

Analyse d'image de filtres à haute densité : méthodes d'acquisition et de traitement

Image analysis of high density arrays: acquisition and treatment methods

B. MEUNIER, K. SUDRE, I. CASSAR-MALEK, A. LISTRAT, J-F. HOCQUETTE
INRA, Unité de Recherches sur les Herbivores, 63122 St-Genès-Champanelle, France

INTRODUCTION

L'analyse du transcriptome est basée sur l'hybridation moléculaire de cibles complexes (ARN extraits d'un tissu, rétrotranscrits et marqués) avec un grand nombre de sondes (ADNc correspondant à des fragments de gènes) préalablement déposées sur un support (lame de verre, membrane de Nylon). Les résultats apparaissent sous la forme de "spots" dont l'intensité correspond au niveau d'expression des gènes. L'acquisition de l'image de ces spots est réalisée par un scanner et l'analyse de l'image est généralement effectuée par un logiciel dédié.

Le premier objectif de ce travail est de donner quelques recommandations concernant le dépôt des ADNc sur le support, l'acquisition et l'analyse de l'image pour optimiser la qualité des résultats. Ensuite quelques techniques d'extraction des données déclarées fiables sont illustrées par des exemples présentant des biais différents.

1. MATERIEL ET METHODES

Environ 400 ADNc d'intérêt ont été déposés en 2 ou 3 réplicats associés à des témoins négatifs (eau, tampon, ADNc d'*Arabidopsis*). Ce dépôt a été effectué sur lames ou sur membranes selon le marquage utilisé, respectivement fluorescence (Cy3/Cy5) ou radioactivité (^{33}P). Un diamètre minimal de spot de 10 fois la résolution du scanner et un pas minimal entre 2 spots de 2 fois leur diamètre ont été respectés afin d'assurer la détection des spots et une mesure satisfaisante de leur intensité. L'acquisition de l'image a été réalisée par un scanner (Affymetrix 428 pour Cy3/Cy5 ou PhosphorImager STORM 840 pour le ^{33}P). L'analyse de l'image sous GenePix Pro V4.1 (© Axon Instruments) a permis de calculer l'intensité (F) de chaque spot et l'intensité du bruit de fond (B) autour de ce spot (encore appelé bruit de fond local). Chaque spot a été représenté par un disque dont la position et le diamètre ont été automatiquement ajustés par le logiciel, méthode généralement préconisée. Le bruit de fond local a été déterminé dans un anneau de diamètre 3 fois supérieur à celui du spot après exclusion de celui-ci. Ainsi, la moyenne, l'écart type et la médiane des niveaux de gris de chaque spot (respectivement Fmoy, Fsd et Fmed) ainsi que du bruit de fond local (Bmed, Bsd et Bmoy) ont été quantifiés. Seule la Bmed a été retenue car elle est moins sensible aux variations artefactuelles des niveaux de gris. La moyenne globale (Bglob) du bruit de fond de l'image a été calculée en faisant la moyenne des Bmed. Nous avons défini 3 méthodes de mesure de l'intensité des spots :

$$F_1 = F_{\text{moy}} - B_{\text{med}}$$

$$F_2 = F_{\text{med}} - B_{\text{med}}$$

$$F_3 = F_{\text{med}} - B_{\text{glob}}$$

Une estimation de la dérive des bruits de fond selon les directions X et Y de l'image a été calculée comme étant la pente relative (D%) de chaque droite de régression des représentations $B_{\text{med}}=f(X)$ et $B_{\text{med}}=f(Y)$. Nous avons estimé qu'une dérive (D%) supérieure à 10 % empêchait toute quantification utilisant un bruit de fond global (F₃). L'écart moyen relatif (EMR%) entre les réplicats i a été calculé pour chaque ADNc représenté par ses n réplicats :

$$\text{EMR}\% = (|F_i - F_m| / n) / F_m \text{ avec } F_m = F_i / n$$

La méthode de mesure donnant un EMR% moyen minimal a été considérée comme la plus fiable.

Afin d'estimer la fiabilité de la mesure d'intensité de chaque spot, des critères (C₁, C₂, C₃, C₄) ont été considérés (Wang *et al.*, 2001) :

C₁ : 1 si le rapport signal sur bruit (RSB) est >2 ; 0 sinon ;
 $\text{RSB} = (F_{\text{med}} - B_{\text{med}}) / B_{\text{sd}}$

C₂ : 1 si le nombre de pixels saturés est <30 % ; 0 sinon

C₃ : 1 si l'homogénéité H du bruit est >90 % ; 0 sinon ;

$$H = 100 - (|B_{\text{med}} - B_{\text{moy}}| / \text{MAX}(B_{\text{med}}, B_{\text{moy}})) * 100$$

C₄ : 1 si EMR% <33% ; 0 sinon.

Les spots déclarés fiables doivent présenter un score S ($S = C_1 + C_2 + C_3 + C_4$) supérieur à une valeur seuil au moins égale à 3. Seules les valeurs F₁, F₂ ou F₃ de ces spots fiables sont conservées, normalisées et analysées.

2. RESULTATS

L'analyse a été réalisée sur 10 lames et 16 membranes.

L'étude de la dérive du bruit de fond a permis d'écarter la méthode de mesure F₃ s'appuyant sur un bruit de fond global (tableau 1). En effet, les dérives étaient trop importantes pour les 10 lames et pour 2 membranes sur 16.

Tableau 1 : Dérive du bruit de fond local selon les directions X et Y de l'image (moyenne ± écart type).

	Lames (n=10)	Membranes (n=16)
D% selon X	4,5 (± 6,1)	8,9 (± 5,0)
D% selon Y	19,2 (± 10,2)	11,1 (± 14,6)

Etant donné la présence de nombreux spots saturés sur les lames, seule la méthode de mesure F₂ a été employée puisque l'utilisation des valeurs médianes permet de s'affranchir de la saturation jusqu'à un seuil de 30 % (Wang *et al.*, 2001). Pour les membranes, la méthode F₁ a été choisie au vu de ses plus faibles EMR% (4,1 % pour F₁ vs 4,7 % et 5,0 % pour F₂ et F₃). Nous avons par ailleurs vérifié, par le test de Lilliefors, que la distribution des intensités d'un ADNc déposé en grand nombre (n>20) suivait bien une loi normale, hypothèse nécessaire pour l'application du test-t de Student (Hocquette *et al.*, 2003) Enfin, le calcul du score S a permis de retenir 37,7 % des ADNc pour les lames et 38,8 % pour les membranes.

CONCLUSION

La présence de réplicats ainsi que divers critères de diagnostic sur les valeurs d'intensité des spots et des bruits de fond locaux ont permis de choisir la méthode d'analyse d'image la plus appropriée. Ainsi, les biais consécutifs à chaque expérimentation implique de tester et d'adapter une méthode d'analyse spécifique.

Hocquette, J.F., Barnola, I., Bernard, C., Meunier, B., Sudre, K., Listrat, A., Cassar-Malek, I., 2003. Renc. Rech. Ruminants, 10.
Wang, X., Ghosh, S., Guo, S-W., 2001. Nucleic Acids Res., 29, 15e75.