Research Report

Automatic Quantitation of Hybridization Signals on cDNA Arrays

BioTechniques 32:1386-1397 (June 2002)

**F. Tahi, B. Achddou¹,
C. Decraene, O. Alibert,
H. Guiot¹, C. Auffray, and
G. Piétu**

Genexpress, CNRS FRE 2376,
Villejuif, and ¹COSE, Dugny,
France

ABSTRACT

Large-scale hybridization of simple or complex cDNA probes to cDNA clones arrayed on high-density filters is a method frequently used to determine systematically the expression profiles of thousands of genes. Hybridization signal intensities, which reflect the level of transcription of the corresponding genes, are captured on phosphor screens with an imaging system. We describe a high-throughput system, XdotsReader, that performs automatic detection and quantitation of each signal on hundreds of images. Reproducibility of spot detection and quantitation within filters and between filters has been assessed in analysis of more than 850 000 hybridization signals on 436 filters. The automatic analysis success was greater than 97%, with 424 of the 436 tested filters fully analyzed without any human intervention. XdotsReader is available from the Software Library at www.BioTechniques.com or at <http://www.lami.univ-evry.fr/~tahi/XdotsReader>. It runs on SUN workstations under UNIX (SunOS or Solaris) and on PC under LINUX. No particular hardware is required, and the software is compatible with any other software. It supports the main standard image formats.

INTRODUCTION

With advances in genome projects, a large number of genes have been discovered and characterized in term of sequence and map localization (6). A rapid and efficient description of expression patterns is an important step toward the goal of understanding the biological functions of these genes and changes that occur in cellular processes during development, differentiation, or disease. New analytical tools, such as DNA arrays, make it possible to collect simultaneously the expression of thousands of genes (1,3,7,10,12,15).

A first generation of macroarray platform has been developed in the Genexpress laboratory in which hybridization signals of cDNA clones arrayed on nylon filters are captured on phosphor screens with an imaging plate system (13). Analysis of the generated images requires a powerful system able to quantitate hybridization signals automatically. Many commercial software systems have been recently developed for this issue, and academic systems (2,5) have been proposed. In this paper, we present a fast and convenient system that can analyze efficiently thousands of hybridization signals generated by the cDNA array technology. The software is completely automatic, for localizing as well as quantitating the spots. Therefore, it can run in batch on a set of images. It presents features that allow high flexibility and robustness. It provides different algorithms to ensure efficient analysis of many types of images.

MATERIALS AND METHODS

Hybridization on High-Density Filters

High-density filters (8 × 12 cm) were constructed as previously described (13). PCR products from cDNA inserts are spotted at density of 16 or 25 96-well microplates/filter. In the present study, we used 1339 clones from a human skeletal muscle library (13) and 2304 cDNA clones from a normalized infant brain library (3,16). Preparation and labeling of single-stranded cDNA probes (complex probes) derived from poly (A)⁺ RNA from various tissues as well as hybridization conditions were as described previously (13).

Control filters were used to study the performances of the software. Ten concentrations (50 to 0.1 ng/μL) of phiX DNA (Invitrogen, Cergy Pontoise, France) were prepared by successive dilutions of the original solution. Each dilution was deposited in a 96-well microplate that was spotted twice on the control filters at density 25. One hundred and ninety-two values were generated for each concentration by hybridization with a ³³P-radiolabeled simple probe corresponding to phiX DNA.

Acquisition of hybridization signals was performed using the PhosphorImager® imaging plate system (Amersham Biosciences, Saclay, France). After exposure of 2–8 filters simultaneously, during 15–72 h, scanned 16-bit images were generated on a 4-log dynamic range. Finally, images represent-

ing several filters were imported on a Sun-Unix workstation and were analyzed using our system, *XDotsReader*.

The *XDotsReader* Software

The DNA array analysis consists of detection, location, and quantitation of spots with size *a priori* unknown and positions following an approximate rectangular grid of variable orientation and step. An array contains $L \times C$ blocs composed of $d = D \times D$ spots, where d corresponds to the density. Our system analyzes images of spots in which the geometry is laid out in a matrix of L lines and C columns. For a density d , there are d matrices of $L \times C$ spots, forming a grid.

The analysis is divided into different steps, from image capture to data edition:

Parameter Setting

A parameter menu is provided for defining the analysis parameters. These parameters describe the type of spotting (e.g., the density, the number of lines, and columns in the grids), the vertical and horizontal steps (corresponding to an estimation of the distance between two spots, respectively, in the vertical and the horizontal modes), the minimum and maximum radius of the spots, etc.

Image Cutting

Images obtained after scanning represent several high-density filters. In this step, the borders that delimit the individual filters to analyze are defined on each image. These borders can be determined manually or automatically. In the last case, the program uses an input file containing default coordinates that define the positions and the area of each filter on the phosphorus screen. Therefore, the cutting can be completely automatic, which allows one to achieve it in batch on a set of images.

Array Analysis

As described above, the spots are located on the intersections of a given grid. The analysis can be performed according to two modes: in spot mode,

the spots are detected and their positions are then used to set the grids. This mode requires a sufficient number of spots. In grid mode, the grids are computed first. The spots are then located on the grids. This mode has the ability to manage also very low signals.

Result Edition and Visualization

The software provides several ways to visualize the results of the analysis. Therefore, the user can verify, edit, or possibly modify them. It also provides several tools for extracting qualitative information from the spots.

RESULTS AND DISCUSSION

Hybridization Signal Analysis

The analysis consists to locate first the spots on the image, then to quantitate them.

Spot Location

XdotsReader provides two modes to analyze correctly different types of images: the spot mode and the grid mode. They are a combination of spot detection algorithms and grid position and orientation algorithms. We propose two algorithms for spot detection, *S1* and *S2*, and two algorithms for grid position and orientation, *G1* and *G2*. The spot mode analysis first detects the spots on the image, using one of the two spot detection algorithms *S1* or *S2*, then detects and sets the grids according to the detected spots using the algorithm *G1*. The grid mode first detects the grids according to the image signals using the algorithm *G2*, then detects all the spots occurring in the image, and finally selects the ones matching with the grids (i.e., spots positioned on the grid nodes).

Spot Detection

The spot detection algorithms approximate a spot as a disk of given center and radius. The radius of each spot is computed automatically. Therefore, the number of pixels constituting a spot is considered variable.

The first algorithm *S1* is based on the 2-D Gaussian model (11,14). A set

of pixels represents a spot if a “sufficient” percentage of these pixels have intensities that follow a 2-D Gaussian model. More precisely, we select the spots as follows: for each pixel (i, j) of the image, we compute the higher radius R such that for each $r \leq R$, 95% of the pixels of the circle of radius r and center (i, j) have intensities greater than the intensities of the pixels of the circle of radius $r+1$ and center (i, j) . If R is greater than 2, then the spot of center (i, j) and radius R is selected. Therefore, *S1* detects simultaneously the centers and the radius of the spots.

The second algorithm *S2* uses methods of filtering based on circular neighborhood with adaptable threshold and morphological algorithms (8,9,11,14, 18). The principle of *S2* is, given an image *image1*, to deduce an image *image2* where the pixels of *image1* constituting spots are selected, using successive filterings. In this purpose, *S2* uses: first, the parameter p , equal to the minimum between px and py , where px and py are respectively the horizontal and the vertical steps of the grid. These are given by the user in the parameter menu or calculated by the grid position algorithm in the case of the grid mode analysis. Second, the parameters $rmin$ and $rmax$, the minimum and the maximum radius of the spots respectively. They can be given by the user (in the parameter menu) or defined by the algorithm; in this last case, $rmin$ is equal to 2 and $rmax$ is equal to $p/2$.

The first filtering pass allows one to put in evidence the circular structures of the image. It works as follows. For each pixel (i, j) , it calculates the average of the pixel intensities on circles of radius of size between $rmin$ and $rmax$. It deduces from these averages the maximum and the minimum values called respectively *valmax* and *valmin*. Then, it associates to each pixel (i, j) an intensity value *image1*(i, j) equal to *valmax-valmin*.

The second filtering pass eliminates a part of possible noise, notably the isolated pixels. It works as follows. Let *image1.minI* and *image1.Ec* respectively be the minimum intensity and the standard deviation of the intensities of *image1*. Let *val* equal *image1.minI* + *image1.Ec*/2. For each pixel (i, j) of *image1*, if it exists a circle of center (i, j) [or passing through the point (i, j)]

and radius r , r between $rmin$ and $rmax$, such as all of the circle pixel intensities are greater than val , then $image2(i, j)=image1(i, j)$ else $image2(i, j)=0$.

The third filtering pass erodes the pixels that are too close to the background level, using a statistical threshold. The principle is to set to zero each pixel that has in its neighborhood of size 1 at least six pixels of intensities less than the background intensity. Finally, the set of the detected pixels is decomposed on particles (i.e., sets of related pixels). Each particle is approximated by a disk, allowing one to model the corresponding spot by its center and its radius.

The first algorithm, because it is adapted only for spots following a 2-D model, is sensitive to spot signal noise, for example, when the spot pixels are weaker in the center (spots with a hole). The second algorithm is more robust to noise and to filter deformations. Its filtering phase has a smoothing effect that decreases noise sensitivity.

Grid Position and Orientation

The algorithm *G1* is based on the position of the spots initially obtained. The algorithm *G2* is based only on the image signals. Both of them compute grid direction, origin, and steps independently on the x and y axis. They use the same procedure for initializing the grid, but two different procedures, *PG1* and *PG2*, for positioning it. The initialization procedure calculates the origin (ox , oy), the angle (ax , ay), and the step (px , py) of the grid. These parameters are then used by the grid positioning procedures.

Initialization procedure. To calculate the origin, the angle, and the step of the grid, we make vary a quantity a corresponding to an angle, between $a-5$ degrees and $a+5$ degrees. Then, for each obtained angle, we do the following. We calculate the image profile in the directions a and $a+90$ degrees (i.e., the curve obtained by the sum of the pixel intensities in that directions). We extract the local maximum from the profiles. We calculate the cumulated energy from these maximums to obtain the criteria $K(ax)$ and $K(ay)$, where $K(ax)$ depends on the profile in the direction ax and $K(ay)$ depends on the

profile in the direction ay . The angles ax and ay are the ones having the maximum criteria $K(ax)$, $K(ay)$ respectively. The origin ox (respectively oy) and the step px (respectively py) are then obtained from the positions of the maximums in the corresponding profile.

Positioning procedure PG1. It consists to set the grids by minimizing a

criterion K_I that depends on the distance between detected spots and their theoretical positions on the grids:

$$K_I = \sum_{i=1}^B (X_i - (n_i p_{xx} + m_i p_{yx}))^2 + (Y_i - (n_i p_{xy} + m_i p_{yy}))^2$$

where (X_i, Y_i) is the image position of the spot i , (n_i, m_i) is the grid position of

Figure 1. Example of the quality of image analysis. (A) Software results on two examples of images of size equal to 1536 spots. Left: spot location on a nice image with many lighted spots. The total time of the analysis (grid setting, spot location, and spot quantitation) is 15 s with the first spot detection algorithm and 20 s with the second one. Right: spot location on a bad image with few lighted spots and many dirty marks. The total time of the analysis is 21 s. (B) Localization and quantitation of an irregular spot. Left: spot localization. Right: spot quantitation. (C) Software results on images of commercial arrays with particularly geometries. On the left, image of an Atlas 1.2 filter, having 1221 spots. Its total analysis time is 21 s. On the right, an image of a Coli filter, having 3072 spots. Its total analysis time is 25 s. Bottom: results on an image corresponding to a microarray on a glass slide obtained from <http://www.bioinformatics.duke.edu/CAMDA> and corresponding to Reference 17. It has 1936 spots, and its total analysis time is 81 s.

the spot i , $(p_{xx} p_{xy})$ is the direction vector of the lines, $(p_{yx} p_{yy})$ is the direction vector of the columns, and B is the number of spots.

The algorithm works as follows. It calculates a tolerance threshold t in function of the horizontal and the vertical steps, respectively px and py : $t = \min(px/3, py/3)$. It selects from the set of the spots initially detected on the image the ones that are in a distance less than t to the grid nodes. From the selected spots, it calculates the final origin, step, and orientation of the grid.

Positioning procedure PG2. It sets the grids by minimizing (in the case of black spots on white back) or maximizing (in the case of white spots on black back) a criterion K_2 that depends only on the signals in the image:

$$K_2 = \sum_{i=1}^L \sum_{j=1}^C I(n_i p_{xx} + m_j p_{yx}, n_i p_{xy} + m_j p_{yy})$$

where $I(X, Y)$ is the intensity at the position (X, Y) , (n_i, m_j) is the grid position of

the spot i , $(p_{xx} p_{xy})$ is the direction vector of the lines, $(p_{yx} p_{yy})$ is the direction vector of the columns, L is the number of lines, and C is the number of columns.

The algorithm works as follows. It does variations on the grid origin (Ox, Oy) around the initial position (ox, oy) . For each origin, it does variations on the grid step (Px, Py) around the initial position (px, py) ; for each step, it does variations of the grid angle (Ax, Ay) around the initial position (ax, ay) . For each origin (Ox, Oy) , step (Px, Py) , and angle (Ax, Ay) , it calculates a criterion K_2 obtained by the sum of the intensities of the pixels on the grid nodes. For K_2 maximum, we obtain the final positioning of the grid.

The algorithm $G1$ gives an analytical solution using a standard mathematical minimization method. It is appropriate for filters that contain enough spots to completely determine the position of the grids, at least $3 \times d$ spots, where d is the density (Figure 1A, left). $G2$ gives exhaustive solution using im-

age profiles and grid values in several directions. It uses all the image signals (spots and background levels), which makes it suitable for filters with few spots (Figure 1A, right).

Spot Quantitation

Once the spots are detected and referenced, a representative value is affected to them to quantitate the hybridization signals. For each spot, the following variables are calculated: (i) the mean intensity I_{mean} (average of all pixel intensities of the spot); (ii) the sum of intensities I_s (sum of all pixel intensities of the spot); (iii) the maximum intensity I_m (maximum pixel intensity in the spot); (iv) the weighted mean intensity I_p (sum of all pixel intensities weighted with the inverse distance to the center of the spot); and (v) the radius, the area, and the background level.

To evaluate each type of quantitation (I_m , I_{mean} , I_s , and I_p), a series of experiments were performed. We used

Figure 2. Comparison between the different quantitation modes: maximum intensity (I_m), mean intensity (I_{mean}), sum of the intensities (I_s), and weighted mean intensity (I_p). Using a logarithmic representation, the figure illustrates the variation of the average and the standard deviation of the hybridization signal intensities obtained for each concentration of ϕX DNA, in the four types of quantitation modes.

control filters spotted with 10 concentrations of phiX DNA (50 to 0.1 ng/μL) as described above. For each type of quantitation parameter, we calculated the average of the hybridization signal intensities obtained for each concentration. The variations of these averages in relation to the concentrations are represented in Figure 2. The more linear representation is obtained when the *I_{mean}* quantitation parameter is used.

The main purpose of cDNA array analysis is to quantitate spots to compare their values with reference values or with other spot values. Since filters are not perfect in their geometry and subject to noise and artifacts, and since the process is not always the same (various exposure time, various quantities of material, etc.), the intensity level of the hybridization filters are subject to large variations between filters and also within individual filters. So, it is necessary to have a good estimation of the background level intensity. The subtraction of the background allows one to eliminate the false-positive spots. A local background is computed for each spot. It is calculated as the average of the six minimum pixel intensities found around the spot. Then, all the intensity values of a spot (*I_{mean}*, *I_s*, *I_m*, and *I_p*) can be calculated with or without the local background level of the spot.

REPRODUCIBILITY OF THE ANALYSIS

To test the reproducibility of the hybridization intensity analysis, we used the control filters spotted with phiX DNA described above.

Reproducibility within Filters

To test the reproducibility within filters of our analysis system, we examined the variation of the signal intensities for each phiX concentration. The average of the hybridization intensity values obtained and the standard deviation were calculated (Figure 2). The deviation is more important for the low concentrations of DNA when the hybridization signal is close to the background. This deterioration for low concentrations illustrates the fact that spot detection is less accurate for weak intensities.

Reproducibility between Filters: Normalization Step

To test reproducibility between filters, we examined the results obtained from the hybridization of three filters (F1, F2, and F3) with the same probe. Figure 3A shows a representation of the variation of the signal intensities in relation with the concentrations. The signal intensities observed for the filters F1 and F2 are very close to each other and are greater than those observed for the filter F3. This is due to the differences in exposure time on the PhosphorImager. Similar differences

can occur because of variation in the efficiency of probe labeling. To permit comparisons between results obtained from multiple filters, the hybridization intensity for each spot was normalized by dividing it by the mean intensity of all spots present on the filter. Figure 3B shows the relation between the normalized signal intensities and the spotted cDNA concentrations for the three filters. The obtained curves become quite similar, illustrating the importance of the normalization process and the high level of reproducibility of the method.

Figure 4 gives another demonstration of the reproducibility. The correla-

Figure 3. Reproducibility between filters: normalization process. (A) Illustration of the variation of the *I_{mean}* signal intensities in relation to the DNA concentrations for three filters: F1, F2, and F3. (B) Illustration of the variation of the signals intensities after normalization by the mean of signals intensities of all the spots on the filter considered.

BioComputing/BioInformatics>>>>:

tion coefficient was calculated for scatter plots of intensities for duplicate spots hybridized with a labeled probe on the same filter (Figure 4A) and between two identical filters in the same hybridization tube (Figure 4B). In each case, it was equal or greater than 0.96.

ROBUSTNESS OF THE ANALYSIS

To measure the success rate of the automatic analysis, we examined a total of 436 filters of density 16 or 25, hybridized with complex cDNA probes. Among the 436 filters, 424 filters examined have been analyzed automatically, without human intervention, giving a success rate of 97%. In these images, all spots have been well detected. In the case of the 12 remaining filters, the number of undetected spots varies. In the best case, just one signal was missed. In the worst case,

because of a bad setting of the grid, half of the signals could not be detected. For all these 12 filters, user intervention was needed to correct the results given by the automatic analysis. This intervention concerned essentially the location of the grid. The user can help the software to set the grid correctly in different ways, such as designating one or two reference spots. In some cases, when very few spots are not detected by the software, the user can locate them manually.

DISCUSSION

Our spot detection algorithms use a spot model (disk of given center and radius) that is sufficient to describe and compare most of spots (e.g., spots with regular form) (Figure 1A, left). In the case of spots having a rugged and a nonuniform shape (Figure 1A, right),

we refine our model by selecting, from the disk defining the spot, the pixels that actually correspond to the spot. We propose two methods: first, the “Gaussian” method selects the pixels having an intensity value included between $E - T1 * SD$ and $E + T2 * SD$, with E the mean of the pixel spot intensities, SD the standard deviation, and $T1$ and $T2$ the two thresholds given by the user. Second, the “statistic” method selects the pixels having an intensity value greater than the intensity values of the first $P1\%$ pixels of the spot and less than the intensity values of the last $P2\%$ pixels, where $P1$ and $P2$ are two percentages included between 0 and 100. In the case of using these two methods, the spot is quantitated by the mean or the median of the selected pixel intensities.

An illustration of using the Gaussian method on an irregular spot (spot with a hole) is given in the Figure 1B. The

pixels having intensities corresponding to the background intensities (in the center of the spot) are not considered in the calculation of the spot intensity. Similar results are obtained using the "statistic" method.

The use of a spot model, namely a disk of given center and radius, lends robustness to our spot detection algorithms. Indeed, when no model or a model with less constraint is used to describe the spots, the spot pixels detected will be affected highly by image noise. The comparison between spots becomes less reliable. For example, in Reference 2, spot detection is based on a method that uses an adaptive and an iterative thresholding. This method works well for high-quality images, where spot signals are well distinguished from the local background. However, for spots of weak intensity, it will be impossible to find a threshold to separate all the spot pixels from the background. Also, for spots very closed with high intensity, it will be impossible to find a threshold to separate two spots without truncating a part of them.

In Reference 5, spot detection is based on the Laplacian threshold method. No model for the spots is provided, which leads to unreliable spot detection.

In our grid position algorithms, the steps and the orientation of the grids are computed independently in x and y coordinates, which allow a high flexibility for our analysis system. In Reference 5, an automatic detection of the grid is not available. A manual intervention and a reference image (template) are necessary. This makes any treatment in batch and/or in high scale impractical. The system of Reference 2 is also limited. First, it needs a high number of spots because it is based on a statistical approach. It uses the Hough transform technique (4) to compute the grid direction and the histogram of the distances between spots to compute the grid steps.

The running time of the software depends on the image size and the number of spots in the image. For example, with a Pentium 400 MHz, for an image of size equal to 650×450 pixels and containing 2400 spots, the software needs 30–40 s to detect and quantitate all the spots. For an image containing 10 000 spots, it needs 1–2 min.

CONCLUSION

XdotsReader is proved to be efficient. The automatic analysis succeeds even when the filters are of low quality, due to deformations or spotting defaults. Indeed, the algorithms present features that allow one to compensate for such problems: the individual location of each grid, the independent processing of the steps and the angles in x and y axis, and the identification of spots on the grids with a large tolerance concerning the theoretical positions. Also, our system has the advantage to be independent of the hybridization and signal acquisition process. All the information required to analyze automatically can be defined beforehand in the parameter menu. Complex data analysis tools are integrated in the software, which proposes different ways to compute the background level, to normalize spot values, and to compare spots and filters.

Because of its ability to analyze ar-

rays with no *a priori* assumption of size and density, our system is suited to quantitate hybridization signals on macroarrays such as those used in this study, but also on microarrays on glass slides or oligonucleotide arrays on DNA chips.

Finally, *XdotsReader* has four main advantages compared to the existing systems. First, it is capable to analyze a large range of images. The user can choose the number of lines, the number of columns, the geometry of the blocs, the presence and the size of the spaces between the blocs, and the presence or the absence of spots at some positions. Our software can for example analyze images provided from the Genexpress platform: 12 lines, 16 columns, blocs of 5×5 and 4×4 , 1 space (Figure 1A); Atlas 1.2 images from BD Biosciences Clontech (Palo Alto, CA, USA): 32 lines, 48 columns, blocs of 1, with several empty positions (Figure 1C); Subtilis images from Eurogentec (Seraing, Belgium): 16 lines, 24 columns, bloc of

Figure 4. Distribution of hybridization intensity (*mean*) values. (A) Distribution between duplicated spots hybridized with a labeled probe on the same filter. *r* is the coefficient of correlation. (B) Distribution between two identical filters in the same hybridization tube.

5 × 5, 1 space, with several empty positions; Coli images from Eurogentec (Seraing, Belgium): 16 lines, 24 columns, blocs in staggered rows (Figure 1C). It can also analyze microarray images (1936 spots on a glass slide) obtained from <http://www.bioinformatics.duke.edu/CAMDA> (17) (Figure 1C). When less precise parameters are allowed (as is the case for most existing software), the solution for the user is to modify manually the initial grid, which is always boring and slow.

Second, *XDotsReader* is capable of detecting the spots using several methods. This allows one to adapt the analysis to a large range of images. The area representing the spots will be then centered on the spots but also adapted to the size of the spots. Other software set areas of fixed size on the spots in some cases centered on the spots. However, a poor relation between the area supposed to represent a spot and the real spot could carry quantitation problems.

Third, *XDotsReader* is capable of detecting the grids automatically. The positioning of the grids ensures the reproducibility of the analysis, to treat in batch the analyses, and therefore to get a gain in time, compared to manual positioning.

Fourth, *XDotsReader* proposes two quantitation methods. Therefore, the user can choose the most adequate method to his or her set of images, to provide a more precise and a more pertinent quantitation.

XdotsReader is available from the Software Library at www.BioTechniques.com or at <http://www.lami.univ-evry.fr/~tahi/XDotsReader>. It runs on SUN workstations under UNIX (SunOS or Solaris) and on PC under LINUX. No particular hardware is required, and the software is compatible with any other software. It supports the main standard image formats. Main text editor output formats are integrated, such as postscript, Excel®, gif, or jpeg.

ACKNOWLEDGMENTS

This work was supported by the CNRS, AFM, and by grants from the EU BIOMED2 program (EURO-IMAGE Consortium, BMH4-CT-97-2284) to C.A.

REFERENCES

1. Alizadeh, A.A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503-511.
2. Audic, S. and G. Zanetti. 1995. Automatic reading of hybridization filter images. *Comput. Appl. Biosci.* 11:489-495.
3. Auffray, C., G. Behar, F. Bois, C. Bouchier, C. Da Silva, M.D. Devignes, S. Duprat, R. Houlgatte, et al. 1995. IMAGE: molecular integration of the analysis of the human genome and its expression. *C. R. Acad. Sci. III* 318:263-272.
4. Gonzales, R. and P. Winz. 1987. *Digital Image Processing*, 2nd ed. Addison-Wesley Reading, MA.
5. Granjeaud, S., C. Nguyen, D. Rocha, R. Lutton, and B.R. Jordan. 1996. From hybridization image to numerical values: a practical, high throughput quantification system for high density filter hybridizations. *Genet. Anal.* 12:151-162.
6. Houlgatte, R., R. Mariage-Samson, S. Duprat, A. Tessier, S. Bentolila, B. Lamy, and C. Auffray. 1995. The Genexpress Index: a resource for gene discovery and the genic map of the human genome. *Genome Res.* 5:272-304.
7. Hughes, T.R., C.J. Roberts, H. Dai, A.R. Jones, M.R. Meyer, D. Slade, J. Burchard, S. Dow, et al. 2000. Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.* 25:333-337.
8. Lee, D. and T. Pavlidis. 1988. Edge detection through residual analysis. *IEEE Computer Vision*.
9. Monga, O. and R. Deriche. 1989. 3d edge detection using recursive filtering. Application to scanner images. *IEEE Conference on Computer Vision and Pattern Recognition*.
10. Nacht, M., A.T. Ferguson, W. Zhang, J.M. Petroziello, B.P. Cook, Y.H. Gao, S. Maguire, D. Riley, et al. 1999. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res.* 59:5464-5470.
11. Pavlidis, T. 1980. *Structural Pattern Recognition*. Springer-Verlag, New York.
12. Perou, C.M., T. Sorlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, C.A. Rees, J.R. Pollack, D.T. Ross, et al. 2000. Molecular portraits of human breast tumours. *Nature* 406:747-752.
13. Pietu, G., O. Alibert, V. Guichard, B. Lamy, F. Bois, E. Leroy, R. Mariage-Samson, R. Houlgatte, et al. 1996. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* 6:492-503.
14. Rosenfield, A. and A. Kark. 1982. *Digital Picture Processing*. Academic Press, New York.
15. Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.
16. Soares, M.B., M.F. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* 91:9228-9232.
17. Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9:3273-3297.
18. Torre, V. and T. Poggio. 1986. On edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* 8:147-163.

Received 18 July 2000; accepted 17 April 2002.

Address correspondence to:

Dr. Fariza Tah
Laboratoire LaMI - UMR 8042
CNRS Université d'Evry-Val-d'Essonne/
Genopole
523, place des Terrasses, 91000 Evry, France
e-mail: tahi@lami.univ-evry.fr