Journal de la Société de Biologie, 196 (4), 303-307 (2002)

Étude des transcriptomes par analyse sérielle de l'expression des gènes

par Jacques Marti, David Piquemal, Laurent Manchon et Thérèse Commes

Institut de Génétique Humaine, UPR CNRS 1142, 141 rue de la Cardonille, 34396 Montpellier Cedex 5

Reçu le 18 juin 2002

RÉSUMÉ

L'Analyse Sérielle de l'Expression des Gènes (acronyme anglais SAGE) permet d'identifier l'ensemble des gènes exprimés dans un échantillon cellulaire ou tissulaire. Elle se fonde sur l'analyse séquentielle d'un grand nombre de courts fragments d'ADNc, dont chacun représente la signature d'un gène. Leur dénombrement donne une mesure précise de leur niveau d'expression. Ce système analytique ouvert n'exige aucune hypothèse préalable sur les gènes exprimés. L'analyse peut donc révéler des produits de transcription encore inconnus, contrairement aux systèmes fondés sur l'hybridation à des collections de sondes nucléiques, exigeant une caractérisation préalable des gènes analysés. Ses caractéristiques font de la méthode SAGE un outil de découverte de nouveaux gènes et de marqueurs potentiels d'états pathologiques. Elle permet de sélectionner rapidement de nouveaux paramètres de diagnostic et d'évaluation de l'efficacité d'agents thérapeutiques. Toutes les données SAGE peuvent être réunies dans une base informatique unique, où chaque nouvelle analyse bénéficie de l'ensemble des données précédentes. Cette première étape dans l'inventaire détaillé des composants cellulaires ouvre de nouvelles perspectives pour la modélisation *in silico* des fonctions biologiques.

SUMMARY Study of transcriptomes by serial analysis of gene expression

The availability of the sequences for whole genomes is changing our understanding of cell biology. Functional genomics refers to the comprehensive analysis, at the protein level (proteome) and at the mRNA level (transcriptome) of all events associated with the expression of whole sets of genes. New methods have been developped for transcriptome analysis. Serial Analysis of Gene Expression (SAGE) is based on the massive sequential analysis of short cDNA sequence tags. Each tag is derived from a defined position within a transcript. Its size (14bp) is sufficient to identify the corresponding gene and the number of times each tag is observed provides an accurate measurement of its expression level. Since tag populations can be widely amplified without altering their relative proportions, SAGE may be performed with minute amounts of biological extract. Dealing with the mass of data generated by SAGE necessitates computer analysis. A software is required to automatically detect and count tags from sequence files. Criterias allowing to assess the quality of experimental data can be included at this stage. To identify the corresponding genes, a database is created registering all virtual tags susceptible to be observed, based on the present status of the genome knowledge. By using currently available database functions, it is easy to match experimental and virtual tags, thus generating a new database registering identified tags, together with their expression levels. As an open system, SAGE is able to reveal new, yet unknown, transcripts. Their identification will become increasingly easier with the progress of genome annotation. However, their direct characterization can be attempted, since tag information may be sufficient to design primers allowing to extend unknown sequences. A major advantage of SAGE is that, by measuring expression levels without reference to an arbitrary standard, data are definitively acquired and cumulative. All publicly available data can thus be stored in a unique database, facilitating whole-genome analysis of differential expression between cell types, normal and diseased samples, or samples with and without drug treatment. SAGE data are readily amenable to statistical comparisons, allowing to determine the level of confidence of the observed variations. A major limitation of SAGE is that, because each analyse is

obligatory performed on the whole set of expressed genes, it can hardly be performed on multiple samples, for example in kinetics studies or to compare the effects of large numbers of drugs. To overcome this limitation, high-throughput detection of a subset of mRNAs is more rapidly performed by parallel hybridization of mRNAs on arrays of nucleic acids immobilized on solid supports. From this point of view, a SAGE platform is a powerful instrument for selecting the most informative subset of genes, assembling them to design microarrays dedicated to a specific problem and calibrating measurement by comparison with a standard cell model for which SAGE data are available. This approach is an attractive alternative to strategies based exclusively on pangenomic arrays. A very large amount of SAGE data are already available and the problem is now to extract their biological meaning. Knowledge on metabolic pathways is already organized so that its successfull integration in a SAGE platform can be undertaken. For other cell components and pathways, the problem lies on the lack of controled vocabulary to describe gene activites, starting form a clear definition of the concept of biological function itself. Progress in gene and cell ontology is expected to facilitate computer-based extraction of biological knowledge from existing and forthcoming SAGE data.

DE LA GÉNOMIQUE STRUCTURALE À L'ANALYSE DES TRANSCRIPTOMES

De lourds investissements ont été réalisés dans l'analyse moléculaire des génomes, en particulier le génome humain. Bien que ce travail ne soit pas terminé, il est désormais possible d'envisager de nouvelles applications exploitant les données déjà disponibles. Il n'existe pas de méthode permettant de prédire directement le phénotype observé à l'échelle cellulaire ou tissulaire, en se fondant uniquement sur l'analyse des génomes. Il a donc fallu créer de nouveaux outils d'analyse pour compléter les données de génomique structurale. Ils permettent à présent l'émergence d'une génomique fonctionnelle dont l'ambition est de prendre en compte, dans l'étude d'un phénomène biologique, l'ensemble des agents impliqués. La première phase consiste à dresser l'inventaire de tous les composants du système, en particulier les protéines (protéome), puis à établir l'ensemble de leurs interactions physiques et fonctionnelles grâce à de nouvelles approches, décrites par d'autres néologismes : interactome, métabolome, physiome, etc... L'analyse des protéomes est limitée par la lourdeur et le manque de sensibilité des techniques d'analyse. Par contre, la possibilité d'amplifier à volonté l'ADN par PCR, la puissance des instruments d'analyse séquentielle, le fait que chaque fragment puisse être lui-même utilisé comme sonde spécifique, permettent désormais d'envisager l'analyse complète des produits de transcription des gènes, ou transcriptomes. Alors que le génome est reproduit à l'identique dans chaque cellule, il existe autant de transcriptomes que de types cellulaires. Chaque analyse implique l'identification et le dosage de plus de dix mille ARNm différents. Les méthodes analytiques doivent donc assurer le meilleur compromis entre qualité et rapidité, soit en utilisant des systèmes massivement parallèles, soit en accélérant le débit d'une analyse sérielle.

Les méthodes massivement parallèles sont fondées sur l'hybridation simultanée des produits de transcription à de grandes collections de sondes nucléiques, l'analyse étant basée sur l'exploitation des signaux de fluorescence ou de radioactivité. Les sondes sont des fragments d'ADN néo-synthétisés ou amplifiés à partir de plasmides et fixés sur un support solide (« macroarrays » sur des filtres de nylon ou « microarrays » sur des lames de microscope). Dans le procédé Affymetrix, les oligonucléotides sont synthétisés directement sur le support.

D'autres méthodes impliquent l'analyse individuelle d'un grand nombre de fragments dérivés des produits de transcription, la quantification résultant directement de leur dénombrement. Sur ce principe, la méthode SAGE (*Serial Analysis of Gene Expression*) a été développée en 1995 par Velculescu, Vogelstein et Kinzler à la Johns Hopkins University (Baltimore MD, USA) (Velculescu *et al.*, 1995). Les améliorations considérables apportées en 1999 par J-M. Elalouf (CEA, Saclay, France) ont débouché sur un protocole standard permettant de réaliser une analyse complète avec des quantités très faibles d'ARN (*Serial Analysis of Downsized Extracts*, SADE) (Virlon *et al.*, 1999).

Ces nouveaux systèmes analytiques à haut débit produisent une masse de données brutes qu'il est impossible d'appréhender directement. En parallèle, il a donc fallu développer une panoplie d'outils informatiques pour traiter les données et en extraire les informations pertinentes, impliquant la conception de nouveaux logiciels, la construction de bases de données et la définition de nouveaux standards de représentation symbolique des connaissances.

PRINCIPE DE LA MÉTHODE SAGE

Il n'est pas nécessaire de séquencer la totalité de l'ADN complémentaire (ADNc) rétro-transcrit à partir d'un ARNm pour obtenir une signature (tag) du gène correspondant. Partant de ce constat, plusieurs grands programmes ont été lancés pour séquencer massivement des fragments de séquences exprimées (ESTs, pour

Expressed Sequence Tags) afin d'identifier rapidement le plus grand nombre possible de produits transcription.

La méthode SAGE reprend ce principe, mais en limitant la taille des tags au strict minimum, ce qui réduit d'autant la durée et les coûts d'analyse. L'identification n'est pas seulement fondée sur la séquence, mais également sur la position des tags, extraits de régions topologiquement bien définies. Pour cela, les ARNm sont d'abord capturés sur des billes magnétiques recouvertes d'oligo-dT, où ils sont directement convertis en ADNc. Ceux-ci sont digérés par une enzyme de restriction (enzyme d'ancrage) reconnaissant 4 nucléotides (GATC pour Sau3A1 ou CATG pour NlaIII). Les combinaisons de 4 bases étant très fréquentes (1/256), la plupart des ADNc sont fragmentés et seul un petit nombre de gènes échappent à l'analyse. Le fragment conservé est copié sur l'extrémité 3′ de l'ARNm, dans la zone où la synthèse de l'ADNc est optimale. Après lavage et élimination d'environ 80 % des produits de digestion, seul ce fragment reste fixé aux billes magnétiques. Son extrêmité cohésive (GATC), produite par la coupure enzymatique, permet d'y lier un adaptateur synthétique. Celui-ci apporte le site de fixation d'une enzyme de type IIs, généralement BsmFI, directement au contact du site GATC. Cette enzyme clive l'ADN 10 à 11 bases en aval du site GATC, excisant un tag de 14 ou 15 paires de bases (GATC plus 10 ou 11 bases spécifiques). Les tags ainsi libérés sont recueillis et leurs extrémités à bout franc reconstituées.

L'ensemble de cette opération est réalisée en double, sur deux échantillons identiques, en utilisant deux adaptateurs différents. Les deux préparations sont mélangées et les fragments liés deux à deux pour former des ditags. L'intérêt de cette opération est de pouvoir amplifier les ditags par PCR, en utilisant des amorces spécifiques de chacun des deux adaptateurs. Comme les tags sont très courts et assemblés au hasard, la composition des ditags est aléatoire, ce qui limite les variations inhérentes à la composition chimique et aux particularités structurales de chaque tag. L'amplification est donc homogène et respecte les proportions initiales des différents tags. De ce fait, la méthode SAGE ne requiert qu'une faible quantité de matériel. Les analyses sont réalisées habituellement à partir de 10 à 20 µg d'ARN total, mais des échantillons de taille plus faible peuvent être envisagés, jusqu'à la cellule unique.

A l'issue de l'amplification, les adaptateurs synthétiques sont éliminés par digestion enzymatique au niveau des sites GATC (ou CATG selon l'enzyme d'ancrage utilisée). Les ditags purifiés ne contiennent plus que les séquences biologiques issues des ADNc initiaux. Il sont alors assemblés sous l'action d'une ligase, pour former des concatémères dont la taille, dans un souci d'économie, doit être adaptée aux performances des séquenceurs d'ADN (de 25 tags à 30 tags par concatémère). Ces concatémères sont insérés dans un vecteur de clonage utilisé pour construire une banque où chaque clone bactérien contiendra une combinaison unique de tags. La croissance des colonies apporte un nouveau cycle d'amplification. Ici encore, la composition aléatoire des concatémères garantit un comportement homogène et une distribution des tags reproduisant celle des ARNm initiaux. En contrepartie, le brassage aléatoire des tags ne permet pas de circonscrire l'analyse à un sous-ensemble de gènes. Une banque SAGE contient typiquement de 1 à 2 millions de tags, répartis dans environ 50 000 colonies bactériennes. Dans la plupart des projets, l'analyse de 50 000 tags, issus d'environ 2 000 colonies et permettant d'identifier environ 10 000 gènes différents, s'avère suffisante. La banque peut être conservée pour des analyses plus approfondies.

ANALYSE INFORMATIQUE DES DONNÉES SAGE

Le résultat brut d'une analyse SAGE est un fichier-texte où les chaînes de caractères représentant les séquences des ditags sont ponctuées par les quatre lettres GATC (ou CATG). La première étape est de mesurer la longueur des ditags : l'enzyme BsmFI ne coupant pas tous les tags à la même longueur, seuls ceux auxquels au moins dix nucléotides peuvent être assignés sans ambiguïté sont conservés. Un faible pourcentage de ditags trop courts sont donc éliminés (< 2 %). A l'inverse, certains ditags ont apparemment deux fois la longueur attendue, en raison de l'absence de site séparateur. Celle-ci provenant d'une erreur de séquence, leur dénombrement donne une évaluation du taux d'erreurs de séquençage.

A partir d'un extrait cellulaire contenant des milliers d'ARNm différents, la probabilité de synthétiser plusieurs fois le même ditag est très faible. L'observation répétée de ditags identiques peut refléter la présence normale d'ARNm très abondants. Elle peut également résulter d'un artefact. En effet, dans la construction d'une banque SAGE à partir d'un échantillon très réduit, le risque majeur est de n'amplifier que les tags issus des ARNm les plus abondants et d'obtenir une banque de complexité réduite. Généralement, le taux de ditags identiques est inférieur à 5 % (Cheval *et al.* 2000). Au-delà, il faut s'interroger sur la représentativité de la banque et sur l'intérêt d'inclure ces tags dans le dénombrement final.

Cette première partie de l'analyse se termine par la construction d'une liste tabulée des tags sélectionnés, enregistrant leur séquence et le nombre de fois où ils ont été observés. Ces résultats sont définitivement acquis. Ils ne demandent aucune calibration relativement à un standard externe et sont directement exploitables, aussi bien sur le plan quantitatif que qualitatif. Si besoin est, il est toujours possible d'accroître la profondeur d'analyse en séquençant de nouveaux clones à partir de la même banque.

Pour identifier les gènes correspondants aux tags, la stratégie la plus efficace est de construire une base de données où sont enregistrés tous les tags potentiellement observables en l'état actuel des connaissances. Cette opération revient à réaliser *in silico* une analyse SAGE virtuelle portant sur l'ensemble des produits de transcription

possible. Pour un génome parfaitement annoté, on obtiendrait une relation biunivoque entre les tags et les exons dont ils sont issus, avec un taux d'ambiguïtés résiduel concernant les gènes ou variants d'épissage dont les tags SAGE sont identiques. En l'état actuel d'annotation du génome humain, on peut estimer à moins de 50 % le nombre de gènes dont le tag est identifié sans ambiguïté. On sait déjà que certaines ambiguïtés ne pourront pas être levées directement, du fait de la présence de séquences répétées, essentiellement des séquences *Alu*, dans la région 3'd'ARNm issus de gènes par ailleurs totalement différents. Néanmoins ces séquences sont faciles à repérer et les tags correspondants sont enregistrés dans un fichier séparé.

Pour les gènes encore imparfaitement annotés, la solution est d'exploiter les grandes collections de séquences d'ADNc et de fragments de séquences exprimées (ESTs). Leur abondance (> 3 millions) chez l'Homme montre que de nombreux fragments ont été séquencés de façon redondante, obligeant à regrouper l'ensemble des séquences identifiant un même gène. La collection *UniGene (National Center for Biotechnology Information*, NCBI ; Bethesda, MD, USA) (http://www.ncbi.nlm.nih.gov/UniGene/) rassemble ainsi 96 000 groupes, ou clusters, chacun défini par une séquence unique. L'analyse informatique de ces séquences (Hs.seq.uniq) permet d'extraire le tag représentatif de chaque cluster. Un logiciel courant de gestion de base de données autorise une confrontation immédiate des tags prédits et des tags expérimentaux, alimentant automatiquement une base où peuvent être compilées toutes les données SAGE existantes. Cette base de données peut être exploitée selon de multiples modalités.

UTILISATION DES DONNÉES SAGE POUR LA RECHERCHE DE NOUVEAUX GÈNES

Dans la mesure où la méthode SAGE ne requiert aucune connaissance préalable du génome, elle permet à l'inverse d'identifier de nouveaux gènes. Chez l'Homme, la masse de données actuelles suggère que, pratiquement, tous les produits d'expression ont été repérés. Cependant, on observe encore une proportion significative de tags SAGE non identifiés. Il est probable que certains types cellulaires ont échappé à ces analyses, notamment des cellules spécialisées ou les rares cellules souches à l'origine de différents tissus. Le problème du nombre de gènes et de produits de transcription encore inconnus reste donc d'actualité.

Dans un programme de recherche de nouveaux gènes, le problème est de savoir à quel moment cesser l'acquisition de nouvelles séquences : il n'a pas de solution évidente. En principe, la pente de la courbe représentant le nombre de tags nouveaux en fonction du nombre de tags séquencés devrait s'annuler. Elle reste positive à cause d'erreurs de séquence : un taux d'erreur de 1 % sur chacun des 10 nucléotides introduit une erreur dans un tag sur 10, soit environ 5 000 tags dans une analyse de 50 000. Une insertion de nucléotides erronés a pu également survenir au cours de l'amplification des ditags par PCR. Le risque d'observer plusieurs fois la même erreur sur la même position est limité, mais les tags observés une seule fois restent sujets à caution.

Outre ce bruit de fond, la recherche de nouveaux gènes implique l'élimination préalable de quelques artefacts. Des tags correspondant à des fragments d'ARN ribosomal ou dans l'orientation inverse du tag attendu, peuvent signaler une souillure de la préparation initiale par des produits de dégradation mal éliminés. D'autre part, il n'est pas possible de digérer les ADNc avec une efficacité absolue. Des tags provenant du site en 5' du site de coupure attendu sont donc prévisibles, mais de ce fait même ils sont aisément détectés. On ne les observe que dans le cas d'ARNm abondants, avec une fréquence moyenne inférieure à 1 %. Un taux élevé suggèrera l'existence d'un variant d'épissage ou d'un site de polyadénylation alternatif, qu'il est possible de confirmer expérimentalement. Un intérêt de l'approche SAGE est que les résultats peuvent être régulièrement critiqués pour assainir et consolider les données.

Une fois arrêtée la liste des tags candidats, des solutions techniques existent pour poursuivre leur caractérisation. L'élongation de la séquence en 3' peut être réalisée par PCR en utilisant une amorce spécifique ancrée sur la séquence du tag et une amorce universelle s'hybridant sur l'extrémité polyA. Cette stratégie et ses variantes permettent d'obtenir des séquences plus longues, de l'ordre de 50 à 100 nucléotides, qui peuvent être utilisées pour lever les ambiguïtés dans le cas de tags identifiant plusieurs gènes, confirmer l'identification d'ESTs ou d'ADNc, rechercher des gènes entièrement inconnus par confrontation avec la séquence du génome complet ou par comparaison avec les gènes déjà connus dans d'autres espèces.

ANALYSE COMPARÉE DES TRANSCRIPTOMES

Un avantage de la méthode SAGE est que chaque nouvelle analyse s'enrichit de la confrontation avec les précédentes. Le format des données SAGE permet de comparer automatiquement les niveaux d'expression entre banques issues du même laboratoire ou de laboratoires différents, en utilisant la séquence du tag lui-même comme identifiant. On peut ainsi distinguer les gènes plus spécifiquement exprimés dans un type cellulaire, une situation physiologique ou un état physio-pathologique donné. La valeur de ces comparaisons doit être appréciée par une analyse statistique des niveaux d'expression. Lors du séquençage des concatémères, l'observation successive des différents tags peut être assimilée à un tirage au hasard dans une population de grandes dimensions, selon une distribution binomiale. Suivant ce principe, on sait calculer la probabilité P que les variations de niveau d'expression d'un même gène dans deux situations résultent de fluctuations aléatoires

(Alric and Marti 2000). Plus P est faible, plus la signification biologique de la variation observée est grande. L'expérience montre que des valeurs de $P<10^{-3}$ peuvent être observées à partir de l'analyse de 20 000 à 30 000 tags par transcriptome identifiant plus de 5 000 gènes. Une analyse SAGE de faible profondeur peut donc s'avérer suffisante pour mettre en évidence des effets biologiques significatifs, sélectionner les marqueurs de diagnostic les plus spécifiques ou identifier des paramètres permettant d'évaluer l'action d'un agent toxique ou d'intérêt thérapeutique.

APPLICATION À LA GÉNOMIQUE FONCTIONNELLE

La connaissance des profils d'expression de gènes ne suffit pas à donner une représentation complète des composants cellulaires. Partant du transcriptome, déduire le protéome d'une cellule supposerait que l'on dispose de l'ensemble des informations relatives à la durée de vie des ARNm, à l'efficacité de traduction et à la durée de vie des protéines. Or, chacun de ces processus est contrôlé par d'autres protéines, organisées en réseaux interactifs, dont il faudrait savoir modéliser le fonctionnement dynamique, alors que l'existence de ces réseaux est déduite de l'analyse elle-même. La connexion complète transcriptome-protéome-interactome n'est encore envisageable qu'à l'échelle de microorganismes.

En deçà d'un objectif aussi ambitieux, des prédictions locales peuvent être formulées, concernant des groupes de gènes dont l'expression concertée suggère qu'ils sont contrôlés par des mécanismes communs ou qu'ils relèvent d'une même fonction cellulaire. Cette stratégie s'applique notamment aux voies métaboliques. Plusieurs projets visent à intégrer les connaissances relatives au métabolome. Ils portent principalement sur les bactéries, mais la base informatique KEGG gère également les connaissances relatives à l'organisme humain (*Kyoto Encyclopedia of Genes and Genomes*, Université de Kyoto, Japon) (http://www.genome.ad.jp/kegg/). Son environnement graphique permet d'intégrer automatiquement les données SAGE dans le schéma des voies métaboliques d'une cellule, donnant une vue synthétique de son répertoire de fonctions.

Une représentation automatique des voies signalétiques n'est pas encore disponible, faute d'une description exacte de leurs composants. Pour de nombreux gènes, la principale difficulté reste encore de coder leur activité biologique selon un vocabulaire contrôlé, admis par la communauté des biologistes, avant même de concevoir une interface homme-machine interactive et assez conviviale pour ne pas faire obstacle au travail du biologiste. Les données SAGE étant définitivement acquises, l'utilisation de cette approche pour élaborer de nouvelles hypothèses de travail et planifier les futures expériences représente déjà un apport considérable, contribuant dès à présent au développement de la génomique fonctionnelle.

CONCLUSION

Les profils d'expression de gènes obtenus par la méthode SAGE donnent une approximation réaliste d'un transcriptome cellulaire. La lourdeur de chaque analyse est la contrepartie de sa robustesse, de sa précision sur le plan quantitatif et de sa quasi-exhaustivité sur le plan qualitatif. Les performances des séquenceurs d'ADN constituent à ce jour le principal facteur limitant mais on peut prévoir une accélération du débit analytique dans les prochaines années. Pour l'instant, la méthode SAGE ne permet pas de comparer rapidement un grand nombre d'échantillons ou de réaliser l'étude cinétique détaillée d'un phénomène. Par contre, une base intégrant l'ensemble des données SAGE constitue probablement l'outil le plus efficace pour sélectionner les paramètres décisifs et limiter le nombre de sondes nucléiques au minimum nécessaire pour concevoir et calibrer des outils analytiques (macro- et micro-arrays) dédiés à l'étude d'un problème spécifique. Il serait intéressant que cette stratégie soit évaluée méthodiquement, relativement à celles fondées uniquement sur l'utilisation de puces à ADN pangénomiques, qui posent d'autres problèmes d'authentification des signaux, de validation des mesures ou de choix d'un extrait cellulaire de référence pour des comparaisons inter-laboratoires. L'exploitation des données SAGE est tributaire des progrès en bioinformatique. Le problème actuel est de pouvoir, au sens littéral du terme, prendre connaissance de la masse de données acquises, que la simple lecture des fichiers textuels ne permet pas de maîtriser. De nouveaux systèmes de gestion des connaissances sont nécessaires pour organiser les données, rassembler les gènes en fonction de leur appartenance à un complexe sub-cellulaire, une voie métabolique ou signalétique commune, et visualiser les résultats pour faciliter le développement de nouvelles applications.

BIBLIOGRAPHIE

Alric J. & Marti J., *JOBIM Journée Ouverte à la Biologie, l'Informatique et les Mathématiques*, Actes du Colloque. G. Caraux, O. Gascuel et M-F Sagot Edrs, 2000, pp. 15-21.

Cheval L., Virlon B. Elalouf J.M., « SADE : a microassay for serial analysis of gene expression ». In : *Functional Genomics : a pratical approach*. Edited by S.Hunt and J.-P. Livesey. Oxford University Press 2000, 2000, 139-163.

Velculescu V. E., Zhang L., Vogelstein B., and Kinzler K. W., « Serial analysis of gene expression [see comments] ». *Science*, 1995, 270(5235), 484-487.

Virlon B., Cheval L., Buhler J. M., Billon E., Doucet A., and Elalouf J. M. « Serial microanalysis of renal transcriptomes ». *Proc Natl Acad Sci U S A*, 1999, 96(26), 15286-15291.

Des informations complémentaires sont disponibles sur le site http://www.sagenet.org/.