

DNA histogram interpretation based on statistical approaches

G. Haroske*, V. Dimmer, W. Meyer and K.D. Kunze

Abstract. Image cytometric DNA measurements provide data which are most often interpreted as equivalent to the chromosomal ploidy although the chromosomal and the DNA ploidy are not identical. The common link between them is the cell cycle. Therefore, if destined for DNA ploidy interpretations, the DNA cytometry should be performed on a population-oriented stochastic basis. Using stochastic sampling the data can be interpreted by applying the rules of stochastic processes. A set of statistical methods is given that enables a DNA histogram to be interpreted objectively and without human interaction. These statistics analyse the precision and accuracy of the entire measurement process. They give in error probabilities for accepting a measurement as reliable, for recognition of stemlines, stemline aneuploidy, and for evaluating so-called rare events. Nearly 300 image cytometric DNA measurements from breast cancers and rat liver imprints examples have been selected to demonstrate the efficiency of the statistics in each step of interpreting DNA histograms.

Keywords: DNA image cytometry, statistics, diagnostics, cancer, quality control

1. Introduction

The ESACP consensus [6] on image cytometric DNA measurements provides a basis for standardization of the measurements and several details of the measurement conditions and preparation of specimens. However, the diagnostic and/or prognostic interpretation of the results of DNA image cytometry remains difficult. Compared with other quantitative morphological techniques, DNA cytometry has the advantage of being backed by a clearly understandable biological model of the cell cycle. From a statistical point of view, several parameters for a cell cycle model are possible to describe DNA-stemlines and fractions of the cell cycle. The question raised by the statistics is how these parameters fit the current DNA-histogram.

The most crucial point in DNA cytometry is the rescaling of the DNA axis, i.e., the definition of the position of a DNA-histogram peak, which can be given in relative units only. All further interpretations, be it histogram classifications [1], distributional [2,14,18,22,30] or interpretative variables [3,20, 21] of the DNA measurement have to be based on reliable calibration, i.e., also reliable determination of the position of the histogram peaks in terms of the DNA content.

Furthermore, the seemingly simple definition of a peak itself also remains a matter of discussion. It is closely related to the problems of cell selection for measurements, to the representativity of the measured sample as well as to its validity. As far as the representativity is concerned, the question

*Address for correspondence: Gunter Haroske, M.D., Ph.D., Institute of Pathology, University of Technology, Fetscherstr. 74, D-01307 Dresden, Germany. Tel.: +49 351 458 3005; Fax: +49 351 458 4358; E-mail: haroske@rcs.urz.tu-dresden.de.

is whether all possible cellular clones in a sample are really represented in their true proportions in the measurement data. This is essentially out of our control. The question of validity of the data is dependent on whether the number of cells measured allows a valid interpretation of the data set in terms of the underlying model, e.g., whether a histogram spike represents a peak.

For all of these problems appropriate statistics may provide us with sufficient solutions for reliable DNA cytometry interpretation. The present paper was intended to present such statistical approaches and to demonstrate their actual performance in normal and tumour cell samples.

2. Material and methods of measurement

2.1. Specimen preparation and staining

Twenty-seven rat liver imprints (Fischer rats, both sexes, age between the eighth and ten week) and 247 fine needle aspirates from breast cancers were immediately fixed in 4% buffered formaldehyde. Staining by the Feulgen reaction was made after 40 min hydrolysis in 5 M HCl at 22°C followed by pararosanilin Schiff's reagent (CI 42500) for 60 min. For each staining batch at least one rat liver imprint was stained together with up to ten breast cancer specimens.

2.2. DNA image cytometry

The image cytometry device consisted of an Axioplan microscope (Zeiss, Germany) equipped with a 486/66 MHz IBM compatible PC with a MFG frame grabber (Imaging Technology, USA) and a CCD-TV camera XC-77CE (Sony, Japan). The microscope was also equipped with a computer-controlled, motor-driven xy-scanning stage for exact relocation of all nuclei measured (Zeiss, Germany). For the DNA-cytometry a Plan-Neofluar objective 63×, n.a. 1.25, and a green interference filter (570 nm) were used.

The DNA cytometry is based on routines written in C and linked with the OPTIMAS image analysis system (OPTIMAS Inc., USA). The segmentation of each nucleus, focused individually, is performed by a local multi-step extinction thresholding, which leads to the segmentation at the highest gradient of the spatial transmission profile, taking into account a local background correction.

Before the specimens are measured, the cytometry system is calibrated for a dark (transmission 0%) and a bright field state (transmission 100% at full range of the input video signal) in order to adjust it as near as possible to a linear characteristic.

For a highly precise measurement of the IOD two procedures were combined which compensate for the glare and the diffraction errors of the microscopic system to a large extent [17].

In rat liver specimens 150 diploid, tetraploid, and octoploid nuclei, as well as between 10 and 15 mesenchymal cells were interactively analysed; and in breast tumour specimens 250 tumour cells and 15–30 lymphocytes. Only those cells morphologically well preserved and isolated from each other were selected for the measurement. In each field all cells fulfilling these criteria were measured, including the appropriate internal reference cells. The fields were selected at random.

3. Statistics and results

3.1. Test for uniformity of reference cell populations

In DNA image cytometry all values being interpreted are relative values of the DNA content. Due to the limited reproducibility of DNA stains for absorption cytometry [27], the calibration of the

DNA axis has to be made by the use of reference cells with known DNA content for each specimen measured.

For all problems of the accuracy of DNA image cytometry, the reference cells are of utmost importance. Their measurements reflect the adequacy of the sampling, the homogeneity (or even heterogeneity) of the preparation of the specimen, the stability of the measurement settings and of the optoelectronic properties of the cytometry device. The reference cells should indicate the process quality which can be obtained in a given specimen. Therefore, sensitive tests are needed to check the reference cell data for unimodality, excluding outliers from computation of the reference cell measurement precision. In most of the diagnostic specimens only a few internal reference cells can be found. The Shapiro and Wilk's W -statistic [10,23,24] is a test also suited for the problem of comparatively low cell numbers.

The W -test detects both outliers and departures from normality (e.g., by bimodality) very sensitively. The resulting p -values can be interpreted as red flags for the entire measurement. If they are below an agreed level, e.g., 0.05, the diagnostic evaluation of the specimen should be done with care. In Fig. 1 examples for reference cell populations in 4 different specimens are given. In usual DNA histograms the departures from normality are not obvious. Therefore a very small bin size of the histogram for the reference cells was chosen for this demonstration in Fig. 1 also.

3.2. Detection of stemlines

From the biological point of view the DNA stemline was already defined by Sandritter [25] as a peak in a DNA histogram that represents a $G_{0/1}$ -phase of a proliferating population. It must therefore show a second peak or $G_{2/M}$ -phase cells in a duplication position. However, not all of the spikes in a DNA histogram can be regarded as distinct peaks. The decisive criterion is the validity of a cluster of data values in a data set to be statistically different from other, non-clustered data values (peak events vs. out-of-peak events). This difference must be defined by the precision of the measurements and by the distribution characteristic of the data values, usually known as a distribution density (i.e., a histogram).

A spike in the histogram is thus a peak if a local maximum in the distribution density can be statistically defined by estimating the random events in the measurement.

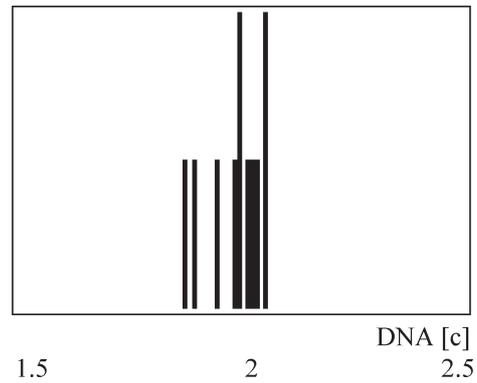
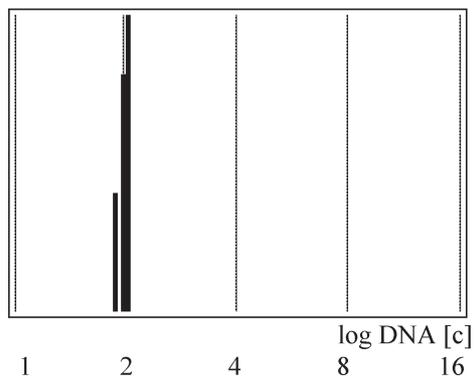
Let us assume a data set of n values with a known standard deviation, SD , then a peak can be found with a given error probability p_{err} (e.g., $p_{\text{err}} < 0.05$) within a region of $\pm SD$ around a real maximum, if the number of events within *this* region is significantly different from zero. The SD can be derived from the precision of the measurements of the appropriate reference cells, e.g., by using the coefficient of variation ($cv = SD/\text{mean}$) of the IOD values from their $G_{0/1}$ -phase-fraction.

The binomial distribution

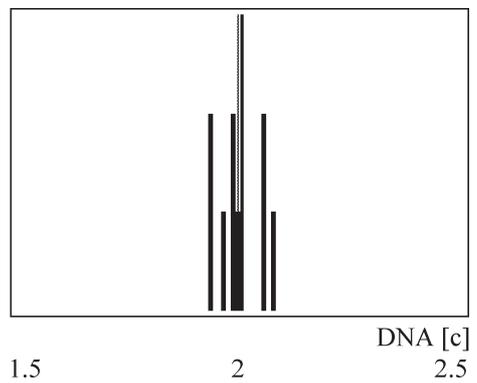
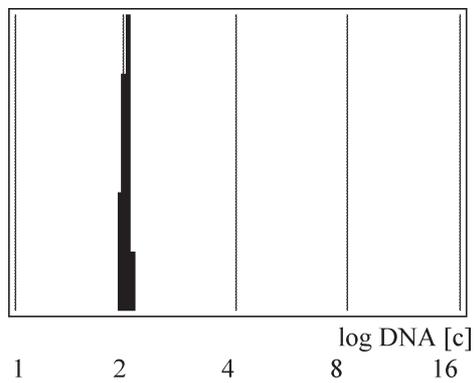
$$p_k^{(n)} = \binom{n}{k} p^k (1-p)^{(n-k)} \quad (1)$$

is used to compare the probability of occurrence of a detected number k of events within a region.

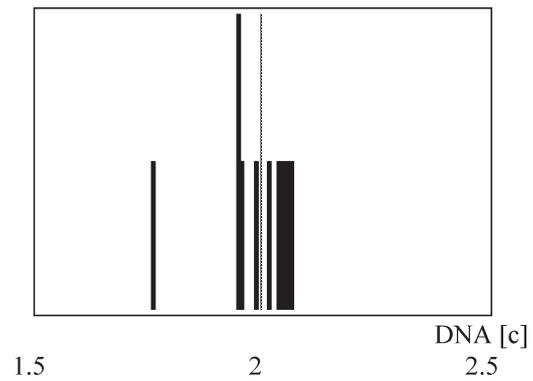
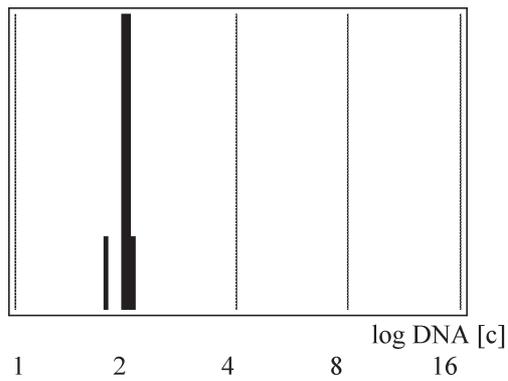
In Fig. 2 an exemplary binomial distribution for $n = 30$ and $p = 0.5$ is shown. In this distribution 95% of all values k are between $k_{\text{low}} = 9$ (with $p_{\text{low}} = 0.025$) and $k_{\text{high}} = 21$ (with $p_{\text{high}} = 0.025$).



(a) $n = 11$, $cv = 2.74\%$, $p = 0.0862$.

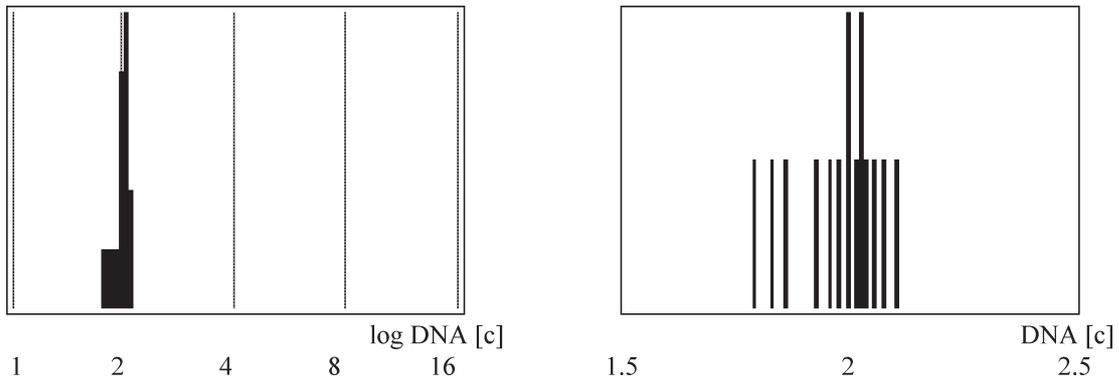


(b) $n = 12$, $cv = 2.16\%$, $p = 0.505$.



(c) $n = 10$, $cv = 2.4\%$, $p = 0.0113$.

Fig. 1. DNA histograms of internal reference cell populations from 4 different breast cancer imprints. Each histogram is shown twice with different scaling. n . . . number of reference cells; cv . . . coefficient of variation (non-parametric estimation from the quantils) of the IOD values; p . . . error probability resulting from the W -statistics. A DNA peak is regarded not to be normally distributed, if $p \leq 0.05$.



(d) $n = 15$, $cv = 4.14\%$, $p = 0.1924$.

Fig. 1. (Continued).

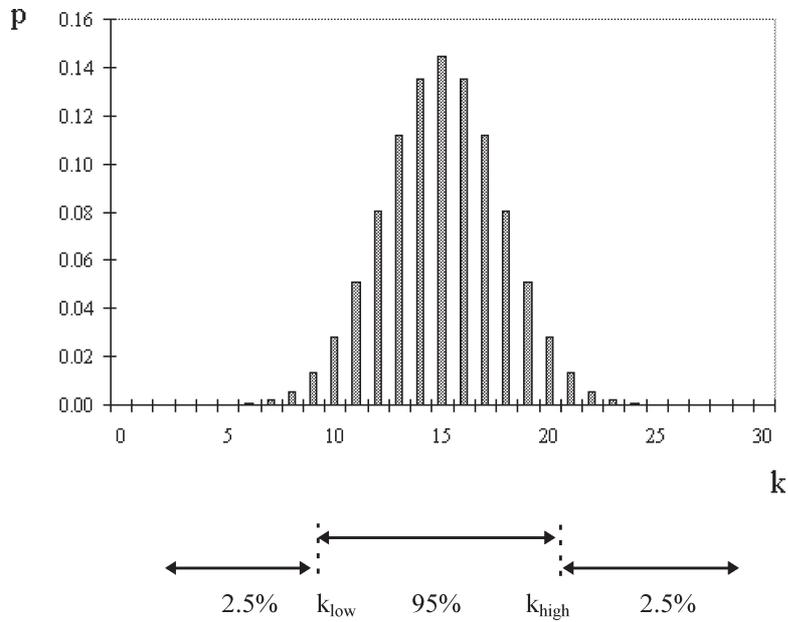


Fig. 2. Binomial distribution for $n = 30$ and $p = 0.5$. In this distribution 95% of all values k are between $k_{low} = 9$ (with $p_{low} = 0.025$) and $k_{high} = 21$ (with $p_{high} = 0.025$).

Values smaller than 9 are not within the same confidence region as values greater than 21. A probability p_0 is calculated which $k_{low} = 0$ for a given number n of nuclei measured:

$$\begin{aligned}
 p_{low} &= p_{0,0}^{(n)} = \binom{n}{0} p_0^0 (1 - p_0)^n, \\
 p_{low} &= (1 - p_0)^n, \\
 p_0 &= 1 - \sqrt[n]{p_{low}}.
 \end{aligned}
 \tag{2}$$

Table 1
Binomial distributions for peak events k_{high} in different total events n

Number of values n	Error probability p	Number of trials	k_{high} according to Eq. (3)	Approximation of k_{high} according to Eq. (4)
10	0.05	10	8	8
		2	6	
		4	7	
		3	7	
20	0.05	20	11	11
		9	10	
		10	10	
50	0.05	50	15	15
		35	14	
		36	14	
100	0.05	100	18	17
		82	17	
		83	17	
200	0.05	200	20	19
		180	20	
500	0.05	500	23	22
		477	23	
1000	0.05	1000	25	24
		975	25	

We now consider the probability $p_{\text{high}} = p_{\text{low}}$ of the occurrence of at least k_{high} events:

$$p_{\text{high}} = 1 - \sum_{k=0}^{k_{\text{high}}} p_{0,k}^{(n)} = 1 - \sum_{k=0}^{k_{\text{high}}} \binom{n}{k} p_0^k (1-p_0)^{(n-k)}. \quad (3)$$

This process has to be repeated until the value of p_{high} is below 0.025 (i.e., half of the error probability desired). The resulting k_{high} is that number of cells not being in the same confidence region with k_{low} .

Let us start the search for the peak at the first (lowest) value of the sorted data values, and move forward value by value making n trials to find a peak.

In each trial we have to check whether the number of values within the \pm SD region around a maximum is greater than k_{high} . In such a region 68% of all cells of a peak would be found. It should be kept in mind that, due to the multiple trials, we have to take into account not the error probability p_{err} , but the error probability p_{err}/n while making n multiple trials (Bonferroni adjustment [8]); i.e., the error probability for each single trial is much lower.

A peak is found if the statistics confirm the occurrence of more than k_{high} events within the test region.

An empirical approximation of (3) for an error probability $p_{\text{err}} = 0.05$ can be given (max. error of $k_{\text{high}} \pm 1$ event):

$$k_{\text{high}} \approx 6 \left(\log_{10} n - \frac{1}{\log_{10} n} \right) + 8. \quad (4)$$

In Table 1 the figures of binomial distributions for peak events are shown, which in given numbers n of data values with a desired error probability p_{err} have a maximum number k_{high} statistically not different from zero. It becomes evident that this procedure is an iterative, converging process.

The table is to be read as follows: If 10 data values have been measured, then 10 trials can be performed. The resulting k_{high} is 8. In the next step 10 minus 8 values are available for the test,

resulting in $k_{\text{high}} = 6$. The third step can use 10 minus 6 values, resulting in $k_{\text{high}} = 7$. The further steps converge at 7.

The maxima found by these procedures can be used as starting variables for any Gaussian fitting to compute the means, variances and counts of the peaks. The resulting peaks are subtracted from the original data measurement and the entire procedure is repeated until no statistically significant peak remains.

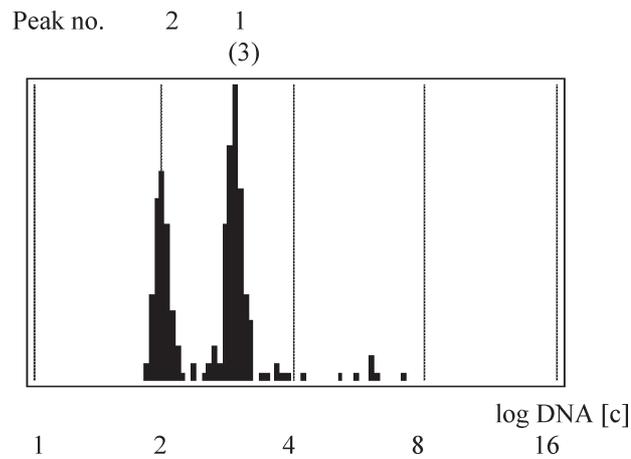
Figures 3–5 show examples of DNA histograms of breast cancers. The effect of the peak recognition approach is shown in each histogram using the simplified formula (4). If Figs 4 and 5 are compared one can demonstrate that the visual impression of the histogram shape can be misleading concerning the recognition of peaks.

The search for a peak should be performed in a cumulative frequency distribution rather than in a frequency density distribution (classical histogram), because in the former no assumptions for the histogram bin size (and therefore for the validity of the histogram classes) are necessary [28].

Finally, a peak is then a stemline if a second peak or also out-of-peak events (see below) are to be found in the duplication region.

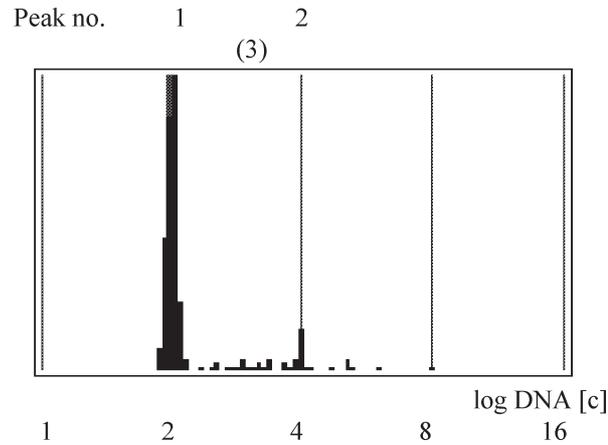
3.3. Test for DNA-stemline abnormality (aneuploidy)

As mentioned above, the calibration of the DNA axis has to be performed by the use of reference cells with a known DNA content, which also have measurement errors. In the theoretical model assumptions for the cell cycle, the reference cells should represent the “truly normal” $G_{0/1}$ -cycle-phase. Therefore, the reference cells have to be included in the model assumptions as well.



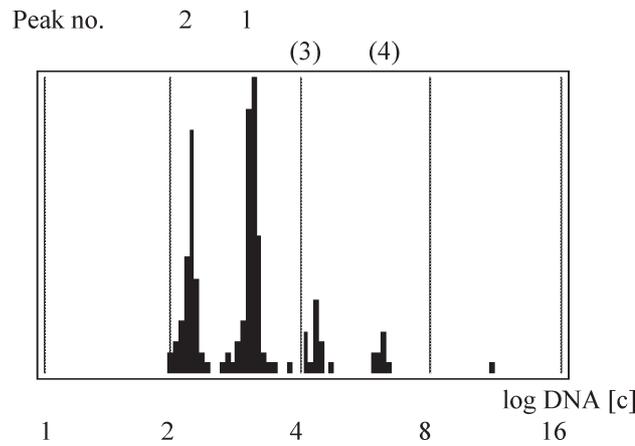
Peak detected	Number of nuclei at start (trials)	Peak position (c)	Peak cv (%)	Number of nuclei within the peak	k_{high} (upper limit of being not a peak)	Significance of being a peak
1	231	2.91	3.95	117	20	yes
2	114	2.00	3.81	88	17	yes
3	26	3.15	2.20	5	13	no

Fig. 3. DNA histogram of a breast cancer with obviously three “spikes”. Two of them are defined statistically as peaks. The details of the statistical figures are given below the histogram.



Peak detected	Number of nuclei at start (trials)	Peak position (c)	Peak cv (%)	Number of nuclei within the peak	k_{high} (upper limit of being not a peak)	Significance of being a peak
1	251	2.00	2.98	200	20	yes
2	51	3.97	1.45	16	15	yes
3	35	2.93	4.49	12	13	no

Fig. 4. DNA histogram of a breast cancer with obviously two “spikes”. Both are defined statistically as peaks. The details of the statistical figures are given below the histogram.



Peak detected	Number of nuclei at start (trials)	Peak position (c)	Peak cv (%)	Number of nuclei within the peak	k_{high} (upper limit of being not a peak)	Significance of being a peak
1	165	3.09	2.81	72	19	yes
2	93	2.20	4.30	55	17	yes
3	38	4.32	2.21	13	14	no
4	25	6.05	2.30	9	12	no

Fig. 5. DNA histogram of a breast cancer with obviously 4 “spikes”. Two of them are defined statistically as peaks. The details of the statistical figures are given below the histogram.

A test to show whether two samples from one case (e.g., internal reference and analysis cells) are from one population, is not reliable, because those samples do not belong to a common population. It is not clear whether the measurement system yields identical IOD values in those two samples because the effects of preparation, glare, diffraction, digitalisation, etc., may be quite different in the different cell types [9,28], as well as in those obtained from the same tissue. By using external reference cells this becomes still more obvious. But the ratio of the mean IOD of the $G_{0/1}$ -phase-fractions from both sample types can be tested in many cases for its statistical properties; it is then the population of these ratios in normal (euploid) cases which is compared to that ratio from a single case. These ratios need not only be ratios between any reference and analysis cells, but ratios between $G_{0/1}$ - and $G_{2/M}$ -phases of analysis cells can also be used to check for polyploidization of a tissue.

For a given DNA-peak there are three error sources for the determination of ploidy:

- the standard errors (SE) of the mean of the analysis cells;
- the SE of the mean of the reference cells (diploid cells from a normal tissue (lymphocytes, granulocytes, rat hepatocytes, etc.); diploid analysis cells in the $G_{0/1}$ -phase, diploid analysis cells in the $G_{2/M}$ -phase, nondiploid analysis cells in the $G_{0/1}$ -phase, nondiploid analysis cells in the $G_{2/M}$ -phase of the same specimen if the polyploidization (i.e., duplication and quadruplication of a stemline) is tested);
- the variance of the tested ratio between analysis cells and reference cells of all normal specimens measured under the same conditions.

Any test statistic has to take into account all of these error sources.

Let

- M_R = mean IOD of $G_{0/1}$ -phase of reference cells;
- S_R = standard deviation of the reference cells' $G_{0/1}$ -phase IOD;
- n_R = number of $G_{0/1}$ -phase reference cells measured;
- M_A = mean IOD of analysis cells (in the peak);
- S_A = standard deviation of analysis cells (in the peak);
- n_A = number of analysis cells (in the peak) measured;

and

- R = IOD ratio M_A/M_R in the specimen under investigation;
- cf_{mean} = mean of the IOD ratios M_A/M_R (so-called corrective factor) in a series of normal specimens, prepared under the same methodological conditions;
- S_{cf} = standard deviation of the IOD ratios M_A/M_R in a series of normal specimens, prepared under the same methodological conditions;

then, a statistical test value t is given by the difference of cf_{mean} and R , divided by the square root of the sum of the variances of the three error sources:

$$t = \frac{cf_{\text{mean}} - R}{\sqrt{S_{cf}^2 + S_A^2/(n_A M_R^2) + M_A^2 S_R^2/(n_R M_R^4)}}. \quad (5)$$

With $cv = S/M$ and $cf_{\text{mean}} \approx R$ the test value t is

$$t = \frac{cf_{\text{mean}} - R}{\sqrt{R^2 cv_{cf}^2 + R^2 cv_R^2/n_R + R^2 cv_A^2/n_A}}, \quad t = \frac{cf_{\text{mean}}/R - 1}{\sqrt{cv_{cf}^2 + cv_R^2/n_R + cv_A^2/n_A}}. \quad (6)$$

Table 2

Actual cf_{mean} and cv_{cf} for different reference/analysis cell combinations in 27 rat liver imprints and 53 diploid breast cancer FNAB with euploid stemlines. The results are separated for rat liver imprints, tumour specimens, and both

Reference system	cf_{mean}	cv_{cf}
<i>Rat liver</i>		
diploid hepatocytes vs. internal reference	1.0131	0.0174
tetraploid hepatocytes vs. internal reference	2.0159	0.0167
octoploid hepatocytes vs. internal reference	3.8992	0.0191
tetraploid hepatocytes vs. diploid hepatocytes	1.9899	0.0052
octoploid hepatocytes vs. diploid hepatocytes	3.8227	0.0276
octoploid hepatocytes vs. tetraploid hepatocytes	1.9208	0.0241
<i>Tumour</i>		
diploid tumour cells vs. internal reference	1.0262	0.0258
tetraploid tumour cells vs. internal reference	2.0138	0.0291
octoploid tumour cells vs. internal reference	4.0338	0.0141
tetraploid tumour cells vs. diploid tumour cells	1.9865	0.0261
octoploid tumour cells vs. diploid tumour cells	4.0242	0.0127
octoploid tumour cells vs. tetraploid tumour cells	1.9624	0.0047
<i>Rat liver + tumour</i>		
diploid epithelials vs. internal reference	1.0217	0.0241
tetraploid epithelials vs. internal reference	2.0147	0.0249
octoploid epithelials vs. internal reference	3.9387	0.0235
tetraploid epithelials vs. diploid epithelials	1.9890	0.0165
octoploid epithelials vs. diploid epithelials	3.8731	0.0332
octoploid epithelials vs. tetraploid epithelials	1.9331	0.0225

The test value t is normally distributed with a mean = 0 and a SD = 1. Then the probability of the null-hypothesis (bilateral) that the ratio tested belongs to the population of “normal” ratios is defined by

$$p(t) = \frac{2}{\sqrt{2\pi}} \int_t^{\infty} e^{-x^2/2} dx. \quad (7)$$

From Eq. (6) it becomes evident how the three kinds of error mentioned above influence the test value and consequently the error probability. The error of cf_{mean} is independent of the number of cells measured. If the cv of the IOD of reference cells is lower than 5% (compare the ESACP consensus [6]) and only 8 reference cells are measured, then its share in the overall error is about 50%. From the statistical point of view it would not be necessary to measure more cells. However, the problems of their representativity are not concerned in these considerations.

If the DNA-stemlines were detected as described above, the measurements usually result in cv_A of analysis cells between 3–4%, and in cv_R of reference cells between 2–3%. Concerning the IOD ratios between reference cells and diploid analysis cells, the measurements are both precise and accurate, with a cv_{cf} of the IOD ratio of about 3%. The mean IOD ratio cf_{mean} is near 1.0, 2.0 and 4.0, respectively. In Table 2 the actual cf_{mean} and cv_{cf} for different reference/analysis cell combinations are given. By analysing the sources of the cv_{cf} in a set of 22 measurements of five rat liver imprints it becomes evident that the preparative variations contributed nearly double the overall cv than the repeated measurements (Table 3).

Table 3
Ratio (cf) between 2c hepatocytes and internal reference cells in repeated measurements

Specimen	No. of meas.	Mean	SD	
RLL142	2	1.065	0.00565685	
RLL143	5	1.0016	0.01894202	
RLL144	5	1.005	0.00707107	
RLL145	4	1.01325	0.02919332	
RLL150	4	1.02025	0.03604049	
RLL153	2	1.0285	0.00070711	
Total	22	1.0161	0.0267	
			0.0148	due to iteration
			0.0222	due to preparation

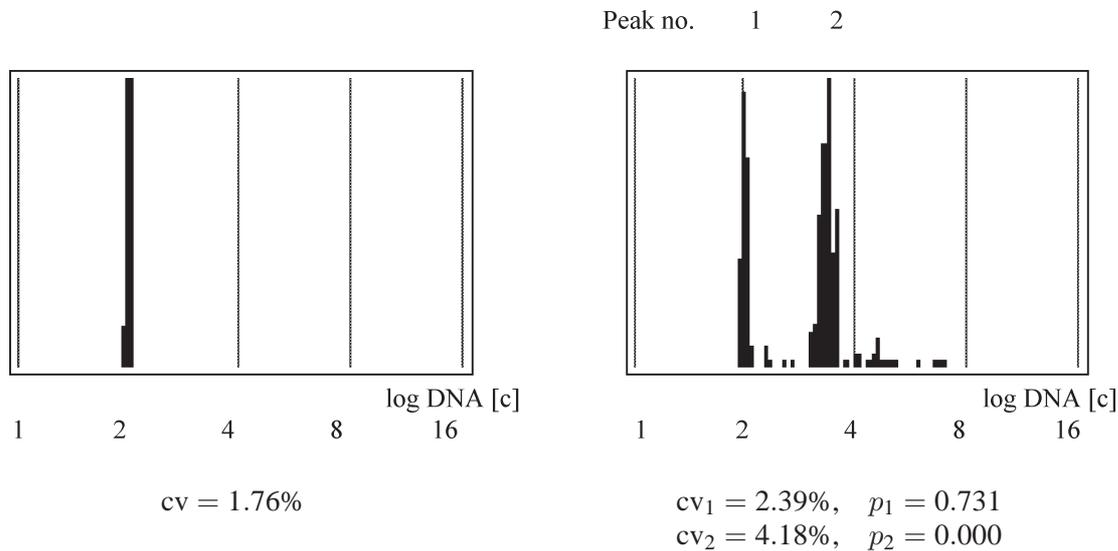


Fig. 6. DNA histograms of the reference cells (left) and the analysis cells (right) of an aneuploid breast cancer. Beside a diploid peak an aneuploid stemline is proven at 3.35 c with $p = 0.00$ (peak no. 2).

If the latter Eqs (6) and (7) are applied to each DNA-stemline, then error probabilities show to what extent the peak is different from the reference system. This is shown for two examples of aneuploid breast cancers in Figs 6 and 7. The error probability is a statistical figure, clearly defined and independent of additional assumptions or conditions. It could also be used for a continuous grading of DNA ploidy, resulting in “normal” grades around the euploid regions.

3.4. Evaluation of out-of-peak events and cell cycle phases

After one or several peaks have been defined in the DNA-histogram, the cells belonging to these peaks as well as the remaining out-of-peak events can be quantified. However, one of the drawbacks of DNA image cytometry is the comparatively low number of cells analysed that makes such a quantification unreliable. Taking into consideration the low cell numbers, the binomial distributions could be helpful again in the interpretation of those data.

Let us assume a number of events e , e.g., between the $G_{0/1}$ - and $G_{2/M}$ -phase of a detected stemline, with a given proportion p , which is derived (for example) from an assumed S-phase fraction, and a

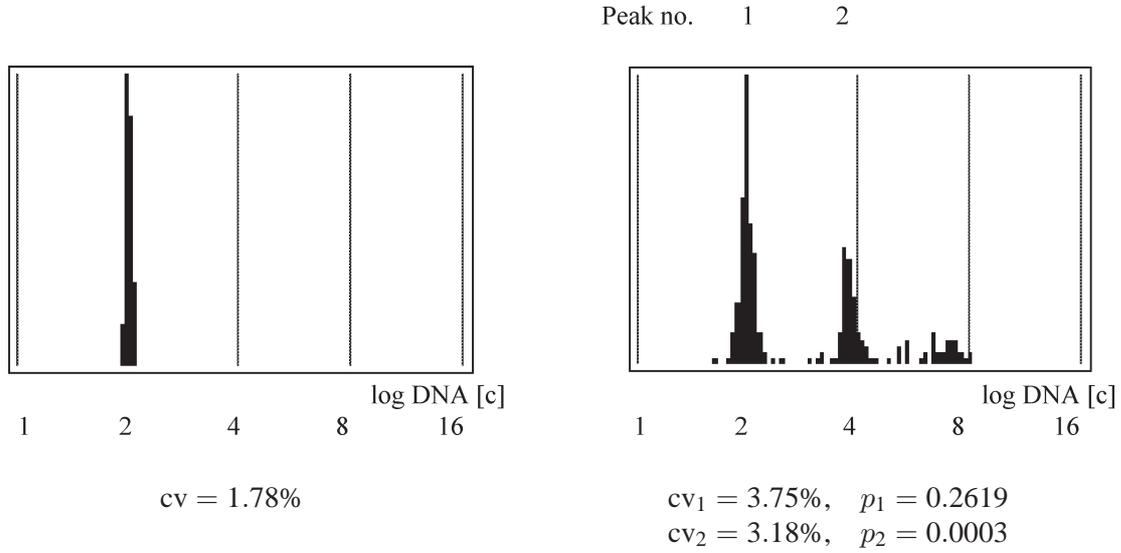


Fig. 7. DNA histograms of the reference cells (left) and the analysis cells (right) of a hypotetraploid breast cancer. Beside a diploid peak an aneuploid stemline is detected at 3.74 c with $p = 0.00003$ (peak no. 2).

Table 4

95% confidence regions of the same probability of occurrence for e events in increasing numbers of n events. In n trials an event occurs with the probability of occurrence p between the lower (n_{low}) and upper (n_{high}) range

n	$p = 0.01$		$p = 0.05$		$p = 0.10$		$p = 0.20$		$p = 0.50$	
	n_{low}	n_{high}								
10	0	0	0	1	0	2	0	4	2	7
20	0	0	0	2	0	4	1	7	6	13
50	0	1	0	5	1	8	5	15	18	31
100	0	2	1	9	5	15	12	27	40	59
200	0	4	4	15	12	28	29	50	86	113
500	1	9	16	34	37	63	83	117	228	271
1000	4	16	37	63	82	118	176	224	469	530
2000	12	28	81	119	174	226	365	434	969	1043

number of cells n in the appropriate stemline (or in the entire population), then the probability that for a given p , e events from n events occur, is determined by

$$p_e^{(n)} = \binom{n}{e} p^e (1-p)^{(n-e)}. \quad (8)$$

The possible number of events with a given error probability p_{err} is in the range between n_{low} and n_{high} with

$$\frac{p_{\text{err}}}{2} \leq \sum_{i=0}^{n_{\text{low}}} p_i^{(n)}; \quad \frac{p_{\text{err}}}{2} \leq \sum_{i=n_{\text{high}}}^n p_i^{(n)}. \quad (9)$$

Table 4 and Fig. 8 show how large the regions are, in those the probability of the occurrence of events from cycle phase fractions as well as out-of-peak events is equal for different absolute numbers of

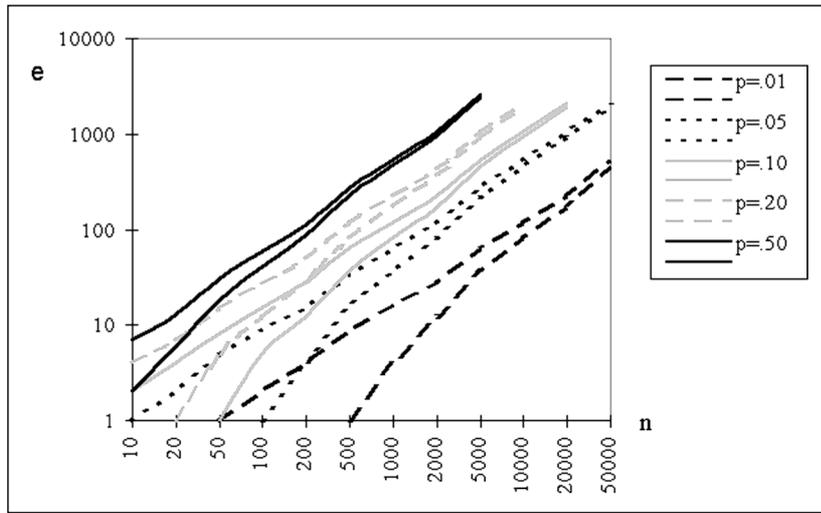


Fig. 8. 95% confidence regions of the same probability of occurrence p for e events in increasing numbers of n events. Each pair of graphs indicate, how in n trials an event occurs with the probability of occurrence p between the lower and upper range (compare with Table 4).

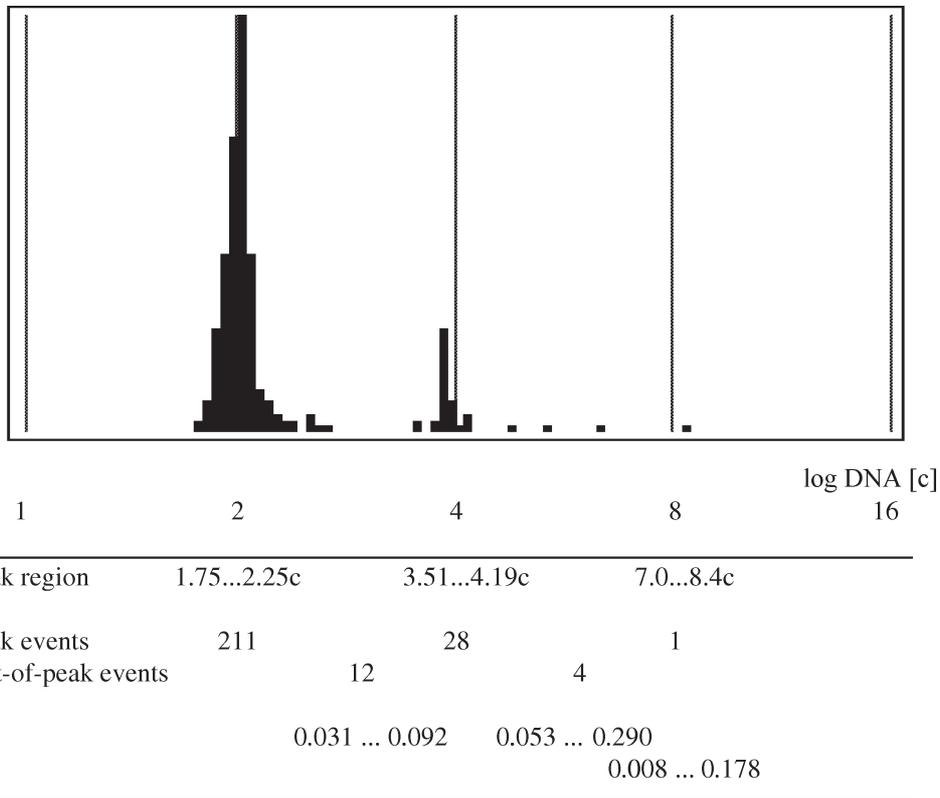


Fig. 9. DNA histogram with out-of-peak events. The regions of the stemline (99.5% of the cells belonging to the stemline) and its polyploidizations are shown with the actual numbers of events and their probabilities of occurrence.

events, or vice versa, the probabilities of occurrence are in the same confidence region for a given event number.

For example, a DNA ploidy analysis of a breast cancer resulted in a DNA histogram shown in Fig. 9. This histogram was evaluated for its cycle phases, peaks, and out-of-peak events as follows:

The DNA histogram shows two peaks and a few out-of-peak events. The S-phase cells in relation to the appropriate diploid and tetraploid $G_{0/1}$ peaks have strongly overlapping regions of their probabilities of occurrence. The single nucleus at 8c can also be the polyploidized $G_{2/M}$ part of 4c. The 4 cells between $G_{2/M}(4c)$ and $G_{2/M}(8c)$ do not have a higher probability of occurrence than the S-phase cells between 2c and 4c, thus not implicating a higher proliferative activity of the tetraploid cycle compared with the diploid.

However, the quantitation of cycle phases in DNA histograms with more than one stemline requires *a priori* knowledge about the $G_{0/1}$ -phase fractions, which cannot be obtained from the histogram alone. A reasonable quantitation should therefore be restricted to histograms with only one stemline.

4. Discussion

Although widely used, the term DNA ploidy has not been defined clearly so far. If it is assumed to be the quantitative cytometric equivalent of the chromosomal ploidy [7], the link between both terms is the cell cycle. Both terms, however, are not identical, as it was stressed by Schulte et al. [28]. Whereas the chromosomal ploidy is theoretically detectable by cytogenetic methods in each single cell, the DNA content of a single cell cannot be equated unequivocally with a certain chromosomal conformation of this cell. Therefore the DNA ploidy is understood here as a cell-cycle related variable, not to be applied to single cells. Consequently, the DNA cytometry should be performed on a population-oriented, stochastic basis. Its evaluations have then to be based on statistical methods. Given a stochastic sampling, the interpretation of the data can be done reliably by applying the rules of stochastic processes. This approach does not reflect directly on the diagnostic meaning of the results of DNA cytometry, i.e., whether a lesion is “normal” or “abnormal”. Certain diagnoses can be also made without these statistical preconditions, e.g., the occurrence *per se* of one cell with an abnormally high DNA content indicates malignancy in certain lesions.

The advantage of statistical approaches is that diagnostic conclusions can be drawn at a specified level of error probabilities, thus making the errors a quantifiable factor. This approach allows both to postulate *a priori* performance levels and to specify *a posteriori* the degree of accuracy of an actual measurement.

The statistics have to include both the precision and the accuracy of the measurements. In the huge body of literature about image cytometric DNA ploidy analysis, only two approaches can be found which considered the inherent measurement errors (or the precision) in a reproducible way for the evaluation and interpretation of the DNA data [5,15]. Yet, in these approaches as well the measurement errors were considered to be confined to limitations in the precision of single cell measurements and of single specimens. In almost all papers a factor for the correction of the mean IOD of the reference cells is given to find the “real” diploid value for the analysis cells. The statistics of this factor remained neglected so far. However, all experience in DNA cytometry is contrary to the assumption that the IOD ratio between reference cells (both internal and external) and analysis cells, the so-called correction factor, could be kept constant [28]. Not only that these correction factors must be determined carefully and extensively for any reference/analysis cell combination, but also the variance of these factors must be measured in comprehensive test sets. It is this variance that

mainly defines the degree of accuracy that can be reached in a DNA ploidy analysis, as it was already demonstrated in a previous paper [17]. All methodological and technological improvements [9,11,12,16,19,26] of the measurement process which lead to an increased precision will also reduce this very important error source, and will therefore improve the accuracy.

On the other hand, interpretation algorithms for DNA-stemlines, referred to in the literature, do really respect the variance, mentioned above, in a rather empirical way: the histogram regions defined as diploid (and tetraploid) are more or less dependent on the variation of the correction factors, although this is not obvious [13,29,31].

But there is another important aspect of a generally increased precision of the measurements. The higher this precision is, the more dependent the interpretations will be on methodological errors in the actual measurements. Therefore, the entire process has to be controlled by statistics to indicate deviations from the expected precision as sensitively as possible. This control concerns the sampling, the preparation effects, the reliability of the instrumentation and of the reference cells. For the first time the statistical procedures proposed in this study allow such a control of the entire process: In a first step the representativity of the reference cells concerning their uniformity is tested. In the next step the precision that can be reached in a given specimen is defined by the coefficient of variation of the reference cells. This precision is used for the estimation of the validity of analysis cell peaks in a third step. In a further step the precision of the actual measurements and the variance of the correction factor, known from previous measurements, are the basis for the interpretation of a stemline of the analysis cells. Finally, the out-of-peak events can be evaluated as to their designation and quantitation of cycle phases.

For each of these steps appropriate statistical measurements have to be used that are sensitive as well as robust enough for the practical use. Thus, the sometimes recommended Kolmogoroff–Smirnov test [5,28] is not sensitive enough for the test of departure from normality in the case of reference cells. But it is a very robust non-parametric test not requiring any assumptions about the type of distributions compared with each other. Therefore, for some problems combinations of test statistics had to be found.

At each step a statistical figure, the error probability, emerges which is continuously distributed between 0.0 and 1.0. Therefore, probability thresholds for each of these steps can be defined, when this is exceeded then a particular measurement is designated as borderline or out of acceptable limits. Usually, these thresholds would be set at 0.05. It should be the aim of scientific committees to agree on the appropriate thresholds in order to ensure a product quality of the DNA image cytometry. All DNA ploidy analyses could be certified by such a procedure.

The consequence could be an automatic quality control and interpretation procedure for image cytometric DNA ploidy analyses, provided by a server or by software manufacturers to all of their users.

Acknowledgements

This work was a part of the EUROPATH project (project no. HC1038 Telematics Applications), supported by DG XIII, European Commission.

References

- [1] G.U. Auer, T.O. Caspersson and A.S. Wallgren, DNA content and survival in mammary carcinoma, *Anal. Quant. Cytol.* **2** (1980), 161–165.

- [2] A. Böcking, C.P. Adler, H.H. Common, M. Hilgarth, B. Granzen and W. Auffermann, Algorithm for DNA-cytophotometric diagnosis and grading of malignancy, *Anal. Quant. Cytol.* **6** (1984), 1–8.
- [3] A. Böcking and W. Auffermann, Algorithm for DNA-cytophotometric diagnosis and grading of malignancy, *Anal. Quant. Cytol. Histol.* **6** (1986), 363.
- [4] A. Böcking, E. Striepecke, H. Auer and L. Fützesi, Static DNA cytometry. Biological background, technique, and diagnostic interpretation, in: *Compendium on the Computerized Cytology and Histology Laboratory*, G.L. Wied, P.H. Bartels, D.L. Rosenthal and U. Schenck, eds, *Tutorials of Cytology*, Chicago, 1994, pp. 107–128.
- [5] A. Böcking, S. Biesterfeld and S. Liu, DNA distribution in gastric cancer and dysplasia, in: *Precancerous Conditions and Lesions of the Stomach*, I.-C. Zhang and K. Kawai, eds, Springer, Berlin, 1993, pp. 103–120.
- [6] A. Böcking, F. Giroud and A. Reith, Consensus report of the ESACP task force on standardization of DNA image cytometry, *Anal. Cell. Pathol.* **8** (1995), 67–74.
- [7] A. Böcking, DNA image cytometry. When and why?, in: *Compendium on the Quality Assurance, Proficiency Testing and Workload Limitations in Clinical Cytology*, G.L. Wied, P.H. Bartels, D.L. Rosenthal and U. Schenck, eds, *Tutorials of Cytology*, Chicago, 1995, pp. 170–188.
- [8] C.E. Bonferroni, Teoria statistica delle classi e calcolo delle probabilita, in: *Volume in Onore di Ricardo Dalla Volta*, Universita di Firenze, 1937, pp. 1–62.
- [9] P. Chieco, A. Jonker, C. Melchiorri, G. Vanni and C.J. Van-Noorden, A users's guide for avoiding errors in absorbance image cytometry: a review with original experimental observations, *Histochem. J.* **26** (1994), 1–19.
- [10] W.J. Dixon, ed., *BMDP Statistical Software Manual*, Vol. 1, Univ. Calif. Press, Berkeley, 1992, pp. 577–578.
- [11] W.A.L. Duijndam, A.W.M. Smeulders, P. van Duijn and A.C. Verweij, Optical errors in scanning stage absorbance cytophotometry. I. Procedures for correcting apparent integrated absorbance values for distributional, glare and diffraction errors, *J. Histochem. Cytochem.* **28** (1980), 388–394.
- [12] W.A.L. Duijndam, Optical errors in scanning stage absorbance cytophotometry. II. Application of correction factors for residual distributional error, glare and diffraction error in practical cytophotometry, *J. Histochem. Cytochem.* **28** (1980), 395–400.
- [13] G. Forsslund and A. Zetterberg, Ploidy levels determinations in high-grade and low-grade malignant variants of prostatic carcinoma, *Cancer Res.* **50** (1990), 4281–4285.
- [14] S.D. Fossa, O. Kaalhus and O. Scott-Knudsen, The clinical and histopathological significance of Feulgen DNA-values in transitional cell carcinoma of the human urinary bladder, *Eur. J. Cancer* **13** (1977), 1155–1162.
- [15] Y.S. Fu, L. Cheng, I. Huang, S. Huang, E. Wiesmeier, F. Wettstein and M. Weissman, DNA analysis of cervical condyloma and intraepithelial neoplasia in specimens obtained by punch biopsy, *Anal. Quant. Cytol. Histol.* **11** (1989), 187–195.
- [16] G.J. Goldstein, Aspects of scanning microdensitometry. I. Stray light (glare), *J. Microsc.* **92** (1970), 1–16.
- [17] G. Haroske, W. Meyer, F. Theissig and K.D. Kunze, Increase of precision and accuracy of DNA cytometry by correcting diffraction and glare errors, *Anal. Cell. Pathol.* **9** (1995), 1–12.
- [18] W. Hiddemann, J. Schumann, M. Andreef, B. Barlogie, C.J. Herman, C.R. Leif, B.H. Mayall, R.F. Murphy and A.A. Sandberg, Convention on nomenclature for DNA cytometry. Committee on nomenclature, Society for Analytical Cytology, *Cancer Genet. Cytogenet.* **13** (1984), 181–183.
- [19] D. Kindermann and C.H. Hilgers, Glare-correction in DNA image cytometry, *Anal. Cell. Pathol.* **6** (1994), 165–180.
- [20] R. Montironi, L. Diamanti, A. Santenelli and M. Scarpelli, Computer aided S-phase fraction determination in DNA static cytometry in breast cancer, *Anal. Quant. Cytol. Histol.* **14** (1992), 379–385.
- [21] M. Opfermann, G. Brugal and P. Vassilakos, Cytometry of breast carcinoma: Significance of ploidy balance and proliferation index, *Cytometry* **8** (1987), 217–224.
- [22] J.J. Ploem-Zaaijer, M. Beyer-Boon, L. Leydte-Veldstra and J.S. Ploem, Cytofluorometric and cytophotometric DNA-measurements of cervical smears using a new bicolor method, in: *Automation of Cancer Cytology and Cell Image Analysis*, N.J. Pressmann and G.L. Wied, eds, *Tutorials of Cytology*, Chicago, 1979, pp. 225–235.
- [23] J.P. Royston, An extension of Shapiro and Wilk's *W*-test for normality to large samples, *Applied Statistics* **31** (1982), 115–124.
- [24] A.E. Sarhan and B.C. Greenberg, *Contributions to Order Statistics*, Wiley, New York, 1962.
- [25] W. Sandritter and M. Carl, Cytophotometric measurements of the DNA content (Feulgen reaction) of malignant human tumors, *Acta Cytol.* **10** (1966), 26–30.
- [26] H. Schimmelpenning, K. Hamper, U.G. Falkmer, J. Caselitz, G. Seifert and G.U. Auer, Methodological aspects of DNA assessment by means of image cytometry in tumors of the salivary glands, *Anal. Quant. Cytol. Histol.* **11** (1989), 379–383.
- [27] E.K.W. Schulte, Standardization of the Feulgen reaction for absorption DNA image cytometry: A review, *Anal. Cell. Pathol.* **3** (1991), 167–182.
- [28] E.K.W. Schulte, D. Seigneuring, F. Giroud and G. Brugal, DNA densitometry, in: *Quantitative Clinical Pathology*, P.W. Hamilton and D.C. Allen, eds, Blackwell Science, 1995, pp. 140–169.

- [29] E. Sprenger, L. Volk and W.E. Michaelis, Die Aussagekraft der Zellkern DNS-Bestimmung bei der Diagnostik des Prostatakarzinoms, *Beitr. Path.* **152** (1974), 370–378.
- [30] B. Stenkvist and G. Strande, Entropy as an algorithm for the statistical description of DNA cytometric data obtained by image analysis microscopy, *Anal. Cell. Path.* **2** (1990), 159–162.
- [31] B. Tribukait, L. Roenstroem and P.L. Eposti, Quantitative and qualitative aspects of flow DNA measurements related to the cytological grade in prostatic carcinoma, *Anal. Quant. Cytol.* **5** (1983), 107–111.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

