Acta Oncologica Vol. 30 No. 8 1991

FROM THE DIVISION OF CLINICAL CYTOLOGY, DEPARTMENT OF PATHOLOGY, KAROLINSKA INSTITUTE AND HOSPITAL, STOCKHOLM, SWEDEN.

IMAGE ANALYSIS CYTOLOGY FOR DNA DETERMINATION IN BREAST AND PROSTATE CANCER

B. STENKVIST and I. JOHANSSON-BROWÉN

Abstract

Nuclear DNA distribution in fine-needle specimens from 112 breast carcinomas and 45 prostatic tumours was studied. The distributions were described statistically with five separate descriptors, namely mean deviation from 2C, percentage of cells exceeding 2.25C and 4.5C respectively, DNA index and entropy (a mathematical measure of degree of scatter of the DNA content of the nuclei). It was shown that information about DNA index only ('diploid' versus 'aneuploid') is too limited and the term 'diploid' even incorrect and misleading for description of human cancer cell populations. Merely an addition of the percentage of cells exceeding 2.25C allowed separation of a majority of the carcinoma specimens from specimens taken from normal tissue. Moreover, in carcinoma specimens 96% of the patients had between 1 and 100% of the cancer nuclei cell population exceeding 4.5C in contrast to all normal specimens, where this percentage was 0. Entropy was clearly correlated with percentage of Ki-67-positive cells, indicating that it contains information on both scatter of nuclear DNA content and proliferation. Plotting of the scatter (entropy) versus DNA index allowed separation of >97% of the carcinoma specimens from specimens taken from normal individuals. With the use of all five descriptors of DNA distribution it was possible to separate all breast cancer specimens and all moderately and poorly differentiated prostate cancer specimens from normal specimens. The discrimination between well-differentiated prostate cancer and hyperplasia constituted a special problem in that it was not possible to confirm the diagnosis well-differentiated prostate cancer by objective measurements in 8% of the patients, even when all five descripitors were used for the analysis. None of the verified carcinoma specimens had a DNA distribution that was truly 'diploid' in the sense of the distributions of cancer nuclei and normal cell nuclei from the actual organ being quite similar.

Key words: Breast cancer, prostate cancer, diagnosis, DNA analysis, statistics, terminology, image analysis cytology.

About 200 years after the invention of the microscope, it was observed that the occurrence of abnormal mitoses and cells in some tumours could have prognostic implications (1). Several visual grading systems were later suggested (2–6). Almost 20 years after the initial studies of cell growth and cell function (7) it was observed that tumours with a highly abnormal nuclear DNA distribution seemed to have a worse prognosis than those with a less scattered distribution (8–11). The measurement of DNA content was originally done with complicated and slow microspectrophotometers, measuring one nucleus at a time. Consequently this methodology was used at only a few research centers (12).

The development of image analysis technology on fine-needle or imprint specimens (IAC) from human cancer cell populations (13)—stoichiometrically stained—made it possible to collect the requested features from 50–100 nuclei per field of vision in a few seconds. This was accomplished because the image segmentation procedures that were developed could isolate the nuclei in fractions of a second. It also allows the microscopist to interact with the final data storage in the computer memory by pointing at artifacts, double nuclei, etc. and to use an algorithm that excludes objects that the observer does not approve (Fig. 1). Such technology allows fast measurements on fine-needle biopsies under visual control and is now proliferating in many clinical laboratories.

Flow cytometry, where a suspension of cell nuclei is prepared and the nuclei are stained with DNA-specific stains, has also been widely used (14, 15). Large amounts of cells in suspensions can be analysed with this technique, but at the stage of preparation it is difficult to achieve perfectly monocellular specimens, which is a prerequisite

Paper presented at Symposium on Flow and Static Cytometry in Umeå, 10–11 September, 1990.

Accepted for publication 20 April 1991.

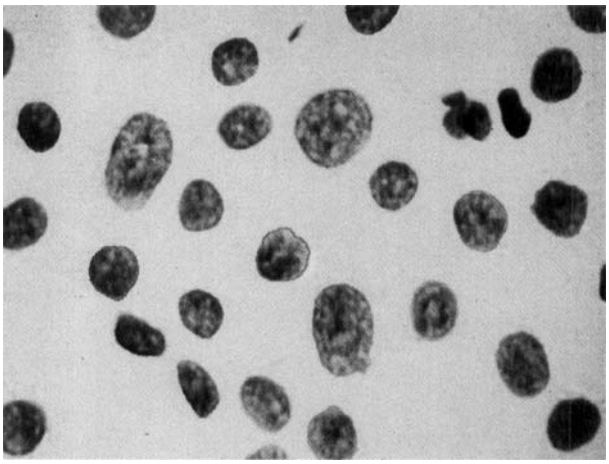

Fig. 1. A digitized image of one field of vision of a Feulgen stained specimen for image analysis cytology. As this image shows, the digital version of the specimen is fully comparable with what can be seen in the microscope, so that the correct type of cells can be chosen for inclusion in the histogram of DNA distribution.

for reproducible and reliable results. In addition, the generated histograms of DNA distributions are difficult to interpret since no visual interaction is possible during the measurement procedure.

The present report deals with image analysis cytology of nuclear DNA in Feulgen-stained fine-needle biopsies from 112 breast carcinomas and 45 clinically diagnosed prostatic enlargements. The study is concentrated on the issue of whether it is possible to identify a distribution of cancer nuclei, which truly cannot be separated from a distribution of normal cell nuclei from the same organ.

Material and Methods

Fine-needle aspiration biopsy procedure. For the prostate, fine-needle biopsy according to the Franzén procedure was employed (16). The syringe holder is designed to fit disposable syringes and to be operated by one hand. The curved needle guide has a ringlike holder at the distal end for insertion of the index finger. The needle is a 22-gauge 20 cm long, flexible, non-disposable needle. Gloves and finger cots are required. Local anaesthesia or bowel preparation was not used. A careful digital examination of the prostate identified locations for biopsy. The needle guide was then mounted on the freshly gloved left index finger (for right-handed persons). The finger cot was used to cover the distal part of the guide. This prevented faecal material from entering the needle when it was introduced into the rectum. In addition the cot will help to keep the guide in position. The target was located with the index finger. The needle, attached to the syringe, was then inserted into the guide and the needle tip was pushed into the target. Full suction was applied and the needle was quickly moved back and forth several times. The negative pressure was released with the needle tip still in the target. After withdrawal from the prostate, the needle was detached from the syringe and the aspirate was pressed out on slides. Smears were then prepared and air-dried or fixed in ethanol. One of the slides was fixed in 4% phosphate-buffered formaldehyde for at least 30 min. This slide was stained according to the Feulgen procedure—1 h hydrolysis in 5 mol/l HCl at 18°C; basic fuchsin from Aldrich Chemie was used.

The fine-needle biopsy of breast carcinomas needed no metallic needle guide. The lesion was held steady with the free hand and the fine-needle aspiration was performed as described above and the smears were air-dried, then fixed and stained by the Feulgen method.

Controls. The determination of the 2C value was accomplished by measuring 30–50 granulocytes or lymphocytes as internal controls from the slide to be measured. Twelve healthy women between 30 and 35 years of age volunteered for fine-needle biopsy of their breasts. In the prostate group 10 patients with hyperplasias were used as reference.

Immunocytochemistry. Monoclonal antibodies were used for immunocytochemical assays. The growth fractions were evaluated with Ki-67 antibody supplied from Dakopatts Inc. Bound monoclonal antibodies were detected with peroxidase-antiperoxidase complex, with avidin biotin glucose oxidase complexes.

Measuring instrumentation and procedure. We have developed specialized application programs for medical microscopy so that our system can execute procedures for DNA analysis (densitometry) at high speed, and analysis of shape and some texture features. The microscopy analysis system presented here is a development of our previously described systems for image microscopy (13, 17, 18) according to ideas previously presented (19). It is not commercially available.

The image analysis system for the microscope includes processing power, input/output channels and user interface as three separate units both in the hardware and in the software. It consists of a high-speed dual ported memory, a special image processor and a display processor with a local display memory. These units are linked together with dedicated high-speed data buses (14 MB/s). In addition, each unit is connected to a standard VME bus, which is used for the loading of programs, special data and control information. A conventional computer is connected to the VME bus as well as disc units and peripherals.

The intended application of the system has been interactive image analysis cytology ever since our first system was designed for such applications in cancer research (13) and for automatized screening of cervical smears (20). This puts strong requirements on high-speed image data flow into the processing system at rates of about 5–10 Mpixels/s sustained for several minutes. Therefore the image memory has two different ports. Through one of these ports data can

be continuously fed into the memory at rates up to 10 Mpixels/s at the same time as data are being read from the other port at rates up to 14 Mpixels/s. Having finished the segmentation algorithm processing and stored away the coordinates of the alarmed objects, the processor resumes its high-speed and analysing operation. The image memory serves as a large FIFO between the data source and the processor allowing asynchronous but still tightly coupled operators. The user interface interacts with the system via a high resolution (1 024 × 1 280) colour display with a non-interlaced refresh rate of 60 Hz. Both the image refresh memory and the graphics memory can be loaded from the image processor. The image memory can accept data at 14 Mpixel/s. The user can interact with displayed images and graphic entities by using a digitizer puck to control a programmable cursor or two cross-hair cursors. The software consists of the microcode for the image processor and the system software in the internal Motorola 68010 host computer.

The cells measured through a 537 nm interference filter and the image was digitized with a CCD camera (Panasonic CCV, model V 56/6) after shading correction. The measurement was made so that the cells selected by the segmentation procedure could be approved by the observer. The microscopic field of vision could be seen on a TV monitor and a decision to digitize the image could be made by the observer, who could then on another monitor observe the segmentation result and select or deselect objects of interest. This was programmed so that selection/deselection of objects could be done by pointing at them with an electronic arrow on the display or at their respective position on a histogram over the object's density ('DNA') at the side of the display (Fig. 1).

Recorded features. As mentioned in the introduction, the measurement procedure gathered DNA values of cells as illustrated by Fig. 1. The procedure consisted in the measurement of 400–1 000 cells of the tumour cell population from each fine-needle biopsy as defined by an experienced cytopathologist.

The second procedure consisted in the production of the following statistics over the cell population:

Mean deviation from 2c was calculated as previously described (21) and according to the formula

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(c_i - 2c)^2}$$

where N = number of cells in the selected interval, c_i = DNA value for the ith cell, 2c = 2C value.

Percentage of cells exceeding 2.25c and percentage of cells exceeding 4.5c were derived from the computer.

DNA-index was calculated according to the formula

$$DI = \frac{\frac{1}{N}\sum_{i=1}^{N} c_i}{2c}$$

Depending on the presence of one or two discernible peaks in the histogram, one or two DNA-indices were calculated according to the above formula. The channel limitation of each peak was determined interactively on the electronic display by the observer. In the study presented here, DNA-index means the index generated by the position of the first peak only.

Entropy (22), which is a concept from mathematical information theory was calculated according to the formula

$$\text{Entropy} = -\sum_{i=1}^{N} p^i \times \log 2(p^i)$$

The formula describes the information content in the histogram and the entropy attains the value 0 if all measurement values of density (DNA in the present situation) become located in exactly the same channel or $\log 2(80) = 6.33$ if the measurement values were evenly distributed over all 80 (N = 80) channels, which is the number in the situation presented here.

Results

Breast carcinoma. In 23% of the specimens (27/112) there was a peak in the histogram describing the distribution of nuclear DNA within 2C + 12.5% (corresponding to a DNA index of 1.00 + 0.125). An example of such a distribution is shown in Fig. 2. As can be seen in this figure the majority of the nuclei are in the 2C region. However, the mean deviation from 2C is 1.18. In addition, the scatter of measurement values is reflected by the entropy value of 3.49, which is higher than we have been able to demonstrate in any of the normal control populations. There are 38.6% of the nuclei that have a value over 2.25C and 7.7% above 4.5C in this particular specimen—values that we did not find in any of the normal control nuclei (see below).

Fig. 3 illustrates that in all breast carcinomas at least 6% of their nuclei exceeded 2.25C and Fig. 4 illustrates the fact that 96% of the breast carcinomas had nuclei exceeding 4.5C with from 1 to 100% of the measured nuclear DNA. We did not obtain such values in any of the controls.

If scatter (entropy) of the measured nuclear DNA values was considered in addition to the DNA index in 2-dimensional plots, a situation was obtained as illustrated by Fig. 5. There was an overlap of only three specimens out of 112 breast carcinomas (2.7%) into the same region as the normal breast nuclei. If the other descriptors were added —namely percentage of nuclei exceeding 2.25 and 4.5C respectively and mean deviation from 2C—in a 5-dimensional analysis none of the breast carcinoma specimens fell within the frame of the measurement values of nuclei from normal breast tissue.

From the Table—a correlation matrix over the statistical descriptors of the DNA measurement values—the descriptors clearly covariate, but still 2-dimensional plots

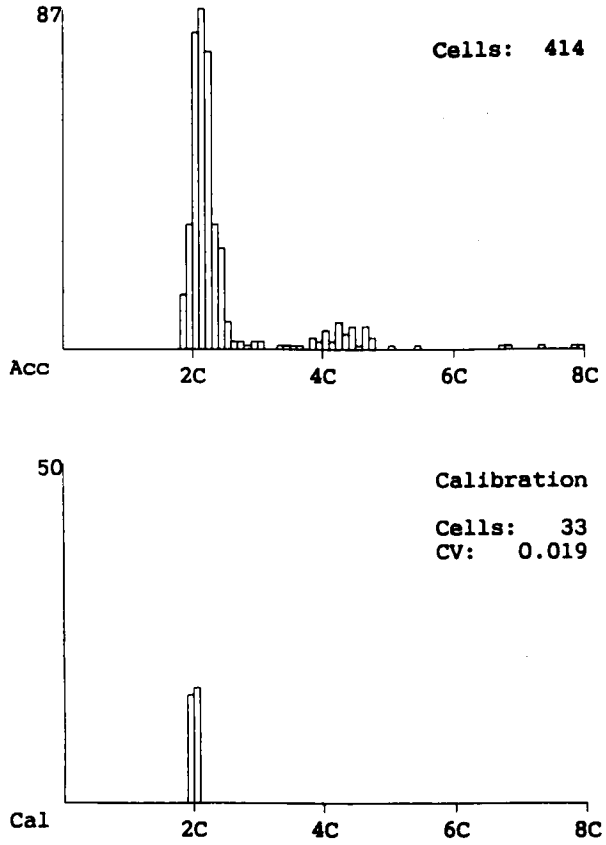

Fig. 2. A histogram of a tumour cell population (upper) and a control population of lymphocytes and granulocytes from the measured slide (lower). The mean deviation from 2C is 1.18. The DNA index is 1.09 but still clearly different from a normal cell population. 38.6% of the cancer cell nuclei exceed 2.25C and 7.7% exceed 4.5C and the histogram entropy is 3.49.

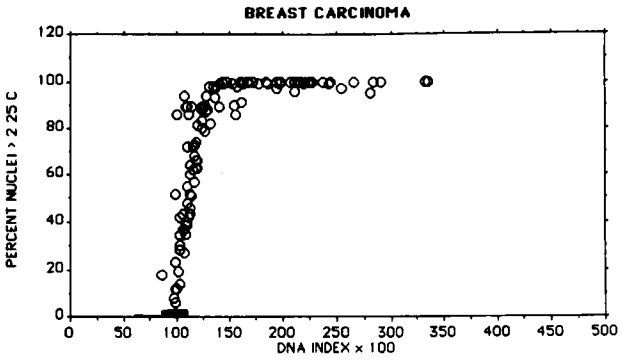

Fig. 3. A scatter plot showing that in all the breast cancer cell populations at least 6% of the cells exceeded 2.25C, whereas this was not the case in any of the cell populations from normal breast tissue (illustrated by the black rectangle at the bottom over DNA index 1.00).

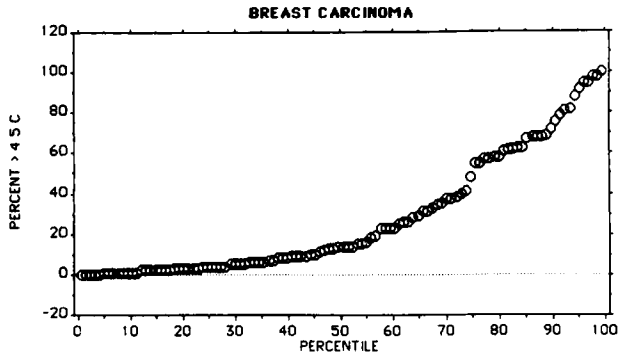

Fig. 4. A scattergram of the percentile distribution of breast cancer cells exceeding 4.5C. Only 5 of 112 breast cancer cell population had 0% > 4.5C, whereas 107/112 (96%) had from 1 to 100% of the tumour cells exceeding 4.5C.

as Fig. 5 contribute substantially to the separation of benign from malignant specimens.

There was also a correlation between the descriptor entropy and the number of Ki-67-positive cells as illustrated by Fig. 6. The correlation coefficient was 0.78.

Prostate. One of the problems in the diagnosis of prostate nodules is to find the demarcation line between atypical hyperplasia and highly differentiated carcinoma. The problem is shown in Fig. 7, where there is an overlap of 36% (5/14) when a scatterplot of hyperplasia—with entropy on the y-axis and DNA index on the x-axis is superimposed on a scatterplot of well-differentiated prostatic carcinoma. Thus it was not possible to separate well-differentiated prostate cancer from hyperplasia by DNA index and entropy only. A follow-up of the patients (and further patients not included in the present study) is ongoing to explore the biological significance of this finding. Fig. 8 illustrates that only one overlap between hyperplasia and moderately/poorly differentiated prostate cancer was encountered, which is 4.8% (1/21). However, if the

Table

Correlation matrix over descriptors of nuclear DNA distribution in breast carcinomas

Percentage of nuclei > 2.25C	1					
Percentage of nuclei > 4.5C	0.638	1				
DNA index (peak 1)	0.66	0.857	1			
DNA index (peak 2)	0.649	0.655	0.742	1		
Mean deviation from 2C	0.713	0.927	0.848	0.737	1	
Entropy	0.728	0.802	0.665	0.66	0.887	1

The descriptors are all correlated with each other at a significant level.

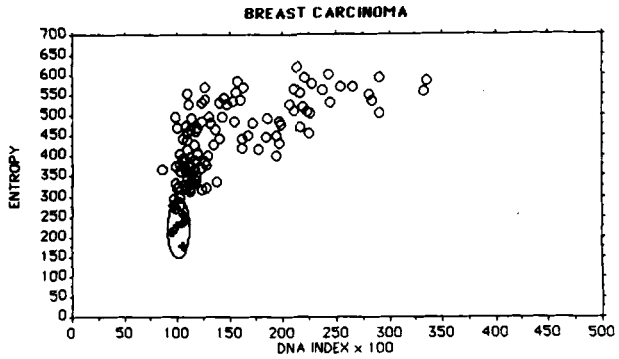

Fig. 5. A scatterplot of the DNA indices versus entropy for 112 breast carcinomas (rings) and normal breast tissue (plus signs inside the ellipse). Only three overlaps were noticed (= 2.7%).

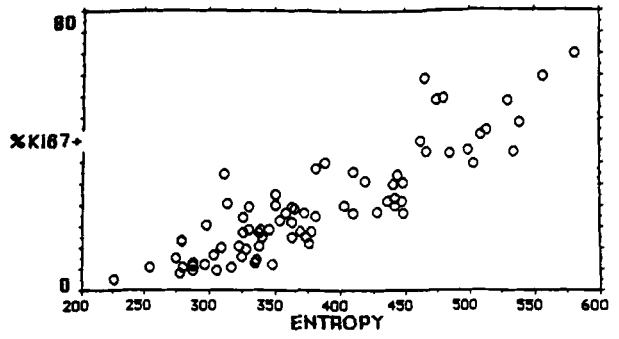

Fig. 6. Scatterplot illustrating the correlation (= 0.78) between entropy (= scatter) and percentage of breast carcinoma cells positive for Ki-67.

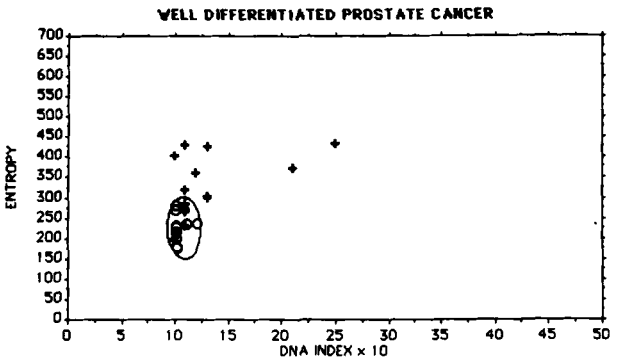

Fig. 7. Scatterplot (DNA index versus entropy) illustrating the overlap between prostatic hyperplasia (rings within the ellipse) and well-differentiated prostate carcinoma (plus signs). Of the well-differentiated prostate cancer 5/14 (36%) were within the ellipse which corresponded to values for the hyperplastic prostate cell populations.

other descriptors were included—namely percentage of nuclei with DNA over 2.25 and 4.5C respectively and mean deviation from 2C—there was a reduction of overlaps to 2/24 (8%) in the hyperplasia well-differentiated prostate cancer situation and no overlap in the hyperplasia—moderaterately/low differentiated prostate cancer situation.

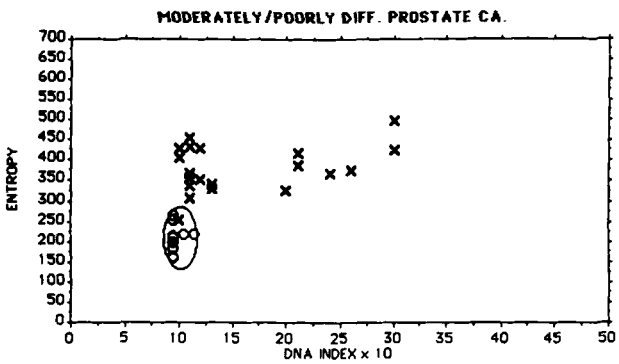

Fig. 8. Scatterplot (DNA index versus entropy) illustrating the sparse overlap between hyperplasia and moderately/poorly differentiated prostate cancer. Only 1/21 (4.8%) of the cancers was within the same range of values as hyperplasia.

Discussion

The terminology that describes the measurement results (often the information content in the histogram over the DNA per nucleus in the measured population) is getting increasingly confusing and it is time that something was done to restore order in that particular field. Not only basically different technologies (mostly IAC and flow cytometry), but also different terminologies seem to be used haphazardly. As a result, the scientific reports from the field are difficult or impossible to intercompare and understand since some use imprecise and profoundly unintelligible descriptors such as 'diploid, peri-diploid, almost diploid, near-diploid, euploid, aneuploid cell populations, etc.' or visual categorizations turned into 'types' of histograms. Other misleading descriptions are also used, such as determination of 'percentage cells in S-phase' by calculating—mostly from flow histograms—the percentage of cells between two peaks implying a cell physiological condition without using any specific marker of the cell population for the identification of truly proliferating cells, such as tritiated thymidine, or markers like bromodeoxyuridine or specific monoclonals against proliferating cell nuclear antigens, for instance Ki-67. These days, when amplified gene expressions can be exactly determined in fine-needle biopsies from patients (23) reports of 'near-diploid' cancers are unsatisfactory. Using a term like 'peri-diploid' is like dialling a long-distance telephone call with two digits wrong, ending up in China instead of England and claiming that it was 'almost right'.

The results obtained in the study presented here indicate that even simple statistical descriptors in 2-dimensional plots can separate the vast majority of breast and prostate cancer specimens from non-cancerous tissue. If five of the descriptors were used, all cancers could be separated from non-cancer except for well-differentiated prostate cancer that could not be clearly demarcated from hyperplasia/atypical hyperplasia—a problem which we have also recognized in a previous study (24).

In the present report it was found that entropy is correlated with Ki-67 positive cells. Since entropy is a measure of histogram scatter, this finding obviously has two explanations, namely that unstable cell populations with DNA values significantly over 2C are not only reflected by a high entropy value but also have an inbuilt coding for proliferation which makes unstable and scattered DNA values covariate with entropy and proliferation markers such as Ki-67 (and the flow cytometric 'S-phase'). Further studies are required to explore these complicated interrelations.

With the exception of the unclear border between well-differentiated cancer of the prostate and hyperplasia, the present study has shown that by using a combination of only two descriptors (entropy and DNA index) the vast majority of breast and prostate cancer specimens can be separated from normal. If all five descriptors are used in a 5-dimensional space, probably all the cancer specimens can be separated from normal. This is promising and increasingly important, in particular with reference to 'early detection' programs, which often reveal minute lesions of the breast and prostate of unclear significance.

The results in the present study also call for a serious discussion of the usefulness of DNA analyses as such in their clinical application. That is, how are DNA analysis results to be presented and interpreted, what methodology is appropriate, and what is the clinical significance? Our results indicate that IAC at least could be used in the future as a diagnostic supportive method. We have previously shown a diagnostic applicability of image analysis cytology to cervical specimens (25). The prognostic implications of the statistical descriptors presented here are now being analysed and will be the subject of a separate article.

ACKNOWLEDGEMENT

This work was supported by grants from the Cancer Society in Stockholm (Project No. 90:133).

Corresponding author: Dr B. Stenkvist, Division of Clinical Cytology, Department of Pathology, Karolinska Institute and Hospital, S-10401 Stockholm, Sweden.

REFERENCES

1. Hansemann D. Über assymetrische Zellteilung in Epithelkrebsen und deren biologische Bedeutung. Virchows Arch [A] 1890; 119: 299–326.
2. Broders AC. The microscopic grading of cancer. In 'Treatment of cancer and allied diseases'. 2nd ed. New York: P. Hoeber, 1940: 55–75.
3. Greenough RB. Varying degrees of malignancy in cancer of the breast. J Cancer Res 1925; 9: 453–63.
4. Patey DH, Scarff, RW. The position of histology in the prognosis of carcinoma of the breast. Lancet 1928; 1: 801–4.
5. Bloom HJG, Richardson WW. Histologic grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957; 11: 359–77.
6. Stenkvist B, Westman-Naeser S, Vegelius J, et al. Analysis of reproducibility of subjective grading systems for breast carcinoma. J Clin Pathol 1979; 32: 979–85.
7. Caspersson TO. Über den chemischen Aufbau der Strukturen des Cellkernes. Scand Arch Physiol 1936; (Suppl 73).
8. Sandritter W. Über den Nukeinsäuregehalt in malignen Geschwülsten. Naturwissenschaften 1952; 39: 46–7.
9. Atkin N. Modal deoxyribonucleic acid value and survival in carcinoma of the breast. Br Med J 1972; 86: 271–85.
10. Böhm N, Sandritter W. DNA in human tumours. A cytophotometric study. Curr Top Pathol 1975; 60: 152–219.
11. Caspersson TO, Auer G, Fallenius A, Kudynowski J. Cytochemical changes in the nucleus during tumour development. Histochem 1983; 15: 337–62.
12. Caspersson TO, Kudynowski J. Cytochemical instrumentation for cytopathological work. Int Exp Pathol 1980; 21: 1–54.
13. Stenkvist B, Westman-Naeser S, Holmquist J, et al: Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations. Cancer Res 1978; 38: 4688–97.
14. Moldavan A. Photo-electric technique for the counting of microscopical cells. Science 1934; 80: 188–9.
15. Headley D, Frielander ML, Taylor IW, et al. Application of DNA flow cytometry to paraffin-embedded archival material for the study of aneuploidy and its clinical significance. Cytometry 1985; 6: 327–33.
16. Franzén S, Giertz G, Zajicek J. Cytological diagnosis of prostatic tumours by transrectal aspiration biopsy: a preliminary report. Br J Urol 1960; 32: 193–6.
17. Bengtsson E, Eriksson O, Jarkrans T, Nordin B, Stenkvist B. CELLO—an interactive system for image analysis. In Bolc L, Kulpa Z, eds. Digital image processing systems. 1981; 109: 21–45.
18. Eriksson O, Bengtsson E, Jarkrans T, Nordin B, Stenkvist B. ILIAD—a high level dialogue system for picture analysis. In: Natural language question and answering systems for image analysis. München: Carl Hanser Verlag and London: MacMillan, 1982: 1–86.
19. Stenkvist B, Pressman N, Bartels P, et al. Biomedical pattern recognition and image analysis of cells and tissues. In: Dahlem Konferenzen Series 'Biomedical image analysis and pattern recognition'. 1979: 333–62.
20. Stenkvist B. Imaging microscopy with artificial intelligence for automatic analysis of papaniolaou smears. In: Clinical cytometry and histometry, 1987: 385–91.
21. Böcking A, Adler CP, Common HH, Hilgarth M, Granzen B, Auffermann W. Algorithm for a DNA-cytophotometric diagnosis and grading of malignancy. Anal Quant Cytol Histol 1984; 6: 1–7.
22. Stenkvist B, Strande G. Entropy as an algorithm for the statistical description of DNA cytometric data obtained by image analysis microscopy. Anal Cell Pathol 1990; 2: 159–65.
23. Lönn U, Lönn S, Nylén U, Winbald G, Stenkvist B. Amplification of oncogenes in mammary carcinoma shown by fine-needle biopsy. Cancer 1991; 67: 1396–400.
24. Stenkvist B, Olding-Stenkvist E. Cytological and DNA characteristics of hyperplasia inflammation and cancer of the prostate. Eur J Cancer Clin Oncol 1990; 26: 261–7.
25. Stenkvist B, Strande G. Analysis of machine-selected cells with an image analysis system in normal and abnormal cervical specimens. Anal Cell Pathol 1989; 2: 1–13.