d'un analogue de nucléotide fluorescent. Dans la condition A un analogue émettant à une longueur d'onde dans le vert est incorporé tandis que dans la condition B un analogue émettant dans le rouge est incorporé. Les ADNc synthétisés dans les conditions A et B sont mélangés et hybridés sur la matrice. Chacun des fluorophores est alors activé à la longueur d'onde appropriée et l'émission de fluorescence (qui reflète l'abondance relative de chacun des ADNc) est mesurée pour les quelque 6 000 coordonnées de la matrice correspondant à un dépôt d'ADN. Le rapport d'émission de fluorescence verte et rouge en un point de la matrice reflète la différence d'expression de cette phase ouverte de lecture entre les conditions physiologiques A et B. Pour valider leur méthodologie, les auteurs présentent trois exemples d'applications. Ils étudient tout d'abord l'expression du génome au cours du passage du métabolisme fermentaire au métabolisme respiratoire lors de l'épuisement du glucose du milieu en aérobiose. Leurs résultats confirment les travaux de nombreux laboratoires depuis des années, montrant que cette transition métabolique s'accompagne, d'une part, de la diminution de l'expression de nombreux gènes (notamment ceux spécifiant des constituants de l'appareil de traduction cytosolique: protéines ribosomiques et autres facteurs) et, d'autre part, d'une augmentation de l'expression des gènes participant au métabolisme respiratoire (constituants de la chaîne respiratoire, du cycle de Krebs, enzymes de la néoglucogénèse et du stockage des carbohydrates). Dans les deux autres exemples, les auteurs

* Les banques de gènes fusionnés sont des banques d'expression consistant en une collection de plasmides portant chacun une fusion entre un gène de levure et le gène *LacZ*. Dans de telles banques, le gène rapporteur est placé sous contrôle des signaux de régulation des gènes de levure et son expression reflète donc celle du gène auquel il est fusionné.

montrent que l'on peut, par cette méthodologie, étudier l'effet de l'absence (*knock-out*) ou de la surexpression de facteurs de transcription, permettant ainsi d'identifier leurs cibles potentielles. Leurs résultats présentés sur un site internet [10] sont consultables librement. Toutes les données quantitatives pour les 6116 phases ouvertes de lecture sont accessibles et permettent de connaître l'expression de n'importe quel gène de levure dans les conditions expérimentales qu'ils ont utilisées. Il reste, certes, de nombreux aspects à optimiser, mais le nombre important de témoins que l'on peut introduire sans alourdir l'expérience rendra possible une meilleure délimitation de la signification des résultats.

Ce qui est ici démontré, c'est que l'on peut désormais étudier l'expression génique à une échelle génomique et que donc, à terme, il sera envisageable de considérer l'expression d'un génome comme une donnée aussi bien établie que sa séquence. On passe ainsi d'une description motivée de l'expression d'un gène à celle, aveugle,

de milliers de gènes, sans justification *a priori* du choix des gènes. La démarche est bien strictement parallèle à celle qui fut à la base du séquençage des EST et des génomes [11]. Ainsi, cet article établit la notion de « transcriptome » et celle-ci s'impose parallèlement à celle de génome. Les conséquences de cette cassure technologique seront sans doute parallèles à celles provoquées par l'apparition des banques de données structurales. D'une part, nombre de laboratoires devront vite, très vite, changer encore une fois leur programmation expérimentale. D'autre part, l'appropriation imaginative de cet océan de données fera sans doute la différence, dans les domaines fondamentaux et appliqués. Le couplage étroit entre biologie et traitement de l'information devient une nécessité à laquelle la politique scientifique devra répondre.

**B.D.F.
M.A.**

1. DeRisi JL, Iyer VL, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; 278: 680-6.

2. Yoshikawa A, Isono K. Chromosome III of *Saccharomyces cerevisiae*: an ordered clone bank, a detailed restriction map and analysis of transcripts suggest the presence of 160 genes. *Yeast* 1990; 6: 383-401.

3. Yoshikawa A, Isono K. Construction of an ordered clone bank, and systematic analysis of the whole transcripts of chromosome VI of *Saccharomyces cerevisiae*. *Nuc Acids Res* 1991; 19: 1189-95.

4. Richard GF, Fairhead C, Dujon B. Complete transcription map of yeast chromosome XI in different life conditions. *J Mol Biol* 1997; 268: 303-21.

5. Boy-Marcotte E, Tadi D, Perrot M, Boucherie H, Jacquet M. High cAMP levels antagonize the reprogramming of gene expression that occurs at the diauxic shift in *Saccharomyces cerevisiae*. *Microbiology* 1996; 142: 459-67.

6. Ruby SW, Szostak JW, Murray AW. Cloning regulated yeast genes from a pool of lacZ fusions. *Methods Enzymol Part C* 1983; 101: 235-69.

7. Dang VD, Valens M, Bolotin-Fukuhara M, Daignan-Fornier B. A genetic screen to isolate genes regulated by the yeast CCAAT-box binding protein Hap2p. *Yeast* 1994; 10: 1273-83.

8. Fondrat C, Kalogeropoulos A. Approaching the function of new genes by detection of their potential upstream activation sequences in *Saccharomyces cerevisiae*: application to chromosome III. *Curr Genet* 1994; 25: 396-406.

9. Bellis M, Casellas P. La puce-ADN : un multi-réacteur de paillasse. *Med Sci* 1997; 13: 1317-24.

10. cmgm.stanford.edu/pbrown/explore/index.html

11. Jordan B. La valse des étiquettes. *Med Sci* 1995; 11: 273-6.