ACTA STEREOL 1994; 13/1: 75-81
PROC 6ECS PRAGUE, 1993
ORIGINAL SCIENTIFIC PAPER

AUTOMATIC CLASSIFICATION OF CELLULAR ELEMENTS OF SOLID TUMORS; APPLICATION TO DNA QUANTITATION

Eric Masson [1,3], Paulette Herlin [2], Isabelle Galle [2], Françoise Duigou [2], Philippe Belhomme [1], Daniel Bloyet [1], Anne-Marie Mandard [2]

[1] LEI/ISMRA F. 14050 Caen Cedex France
[2] Pathology Department – F. Baclesse Center F. 14021 Caen France
[3] Informatic Department – F. Baclesse Center F. 14021 Caen France

ABSTRACT

Image Cytometry (ICM) leads often to more reliable DNA measurement of archival solid tumors than Flow Cytometry (FCM) due to sorting of unwanted nuclear elements (inflammatory and stromal cell nuclei, nuclear debris). Nevertheless, interactive sorting is time consuming and unable to give statistically valuable results in an acceptable delay for clinical oncology. The strategy, first results and output of an automatic cell classifier developed within the framework of the optimization of a DNA analyzer are here presented.

Key words: automatic cell classifier, DNA cytometry, image analyzer, solid tumors .

INTRODUCTION

Though Flow cytometry (FCM) is commonly used for rapid assessment of solid tumor DNA ploidy disturbances, a well known major drawback of this technic is its inability to discriminate debris and normal inflammatory or stromal cells from tumour cells (Friedlander, 1991). Image cytometry allows sorting of unwanted elements but the interactive procedure is very time consuming and unable consequently to deliver statistically valuable results in an acceptable delay for clinical oncology.
For several decades, numerous attempts have been made to provide cytometers with automatic cell selection for cytological prescreening or DNA measurement, nevertheless, as yet, few fully automatic efficient methods and systems are available (Cornelisse and van Driel-Kulker, 1985, Meyer and van Driel, 1986, van Driel-Kulker and Ploem-Zaaijer, 1989, Haroske et al.,1990, Palcic et al., 1992).
The challenge for a successful automatic cell classification is the ability to appraise the wide spectrum of morphological specific features of the various normal and atypical cell types at the various stages of their lifetime.

A flexible and adaptative tool is needed for a putative knowledge based training of the machine. The aim of this work is the description of a DNA dedicated cell classifier and the critical report of the first results obtained.

MATERIAL AND METHODS

1) Image cytometer

The cytometer is provided with a BH2 Olympus microscope, a moving stage, a Matrox PIP 1024 frame grabber and a Sony CCD camera. IOD measurements are done at a resolution of 512 x 512 in 8 bits (1px = 0.11 μm^2). Software is written in C language, runs under UNIX operating system and uses OSF motif graphic interface on 386 or 486 PC (Masson et al., 1992).

2) Biological material

DNA measurements are performed on archival human esophageal and breast cancer samples. Isolated nuclei are prepared by thick paraffin section disaggregation according to van Driel Kulker et

al. (1987)and stained according to Feulgen and Rossenbeck (1924) as previously described (Herlin et al., 1992).

3) Image treatment and cell sorting stages

DNA measurement of sorted cells is achieved as follows, using a six step procedure:

a- grey level image recording of Feulgen stained nuclei at 525 nm.
b- edge detection according to Deriche (1987a, 1987b) and suppression of non maximum points.
c- analysis and closing of connected pixel chains which define the object border-lines.
d- elimination of unclosed objects.
e- extraction of a set of features from each segmented object ie:
*integrated optical density (for final DNA analysis alone)
*thirty eight other parameters for cell sorting (11 form factors, 18 statistic parameters calculated on original and edge enhanced images and 9 textural parameters).The full list of parameters is given in appendix.
f- learning and automatic classification of objects.

This last step requires the collection of a meaningful data base through manual sorting by one or several experts, followed by an automatic classification according to this knowledge base.

4) Data base collection

A specific data base is built for every localization under study (esophagus, breast ...).
Each object is assigned to a predetermined fine group. As an example, up to eleven morphological groups have been identified for the study of human esophageal or breast carcinomas and their normal reference counterpart (fig.1).

- Five groups of cell nuclei to be analyzed: clear epithelial nuclei (1), dark epithelial nuclei (2), nuclei abnormal for their size (8), nuclei abnormal for their texture (9), nuclei abnormal for their size and texture (10).-
- Five groups of nuclei of unwanted cells: lymphocyte nuclei (3), plasma cell nuclei (4), granulocyte nuclei (5), fibroblast or muscle cell nuclei (6), unidentified nonepithelial nuclei (7) (such as nuclei of endothelial cells, mast cells, macrophages, adipocytes ...)
- One group of nuclear debris: nuclear clumps as well as cut nuclei (11).

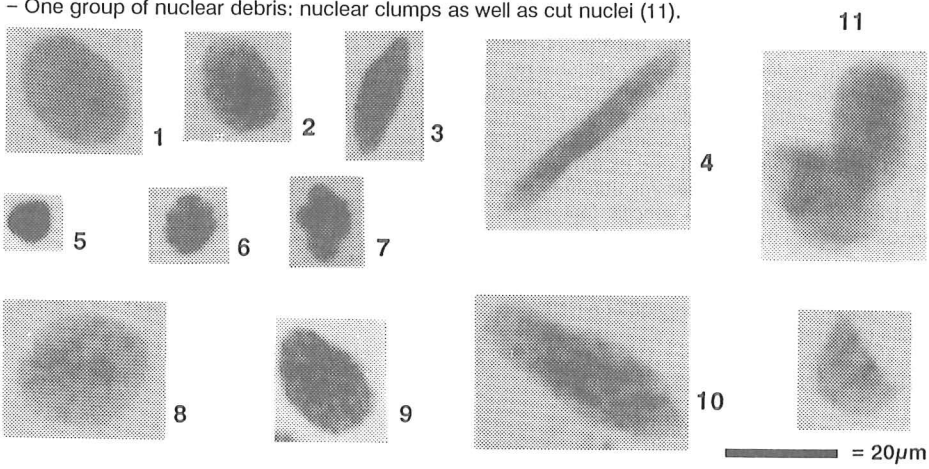

Fig.1: cell categories (human esophagus)

A large data base can be obtained when merging several data sets; a refined data base can be collected when crossing the labels delivered by several experts from the same set of objects.

5) Strategy of automatic classification

It is a three step procedure which goes as follows:

a– extraction of the model

Each group of objects is represented by an ellipsoid in its own 38 parameter reference space. Normal cells (groups 1 to 7) are taken only from normal reference tissue; debris are not learned (group 11).

Each ellipsoid is rescaled to get the same unity value mean squared distance.

b– sorting of an unknown object

The Euclidean distance (Ed) from the object to the center of each rescaled ellipsoid is calculated. If Ed≥2 the object is labelled as debris, if Ed<2 the object is assigned to the closest group (shortest Ed). Labelling is checked with respect to each parameter versus the learned limits of the group (only the upper limits of the abnormal groups 8, 9 and 10 are not taken into account); if the check fails, the object is assigned to the next closest group or labelled as debris in case of a second failure.

c– data analysis

A DNA ploidy histogram is built with respect to the whole cell population to be analyzed. 2c IOD value is determined from group 1 (clear epithelial nuclei from normal reference tissue).

Interactive and fully automatic global results are then compared. The reproducibility quality factor (Q_i) between two, cell by cell, sortings (interactive and / or automatic) is given by the following formulae: $Q_i = 100 \times N_1(i) \cap N_2(i) / 1/2 [N_1(i) + N_2(i)]$

Where $N_1(i)$ and $N_2(i)$ are the numbers of elements obtained for the i^{th} group and the symbol $\cap$ defines the "intersection" operator.

RESULTS

1) Relevance of the sorting strategy

Analysis of the distribution of Euclidean distances of every nucleus of any group in its own reference space (ie normalized ellipsoid) and in the reference space of other groups, gives evidence that the chosen 38 parameter system is able to isolate cells to be analysed (fig.2A, B) from unwanted cells (fig.2C). However, parameter values need to be checked to identify nuclear debris as illustrated by their wide-spread Ed values, in a well-preserved-cell reference space (fig.2D). From fig.2A and B, one must note that if the breakdown of the nuclei population in several fine groups is useful to improve identification of unwanted elements, nuclei to be analyzed overlap in their appearance. This observation justifies by its own a subsequent global analysis of fused groups of epithelial nuclei.

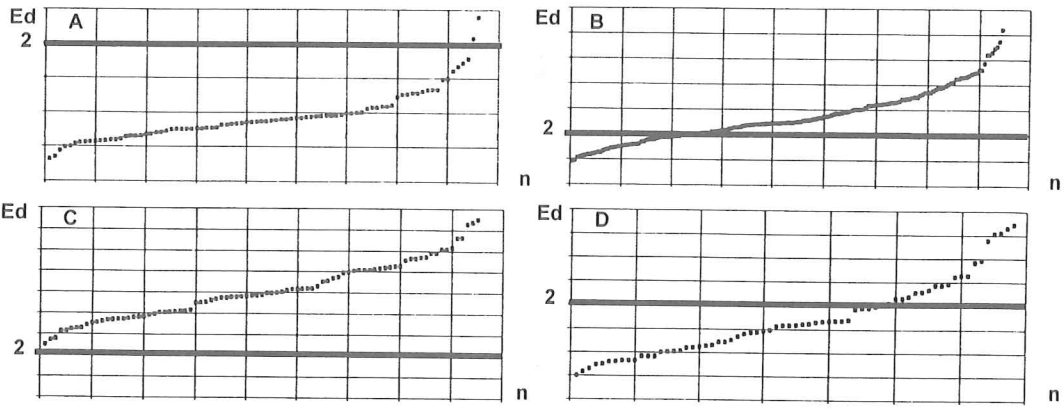

Fig.2: Data base obtained from a normal reference breast tissue, ranked distances of type 1 cells in their own reference space (A), distances of type 2 cells in the reference space of group n°1 (B), distances of type 3 to 7 cells in the reference space of group n°1 (C), distances of type 11 elements (debris) in the reference space of group n°1 (D).

2) Global analysis of the results obtained from fully automatic versus interactive DNA measurement

Fig.3 shows that unwanted elements (hypoploid debris in particular) are eliminated with obviously the same efficiency from automatic or manual sorting.

Informations dealing with the occurrence of aneuploid clones can be collected from an unknown cancer sample with the same efficiency from automatic or manual sorting, when compared to crude FCM and ICM analysis (fig.4).

As illustrated by fig.5, one is able to further purify DNA ploidy curves, by refinement of the characterization of normal looking nuclei.

Though ambiguous and difficult, fine subgrouping of morphologically abnormal nuclei offers promising tool for the identification of aneuploid clones (fig.6).

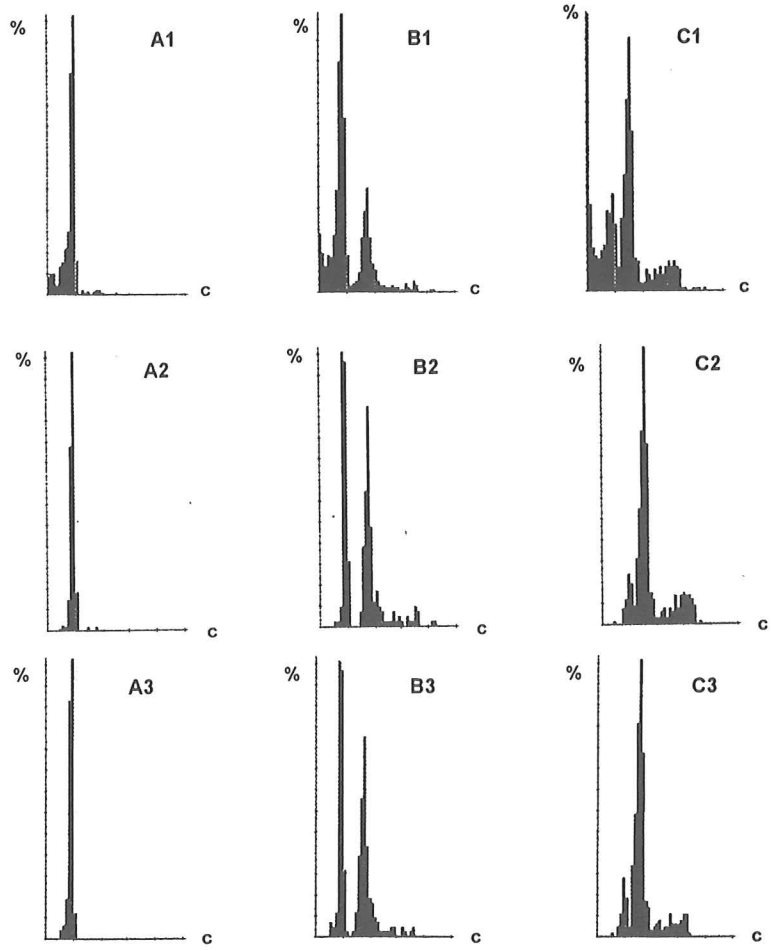

Fig.3: Histograms obtained from normal reference human breast tissue (A) and two high grade breast carcinomas (B,C), crude curves (1), manual sorting (2) and automatic sorting (3). Same images for the learning and test sets (2075 elements).

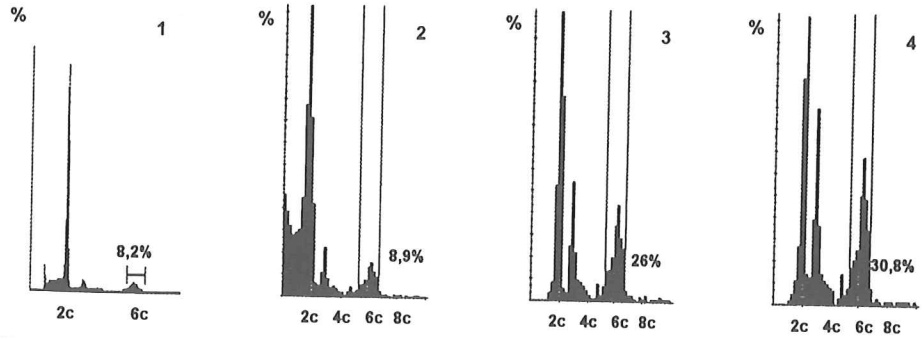

Fig.4: Histograms obtained from an unknown high grade breast carcinoma, crude curves (1: FCM, 2: ICM), manual sorting (3) and automatic sorting by reference with a different knowledge base (4). Learning set: 2075 elements, test set: 2828 elements (including reference sample).

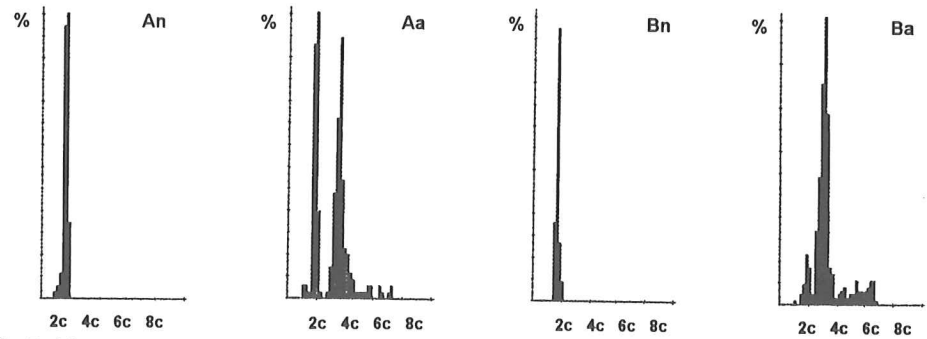

Fig.5: Histograms obtained from two high grade breast carcinomas (A,B), relationship between nuclear morphology and DNA ploidy (nuclei with normal features, group 1 to 7: n; abnormal nuclei, group 8 to 10: a). Same images for the learning and test sets (2075 elements)

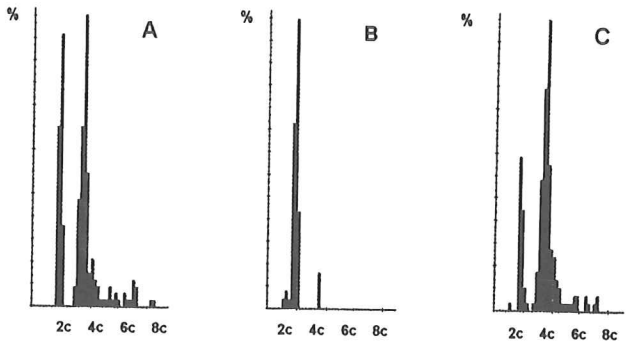

Fig.6: DNA ploidy of abnormal nuclei fine subgroups. Example of a high grade breast carcinoma: whole abnormal nuclear population (A), nuclei abnormal for their size only (B), nuclei abnormal for their texture or their size and texture (C). Same images for the learning and test sets (2075 elements).

3) Cell by cell analysis of interactive and fully automatic cell sorting

Cross-checking of interexpert and automatic-versus-expert labelling of a same large set of objects (5291 elements) shows a fairly good agreement (around 90%) obtained when eliminating unwanted cell nuclei from reference esophageal epithelium, as well as in situ carcinoma. The score of automatic sorting procedure is less satisfactory for analysable cell nuclei of cancer sample (around 75%) when compared to normal epithelium (around 90%) and needs to be improved for nuclear debris (around 70%). Reproducibility of refined group labelling is lower (under 50% for some groups).

4) Output

Up to 5000 sorted elements are required per case to collect 1000 nuclei of cells to be analyzed. At least 200 images per case must thus be recorded (25 to 30 elements per field at a magnification 125).

For the computer used, about 40 seconds are needed to record and process one image; so around 140 minutes are needed per case. A full time running system with an automatic multistage could analyze up to 10 cases per day (ie: one reference sample and nine cancers of the same localization).

CONCLUSIONS AND PERSPECTIVES

Automatic sorting of cell categories for a rapid and accurate ICM DNA measurement of solid tumors seems to be at hand .

Though highly promising, the here presented automatic classifier needs to be improved by adequate choice and enlargment of the knowledge bases and optimization of the number and quality of cellular groups, for every localization under study.

The introduction of weighted parameters and the use of mathematical morphology tools are able to improve debris as well as cell identifications.

Moreover, the chosen classification procedure must score off other classifiers (neural network procedure, decision tree processing...) tested on the same sets of images, work which is actually in progress (Elmoataz et al., 1992, Revenu et al., 1993).

ACKNOWLEDGEMENTS :

The authors are grateful to Mrs M. Michel and A. Baudron and to Miss E. Binet for their technical help.

This work was funded by grants from the "Agence Nationale pour la Valorisation de la Recherche", from the "Ligue Française contre le Cancer", from the "Association pour la Recherche sur le Cancer" and from the "Conseil Régional de Basse Normandie".

Appendix

11 form factors: area, perimeter, compactness, relative perimeter variation, mean symmetry difference, maximum symmetry difference, mean concavity ratio, maximum concavity ratio, extent of concavity, minimum angle between convex contour points, maximum lenght between convex contour points.

18 intensity factors: minimum extinction of the object, maximum extinction of the object, extinction range, skewness and Kurtosis of the extinction density within the object, 1st, 2nd, 3rd and 4th moments of the object for original and edge enhanced images.

9 textural parameters: 4 parameters according to Haralick (1979) for d = 1px at 0° and 90° and for d = $\sqrt{2}$ px at 45°and 135° (entropy, energy, contrast, inverse difference moment), "tex" parameter according to Masson (1992).

REFERENCES

Cornelisse CJ, van Driel-Kulker AMJ. DNA image cytometry on machine-selected breast cancer cells and a comparison between flow cytometry and scanning cytophotometry. Cytometry 1985; 6: 471-477.

Deriche R. Using Canny's criteria to derive a recursively implemented optimal edge detector. Intern J Comp Vision 1987; 167-187.

Deriche R. Optimal edge detector using recursive filtering. IEEE Transaction on Pattern Analysis and Machine Intelligence 1987; 9: 901-909.

van Driel-Kulker AMJ, Ploem-Zaaijer JJ. Image cytometry in automated cervical screening. Anal Cell Path 1989; 1: 63-77.

van Driel-Kulker AMJ, Mesker WE, van Der Burg MJM, Ploem JS. Preparation of cells from paraffin-embedded tissue for cytometry and cytomorphologic evaluation. Anal Quant Cytol Histol 1987; 9: 225-231.

Elmoataz A, Revenu M, Porquet C. Segmentation and classification of various types of cells in cytological images. Proceedings of the 4th Int conf on Image processing and its applications. Maastricht, april 1992; 385-388.

Feulgen R, Rossenbeck H. Mikroskopisch chemischer nachweis einer nukleinsäure von typus der thymonukleinsäure und die darauf beruhende elektive färbung von zellkernen in mikroskopischen präparaten. Z Physiol Chem 1924; 135: 203-224.

Friedlander M. Flow cytometric analysis in quantitative pathology. In : JPA Baak, ed. Quantitative Pathology in Cancer Diagnosis and Prognosis. Berlin Heidelberg New York: Springer Verlag, 1991; 244-268.

Haralick R. Statistical and structural approaches to texture. Proceedings of the IEEE 1979; 67: 786-804.

Haroske G, Bergander St, König R, Meyer W. Application of malignancy-associated changes of the cervical epithelium in a hierarchic classification concept. Anal Cell Path 1990; 2: 189-198.

Herlin P, Duigou F, Masson E, Bloyet D, Michels JJ, Mandard AMM. Image analysis optimization for the routine retrospective measurement of DNA ploidy of solid tumors. Part II. Acta Stereol 1992; 11/suppl 1: 411-417.

Masson E. Cytométrie automatisée par traitement numérique d'images : application à la caractérisation et à la classification de cellules épithéliales normales et tumorales. Thesis, Caen University , 1992.

Masson E, Herlin P, Bloyet D, Duigou F, Mandard AM. Image analysis optimization for the routine retrospective measurement of DNA ploidy of solid tumors. Part I. Acta Stereol 1992; 11/suppl 1: 405-410.

Meyer F, van Driel A. The application of mathematical morphology in cervical cancer screening. In: J Mary, JP Rigaut, eds. Quantitative Image Analysis in Cancer Cytology and Histology. Amsterdam: Elsevier SC Publ, 1986; 33-49.

Palcic B, Mac Aulay C, Shlien S, Treurniet W, Teczan H, Anderson G. Comparison of three different methods for automated classification of cervical cells. Anal Cell Path 1992; 4: 429-441.

Revenu M, Elmoataz A, Porquet C, Cardot H. An automatic system for the classification of cellular categories in cytological images. Proceeding of Intelligent Robot and Computer Vision XII: Algorithms and Technics, SPIE, Boston, sept 1993; 2055: 32-43.