Digital Image Processing for Rapid Analysis of Differentially Expressed Transcripts on High-Density cDNA Arrays

BioTechniques 26:1174-1179 (June 1999)

Boris Adryan, Volker Carlguth and Hans Jochen Decker
Johannes Gutenberg-University, Mainz, Germany

ABSTRACT

Usage of filter arrays is becoming increasingly attractive for many research laboratories involved in determination of gene-expression profiles. However, analysis of numerous spots, representing genes or partial gene sequences (ESTs), is still tedious work involving the ordered analysis of vast amounts of numerical tabular data. We present a rapid and efficient method for the visual identification of differentially expressed targets on high-density cDNA filter arrays using standard laboratory equipment and standard software, which is available for free. The method we introduce provides an inexpensive alternative, and no changes in the experimental set up are required. Our results were verified by densitometric analyses performed with an established system.

INTRODUCTION

Recently, the use of cDNA (micro-) array technology has provided an effective means to analyze gene expression profiles in human and other organisms (3). With these methods, the quantification of numerous transcripts of known full-length gene sequences or of partially known sequences as expressed sequence tags (ESTs) is possible. Unfortunately, high cost still limits the widespread laboratory use of high-density arrays. Although the optimization of array construction has already led to less expensive arrays (4), the high costs of hardware and software for post-hybridization analysis still limits the availability to large institutes. Chen et al. (1) have described an approach for using colorimetric reactions combined with an economical computer system instead of using costly laser-induced fluorescence detection or radioisotope detection, which usually requires expensive phosphor-imaging equipment.

In an attempt to make the technology accessible to most laboratories, without adopting new protocols such as chromophore labeling, we used an approach based completely on digital image processing to allow analysis of autoradiographs from hybridized array filters. Direct and fast analysis of autoradiographs from cDNA arrays without digital image processing is a time-consuming and tedious task involving vast amounts of tabular data. These data have to be densitometrically obtained, ordered and compared to identify differentially expressed targets. Some commercial software systems provided with high-performance imaging systems have features to solve this problem rapidly; however, they can be quite expensive. Here, we introduce a method and software for the rapid identification of differentially expressed transcripts based on standard laboratory equipment and a self-developed software, which requires no changes in the experimental set up. The reproducibility of our approach has been verified by evaluation of several examples.

MATERIALS AND METHODS

First, mRNA was prepared from normal renal tissues and sporadic renal cell carcinomas (RCC). Atlas™ cDNA Filter Arrays were purchased from CLONTECH Laboratories (Palo Alto, CA, USA). These arrays comprise 588 cDNA elements ordered into different groups and spotted in duplicate onto a positively charged nylon membrane. Essentially, the procedures of [^{32}P]cDNA labeling, array hybridization and washing were performed according to the manufacturers' instructions. BioMax™ MS film and a BioMax TranScreen™ Intensifying Screen were used (Scientific Imaging Systems [Eastman Kodak], New Haven, CT, USA).

For comparison with our approach, the exposed autoradiographs were scanned at 100-µm pixel size on a Storm® Blot Imaging System (Molecular Dynamics, Sunnyvale, CA, USA). Grids composed of 7 columns and 14 rows were laid over the image generated and analyzed by ImageQuant® soft-

Table 1. Comparison of Differentially Expressed Genes Identified by Pseudo-Colors (see Figure 2)

a Up-Regulated in Tumor	Array Position	Normal Value[a]	Tumor Value[b]	N/T Ratio
1	A1d	31	138	0.22
2	B1b	60	205	0.29
3	E4e	112	147	0.76
4	E6i	n.d.[c]	61	--
5	F1b	6	93	0.06
6	F1c	1.6	22	0.07
7	F5i	n.d.[c]	24	--
b **Down-Regulated in Tumor**	**Array Position**	**Normal Value[a]**	**Tumor Value[b]**	**N/T Ratio**
8	C3c	199	38	5.24
9	C2n	387	20	19.53
10	E6b	442	81	5.45
11	F2i	187	49	3.82

These visually identified spots were quantitated using the commercially available software (ImageQuant). The tumor values were normalized to the average of three housekeeping genes (*ubiquitin*, *GAPDH* and *β-actin*). It appears that our image reconstructions allow for identification of expression differentials at about 30% (spot 3).

[a]N, reference.

[b]T, corrected.

[c]Not determined.

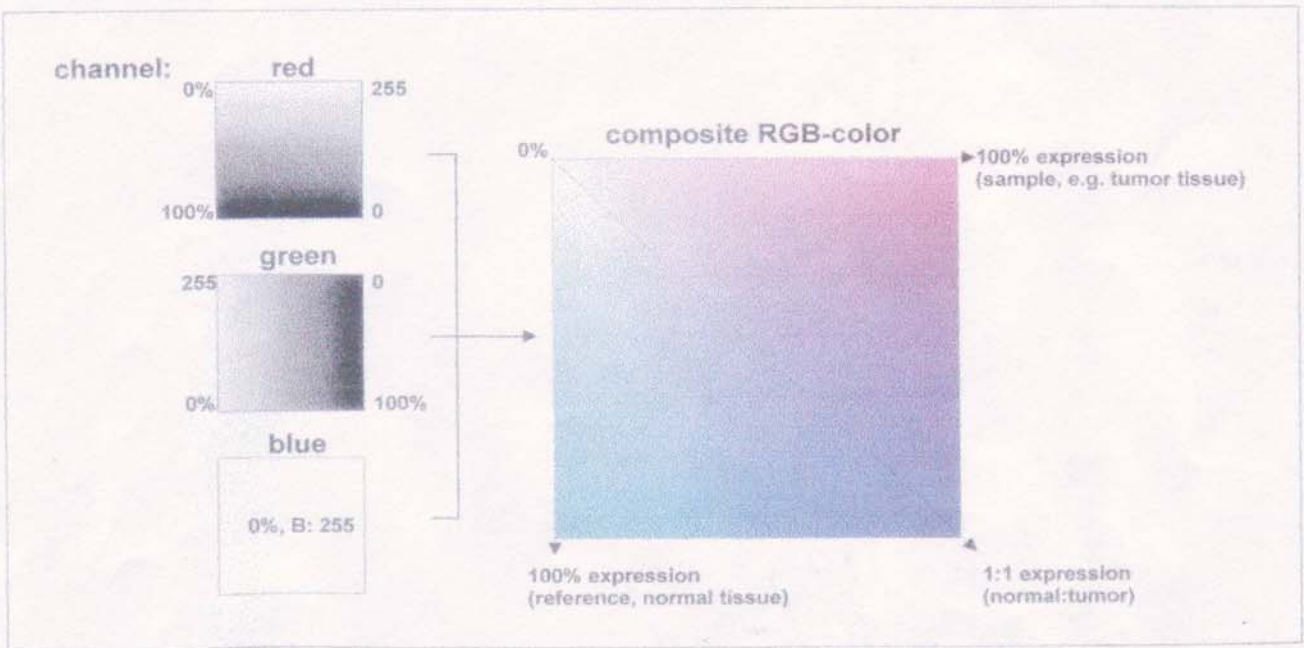

Figure 1. Pseudo-color transformation—principle of defining composite colors. The RGB system is the most popular method for defining a projected color, e.g., in computer systems. For example, pure red is defined by red = 255, green = 0 and blue = 0. Pure black has red, green and blue values of 0, and pure white has red, green and blue values of 255. The color vector shown here contains a non-variable background (blue = 255), thus creating a derived CMYK (cyan, magenta, yellow, black) vector. That was done for aesthetic reasons and is not cogently necessary.

ware (Molecular Dynamics) using the "local background" correction method. Intensity values of normal and tumor tissue for each gene were normalized using the averaged factors determined for three housekeeping genes (*ubiquitin*, glyceraldehyde-3-phosphate dehydrogenase [*GAPDH*] and *β-actin*).

Our method is based on typical laboratory computer equipment with a standard scanner. In our laboratory, we use the following: CPU: Pentium® 200, 64 MByte random-access memory (RAM), 4 Gbyte hard drive; scanner: Mustek® Paragon® 1200 SP; and operating system: Microsoft® Windows® NT 4.0. Both autoradiographs were scanned with moderate resolution (ca. 300 dpi), with at least 8-bit color depth (256 gray levels) and the same scanning options (e.g., contrast and brightness). To keep a linear scaling of gray values, the images were generated without gamma correction. Before scanning, the autoradiographs are aligned to allow exactly matched scanning. An accuracy down to 1 pixel is desirable, since the visual interpretation of the image might be compromised by less accurate alignment. In practice, however, this accuracy is very hard to achieve. We describe a procedure to align the filters on our Web site (see Availability). This procedure allows accuracy of 4 to 5 pixels, depending on the analysts' skills. This level of accuracy appears to be sufficient.

Before performing the pseudo-color composition, the following procedure was applied: *(i)* To equalize inconsistent exposure of x-ray films, the housekeeping levels of both the normal and the tumor-tissue blots were determined. Therefore, the signal levels of three housekeeping spots were averaged for each blot, considering their local background. *(ii)* The signal strength of the lower intensity spot was then normalized to the higher one. This facilitated producing comparative signal strength and avoiding the saturation of darker image areas. *(iii)* Background was removed by subtracting the optical density, calculated as an average of a defined blank area. *(iv)* For better visualization of the results, the gray values were scaled to cover the complete gray scale from 0–255. *(v)* The image composition was performed according to standard principles of digital image processing (2) using a pseudo-color transformation into the RGB system (Figure 1). In our program, these calculations and normalization are performed automatically.

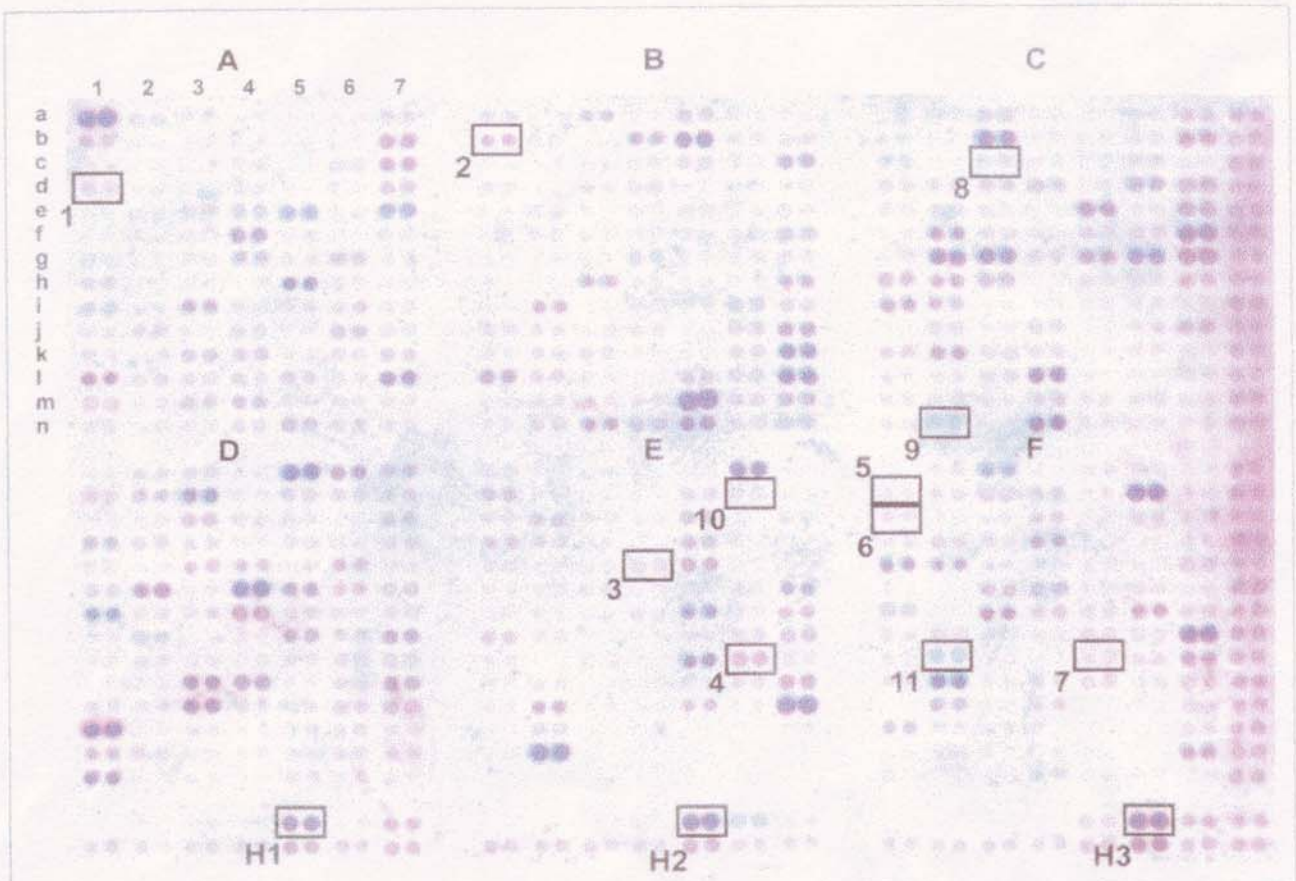

Figure 2. Merged pseudo-color image from digitized autoradiographs. Cyan component is attributable to the hybridization signature of reference-tissue mRNA (normal renal tissue), while the magenta component refers to the sample mRNA species (renal cell carcinoma), as determined by parallel array hybridization experiments. Numbered boxes (1–11) indicate differentially expressed targets identified by visual examination. H, housekeeping genes (H1, *ubiquitin*; H2, *GAPDH*; and H3, *β-actin*). DNA spots below the housekeeping genes as well as the rightmost double spots are genomic control DNA.

Table 2. Comparison of Our Averaged Densitometric Results with an Established Imaging System

	Clear Cell RCC		Chromophobe RCC	
	Our approach	ImageQuant	Our approach	ImageQuant
Up-regulation	6 (1.0%)	82 (13.9%)	39 (6.6%)	86 (14.6%)
Down-regulation	16 (2.7%)	68 (11.5%)	7 (1.1%)	62 (10.5%)
Not definable	0 (0.0%)	156 (26.5%)	(0.0%)	147 (25.0%)

Two subgroups (clear cell and chromophobe RCC) were investigated. 588 data sets (different genes) were evaluated for different gene expression in tumor tissue samples. Numbers are the differentially expressed spots identified, with their percentages to total spots in parentheses. Not definable data sets are those with "negative" gray values or ratios. Our approach provides lower detection level, but all spots were definable.

Different channels of the color system are assigned to the digitized and processed autoradiographs, creating a merged pseudo-color image (Figure 2). We have developed a Windows 95/NT-compatible software, which combines all image-processing steps into one application in which both visual interpretation and semiquantitative densitometric analysis of reference/sample blot pairs is possible. Densitometric data are obtained using the spots average gray value, corrected by the local background. This method appears to be less sensitive to noise and artifacts than using volume units.

RESULTS AND DISCUSSION

We examined the gene expression profiles of reference tissues (normal renal tissue) and samples (sporadic renal cell car-

cinoma and oncocytoma) using cDNA array hybridization of a high-density filter composed of 588 genes. Densitometric data of 16 autoradiographs (8 pairs normal/tumor: 5 clear cell RCC, 3 chromophobe RCC) produced by our application have been compared with the results obtained for three pairs by the established imaging system (Table 2). Differentially expressed transcripts were identified by the presence of pure or predominant colors. In the example shown (Figure 2 and Table 1, a and b, spot 3), a 30% up-regulation in expression levels could be identified visually and was subsequently confirmed by ImageQuant analysis. The semiquantitatively analyzed data appear to be reliable compared to the data analyzed by the established system. As shown in Table 2, higher sensitivity of the established system might not be advantageous with respect to its inherent sensitivity to background noise; this probably produces more false positives or a high percentage of "not definable" spots. However, the limitations of our approach underlie the common restrictions of autoradiography and 8-bit scanning. Image processing can reduce background noise; however, most often this reduction occurs at the cost of image information and signal strength. Therefore, only genes expressed at medium-high levels might be thus analyzed.

SUMMARY

The method presented aims to improve the identification of differentially expressed targets using a high-density cDNA filter array and standard laboratory equipment (scanner and software). The advantages of our effort are summarized as follows: *(i)* The Windows 95/NT software developed is ready–to–use for analysis of CLONTECH's Human cDNA Array and provides both visual data (merged pseudo-color image) and semiquantitative analysis. *(ii)* It can be applied to existing autoradiographs, and no change in the experimental set up is required. *(iii)* Our software offers an automatic background removal (local/general) and normalization against the housekeeping levels (reference/sample). *(iv)* Tabular data are transferred to a data table with position specification, gene name, expression level and ratio (reference/sample). *(v)* A link to GenBank® through the internet is implemented. *(vi)* All image-processing steps needed for the rapid analysis of cDNA expression filters are combined in one application and are executed automatically. *(vii)* In our hands, the whole procedure, starting from aligning autoradiographs, takes less than half an hour and should therefore offer a considerable advantage compared to other analysis techniques. *(viii)* The application is adaptable for other cDNA arrays.

AVAILABILITY

We offer this software as "freeware", and it can be downloaded from our Web site [uniform resource locator (URL): **http://www.uni-mainz.de/FB/Medizin/Tumorgenetik/array/index.htm**] (Note: the address is case-sensitive). Investigators interested in digital image processing can obtain a free copy of our source codes (Borland Delphi 3.0) for further development. This software is also available from the **Software Library** on the BioTechniques Web site (**http://www.BioTechniques.com**).

ACKNOWLEDGMENTS

We would like to thank Florian Konle for his valuable support in revising our source codes. We also thank Juergen Brieger and Tien Hsu for critical reading of the manuscript. Array hybridizations and ImageQuant analyses were performed by Ruediger Schwerdtle, whom we gratefully acknowledge. The work was supported by grants from Deutsche Forschungsgemeinschaft (DFG) (German Research Council De No. 356/3-3) and the Sander Foundation (No. 93038.2).

REFERENCES

1. **Chen, J.J., R. Wu, P.C. Yang, J.Y. Huang, Y.P. Sher, M.H. Han, W.C. Kao, P.J. Lee et al.** 1998. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics *51*:313-324.
2. **Gonzales, R.C. and P. Wintz.** 1987. Digital Image Processing. Addison-Wesley Publishing, MA.
3. **Schena, M., R.A. Heller, T.P. Theriault, K. Konrad, E. Lachenmeier and R.W. Davis.** 1998. Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol. *16*:301-306.
4. **Schummer, M., W.-I. Ng, P.S. Nelson, R.E. Bumgarner and L. Hood.** 1997. Inexpensive handheld device for the construction of high-density nucleic acid arrays. BioTechniques *23*:1087-1092.

Received 26 January 1999; accepted 15 April 1999.

Address correspondence to:
Boris Adryan
Department of Hematology/Oncology
Laboratory of Tumorgenetics and Molecular Cytogenetics
III. Medical Clinic of the Johannes Gutenberg-University
Obere Zahlbacher Strasse 63
55131 Mainz, Germany
Internet: adryan@mail.uni-mainz.de