15

Use of an Image Analysis System to Count Colonies in Stem Cell Assays of Human Tumors

Bernhardt E. Kressner, Roger R. A. Morton, Alexander E. Martens, Sydney E. Salmon, Daniel D. Von Hoff, and Barbara Soehnlen

As is well demonstrated elsewhere in this text, the clonogenic assay for tumor colony-forming cells has applicability to a broad scope of human tumors and has proved valuable in studies of biology, clinical course, and chemosensitivity of human cancers. The development of this promising new area of clinical research, however, has precipitated a substantial new laboratory problem; namely, the need for automation in counting tumor colonies. This need was not fully apparent until it became clear that the clonogenic assays predicted clinical and biological features of human cancers. In the initial studies, careful qualitative and quantitative evaluations of tumor clusters and colonies in soft agar were conducted by the clinical research laboratory staff of two of the authors (S.E.S., D.D.V.H.). As their studies proceeded, we recognized that there was a major need for a precise automated instrument for selective counting of tumor colonies and therefore initiated a join developmental project with Bausch & Lomb Incorporated on the application of image analysis to this task.

Our experience with visual counting of colonies growing in this assay has defined a number of problems:

1) Visual counting of experiments is time-consuming and therefore very expensive in terms of professional time. It is not uncommon to spend nine hours to count a 100-plate experiment.

2) We have shown (Fig. 1) that in the size range of interest the number of colonies changes very rapidly with size. This means that an error of only ±10% in the visual estimate of colony size can result in +90% to –45% counting error. Since in visual counts small differences in size are very difficult to determine accurately, this problem can contribute to very substantial counting errors.

3) Investigator fatigue becomes a significant factor in counting a series of large assays on one or more specimens, and has been shown to lead to nonreproducibility.

Cloning of Human Tumor Stem Cells, pages 179–193
© 1980 Alan R. Liss, Inc., 150 Fifth Avenue, New York, NY 10011

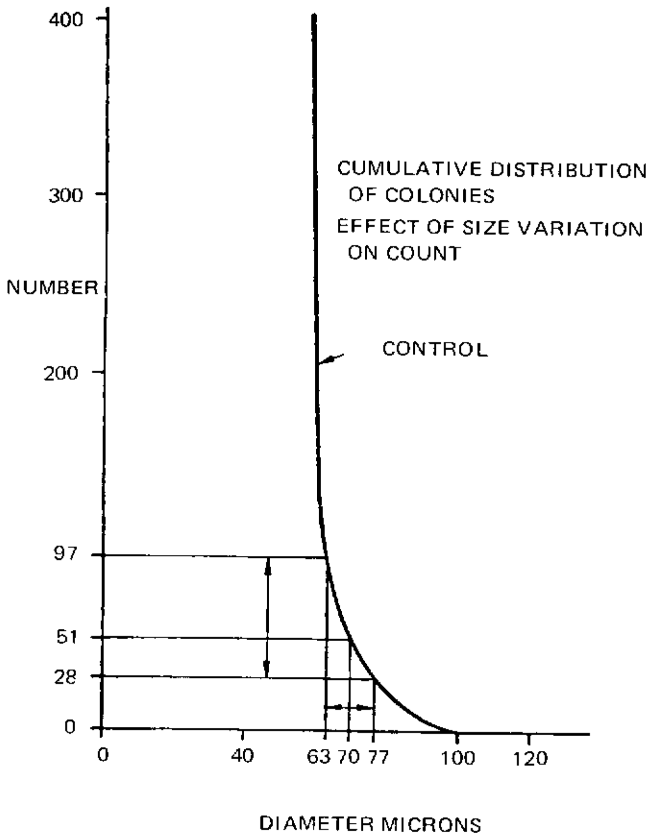


Fig. 1. Plot of cumulative number of colonies versus colony diameter in microns to show extreme sensitivity of counts to changes in size limits.

4) There is variability in counting from one investigator to another, in part because of differences in criteria with respect to cluster and colony size, and additionally because of fatigue, as mentioned above. Errors of 100% and more have been often observed when two people count the same plate.

5) A large amount of time is required to train an individual to count colonies visually with an acceptable accuracy.

6) Standardized definitions, such as size, cut-off criteria and information on the rate of development of clusters and colonies, are highly desirable and virtually require automation if they are to be carried out on more than just a few plates.

Whereas we expected that automated counting equipment could potentially increase counting speed and provide reproducible criteria, we also anticipated substantial difficulties with regard to the instrument's ability to discriminate be-

tween the colonies and extraneous objects. In fact, our concerns were substantiated during our initial attempts to use a relatively simple image-analysis system (The Bausch & Lomb Omnicon Alpha™) in this application.

This instrument's analytical capabilities for differentiating between colonies and other objects were not adequate to produce reliable results. Although the Alpha performed quite well on plates with a large number of colonies (compared to the number of artifacts) it failed to analyze accurately plates with only a few colonies – ie, those with low fractional survival of TCFUs, which are of primary importance in indicating sensitivity to drugs.

We therefore initiated a series of studies using the more powerful, software-based Bausch & Lomb Omnicon™ FAS-II image analysis system with the objective of matching or exceeding human capabilities for correctly identifying tumor clusters and colonies while rejecting artifacts. This objective was attained, and efforts toward reducing counting time and providing additional automated features were initiated.

MATERIAL

Tumor biopsies were prepared and cultured with the method of Hamburger and Salmon [1–3], which is detailed in Chapters 3–10. The standard agar concentrations were used, but many plates (particularly for solid tumors) did not require conditioned media. In addition to plates cultivated at 200,000–500,000 cells/dish, and various reference specimens, additional cell concentrations ($1 \times 10^4 – 1 \times 10^6$) were prepared to test the linearity of the automated equipment in relation to visual counting. Representative cultures from a wide variety of tumor types were investigated (eg, breast, melanoma, ovary, lung, neuroblastoma, and myeloma) from fresh biopsies as well as from human tumor cell lines (myeloid-leukemia HL60, and myeloma 8226). Plates with varying degrees of red cell contamination were also studied, as were plates with a variety of extraneous objects in the agar layers. Colony counts on triplicate samples were performed by two of us (B.S., D.D.V.H.) immediately prior to counting with the FAS-II. Some specimens were also fixed with 3% glutaraldehyde in phosphate-buffered saline so that they could be preserved in a refrigerated humidified chamber for varying periods prior to automatic counting.

EQUIPMENT

The equipment used in these experiments consisted of a microscope equipped with a fast automatic stage and a precision television scanner, the Omnicon FAS-II unit, and a printing terminal. The FAS-II [4], which is a versatile, general-purpose image-analysis system, was modified, as described below, to ensure optimum performance in this application. Also, dedicated software, including special algorithms for object discrimination, was developed.

Both regular and inverted stage microscopes have been used in this application. The inverted stage offers greater convenience in loading and removing the Petri dishes, and also has the very desirable feature of permitting the dish cover to be left in place during counting because the image is formed through the bottom of the dish. Thus the Petri dish contents are continually protected from dust, the drying of the agar is retarded, and the possibility of contamination by biologically active agents (bacteria, mold spores, viruses) is greatly reduced.

The initial machine-counting of the assay plates has been attempted using the regular Omnicon automatic microscope stage designed primarily for operation at high magnification. That stage proved to be unacceptably slow, and a new fast stage was developed specifically for colony counting. This stage can move at speeds of up to 35 mm per second, and its introduction made the sample scanning time a small fraction of the total analysis time.

In its present configuration the stage is equipped with a shallow well designed to accept and locate accurately the round 35 mm plates. This arrangement is designed for manual placement and removal of Petri dishes. Several accessories for handling groups of dishes will be added in the future.

We found experimentally that the optimum magnification, as a compromise among depth of field, resolution, and scanning speed, is 2.5 X. At this magnification, the size of each field corresponding to a single TV frame, on the agar surface is 3.3 X 4.4 mm. The scanning sequence is as follows:

1) From the fixed start position the stage moves the center of the dish directly under the objective to allow for focusing adjustments, if required.

2) The stage moves the inner edge of the meniscus to the optical axis to start the analysis.

3) The TV scanner scans this field to acquire the image for analysis.

4) After completion of the image processing, the stage moves to the next field in the scanning pattern. This pattern of 35 fields has been devised to provide maximum coverage (86% of the useful agar surface area limited by meniscus) of the plate in minimum time, with larger area coverage optionally available.

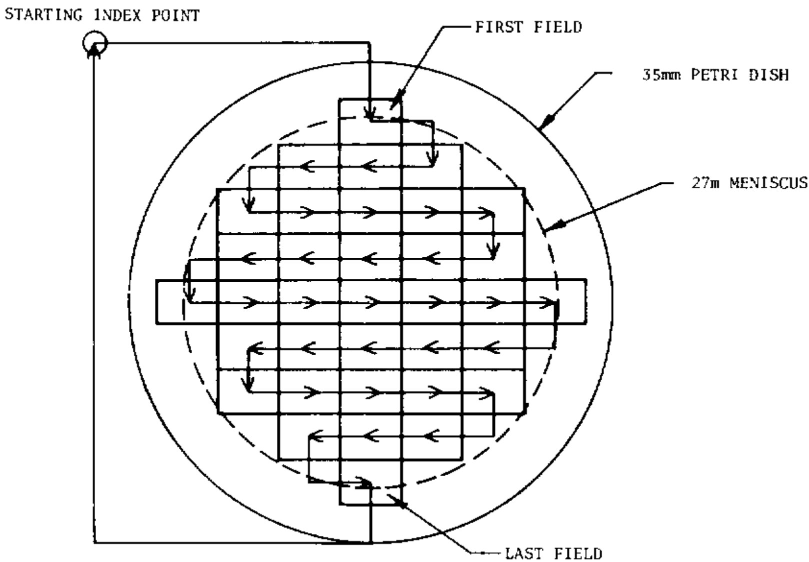
The area for analysis is confined to the flat portion of the agar surface, because the meniscus at the Petri dish walls may cause optical distortion. The overall coverage with this scanning pattern is 51% of the total Petri dish area. Initially, when comparing visual with machine counts to account for the difference in the areas counted, we applied a constant factor to the machine results because we had assumed that the visual counts came from the total area of the plate. As it turned out, investigators doing visual counts also stay within the level area of the agar. For this reason, the constant was dispensed with (Fig. 2).

In order to understand why the FAS-II has been successful in simulating human judgment to perform these analyses, it is necessary to discuss some fundamental concepts. The agar is imaged by the microscope objective onto the photosensitive

STEM CELL ANALYSIS PROGRAM
STAGE MOTION

35 FIELDS

FIELD SIZE: 4.437 X 3.219mm



COVERAGE

51% OF TOTAL AREA

Fig. 2. Diagram of scanning pattern for stage motion in Omnicon stem cell analysis.

surface of a vidicon tube in the scanner, which converts the optical image into electrical video signals. The precision scanner has been designed specifically for image analysis, as distinct from scanners used in closed-circuit or broadcast television. As such, the scanner has a much greater scanning linearity, lower noise, and higher stability to ensure faithful transformation of images into equivalent video signals. These are suitable for extraction of quantitative size, shape, and density information in the FAS-II.

FAS-II is a system in which a computer plays a major role. Nearly all capabilities and functions of this instrument are defined in terms of software programs, which facilitates the adaptation of the FAS-II to a wide range of diverse applications. Consequently, the major tasks in adapting this instrument to colony counting in stem cell assays, in addition to the design of a fast stage and some modifi-

cations in the detection circuitry, were the formulation of the colony selection and measurement algorithms and the software realization of these algorithms.

Besides the computer, the FAS-II includes the precision scanner discussed above, detection circuitry, and special signal processing and interface circuits that allow the computer to accept the video information. The interaction between the system and its operator occurs by means of a keyboard and a television screen which, in addition to the image of the specimen, also displays a variety of messages, queries, and prompters to guide the operator. A printing terminal is also included to generate hard-copy output of results in a desired format.

METHODS

Our initial attempts to use the less powerful image analyzer (Omnicon Alpha) did not yield acceptable results because the instrument was unable to discriminate satisfactorily between the colonies and extraneous objects, such as tissue debris, agar imperfections, bubbles, and occasional colonies of mold. Some of these artifacts are always present in the culture plates despite efforts to prevent them. The limitations of the instrument were most obvious in drug-survival curves.

When visual counts are performed, the individual doing them identifies, on the basis of training and previous experience, which objects in the Petri dish are to be counted or not counted. The decision whether to count or not to count an object is based on this object's appearance – ie, size, shape, optical density, color, and surface texture. Our studies have revealed, however, that shape and optical density are the primary discriminating factors. The person performing a visual count uses a three-step procedure:

1) Decides to accept or reject an object, depending on whether it is a colony or not.

2) If it is a colony, judges whether its size (diameter or maximum chord) exceeds the predetermined minimum size to qualify it for acceptance.

3) If these conditions are satisfied, adds one more count to the running total, and proceeds to the next object.

The techniques for selective automatic counting of stem-cell colonies developed for the FAS-II emulate this procedure.

1) The image from a field on the agar surface is *detected*. This means that the optical density of all objects in that field is analyzed and only those objects that fall into a predetermined optical density range are accepted for further investigation. The detection rejects some of the extraneous objects, while at the same time converting the video signals representing the remaining objects into a digital form for further processing.

Agar substrates are optically nonhomogeneous in transmitted light, so that the optical density of the background against which the objects on the agar surface are detected varies considerably. To deal with this problem and the inherent shading in the optics special circuitry to track the background has been developed to ensure uniform detection across the entire field.

2) The system measures the individual areas of all detected objects, rejecting all that are outside a specified range. This is a fast preselection process that results in a considerable overall time saving by reducing the number of objects to be subjected to the more time-consuming shape measurements.

3) Dimensionless shape parameters (related to elongation and boundary "roughness") are then measured on the remaining objects. These shape parameters and their numerical criteria have been chosen after considerable experimental effort and it appears that they are applicable to a wide range of cancer types. The instrument performs the final selection of the colonies based on the numerical values of these parameters. The four steps of this process are illustrated in Figure 3.

4) After the final selection, the remaining objects now accepted as clusters or colonies are measured. We used two types of measurements: equivalent circular diameter and maximum horizontal chord. The former is derived by measuring the areas and computing diameters of circles corresponding to those areas; the latter is the longest intercept in the direction of the scan. Both of these measurements are reported.

The foregoing four-step procedure is termed "Selective Counting."

RESULTS

From the many experiments and investigations performed over a nine-month period, several findings stand out.

1) Conventional regression analyses on repeated FAS-II versus FAS-II counts showed a four- to eight-fold improvement in stability or reproducibility when compared with visual versus visual counts (Table I).

2) Correlation coefficients relating machine versus visual counts were typically greater than 0.95, with regression coefficients ranging from 0.84 to 0.94. Comparison data from one of the most recent experiments are also listed in Table I. We found, however, that correlation coefficients can be a misleading measure of error, due to the strong influence of the spread or range of counts in a given experiment. Thus individual counts having a certain displacement from a regression line, but which were grouped close to the control counts, produced regression coefficients that differed considerably from those derived from counts that had the same apparent displacement but were grouped farther away from the control value.

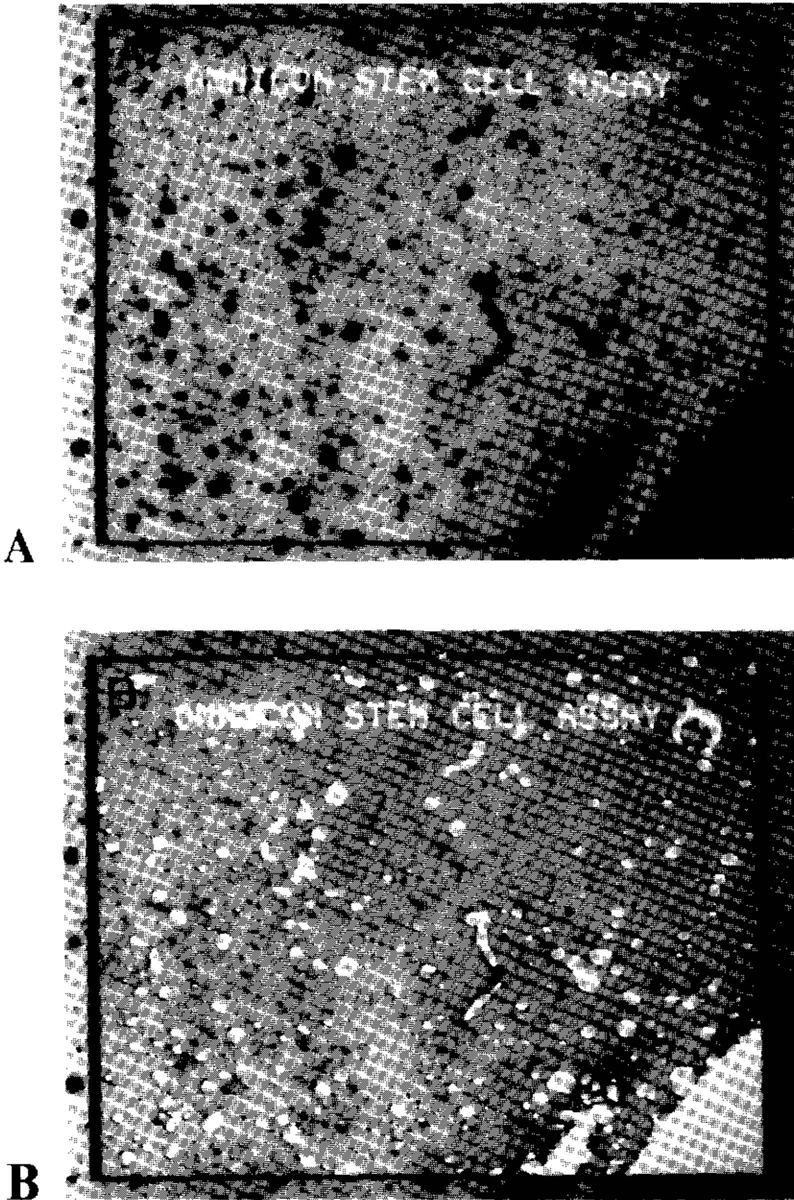
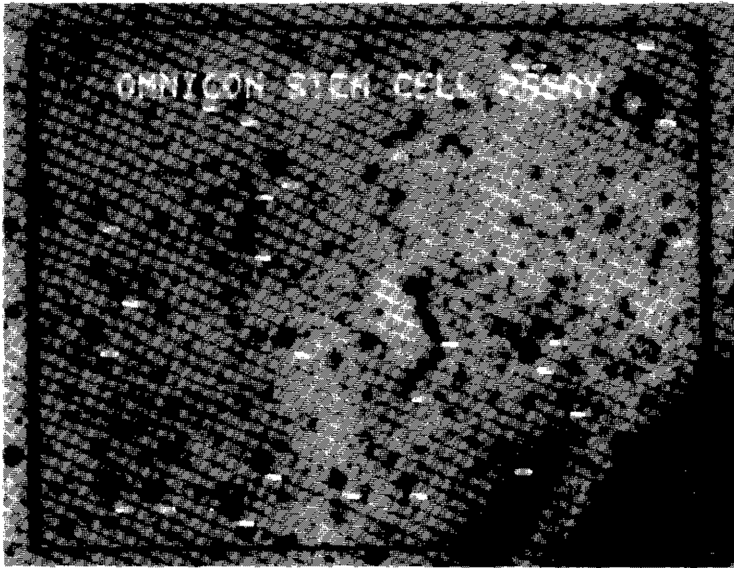
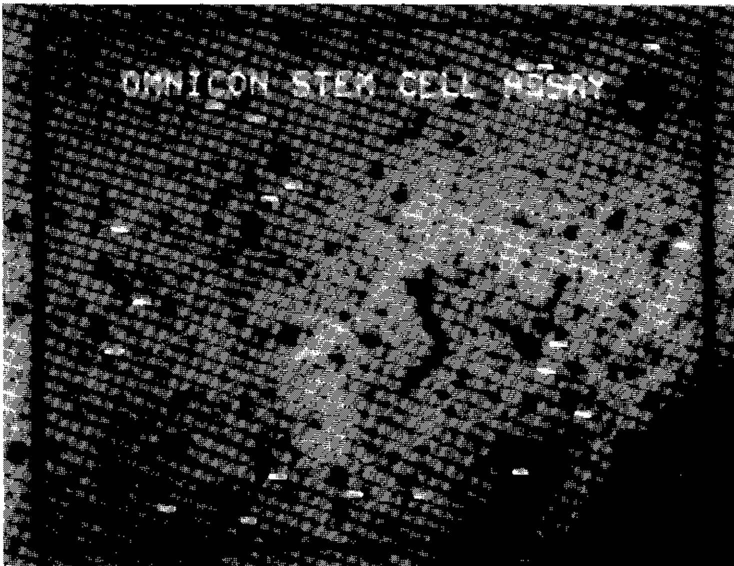


Fig. 3. Four monitor photographs showing the detection and selection process (the white "blips" on panels C and D are tags indicating the selected features): (A) A normal image with colonies, cracks, agar wrinkles, and a bubble; (B) a detected image of the same view;



C



D

(C) tagged features remaining after size selection only; (D) tagged features remaining after size and shape selection. These are now only the colonies of interest.

TABLE I. Tabulation of Two Types of Errors in Comparisons of Machine and Visual Counts

Comparison[a]	Regression coefficient (r^2)	% RSM error
Original counts		
Visual vs visual	0.84	23
FAS-II vs FAS-II	0.98	11
Visual vs FAS-II	–	20
% Survival data		
Visual vs visual	0.91	17
FAS-II vs FAS-II	0.98	8
Visual vs FAS-II	0.94	13

[a]Comparison of visual vs visual and FAS-II vs FAS-II refers to blinded repeat studies with each approach. The analysis has been carried out separately on both original counts and computations of percent survival in relation to controls in drug-treated experiments.

To deal with this problem, a "% RMS error" function was derived for characterizing various comparisons in this application (Fig. 4). Among those advantages, this measure of error represents the overall accuracy of data matching to a greater degree and with more stability than do regression coefficients. Our goal, of course, has been to develop a system to the point where machine counts would match visual counts at least as well as visual counts can match other visual counts. On the basis of either type of error analysis, we have achieved that goal, as is shown quantitatively by the examples in Table I.

More importantly, the ability of the FAS-II to predict the percent survival of TCFU after drug exposure (standard assays as in Chapter 18) closely matches the results of visual (manual) counting. This is shown for three separate studies in Table II.

3) Part of the reason for the relatively poor performance of human counting is shown in Figures 1 and 5. Contrary to the most desired functional relationships for limits in scientific work, those in which a limit value is chosen on a "plateau," the nature of cumulative counting in stem cell assays requires a judgment to be made on the "edge of a cliff!" The problem is of course compounded for an investigator who is required to estimate diameters using a reticle or other eyepiece scale. For example, in Figure 1 a 10% error in estimated diameter at the 70 micrometer level can cause cumulative count values having relative errors of –45% to +90%. It is apparent, therefore, that only an accurate, stable detection system that provides a constant size limit can ensure the required reproducibility, even if other colony recognition factors were not involved. Figure 5 shows a comparison of cumulative counts from a control dish and two drug-treated dishes.

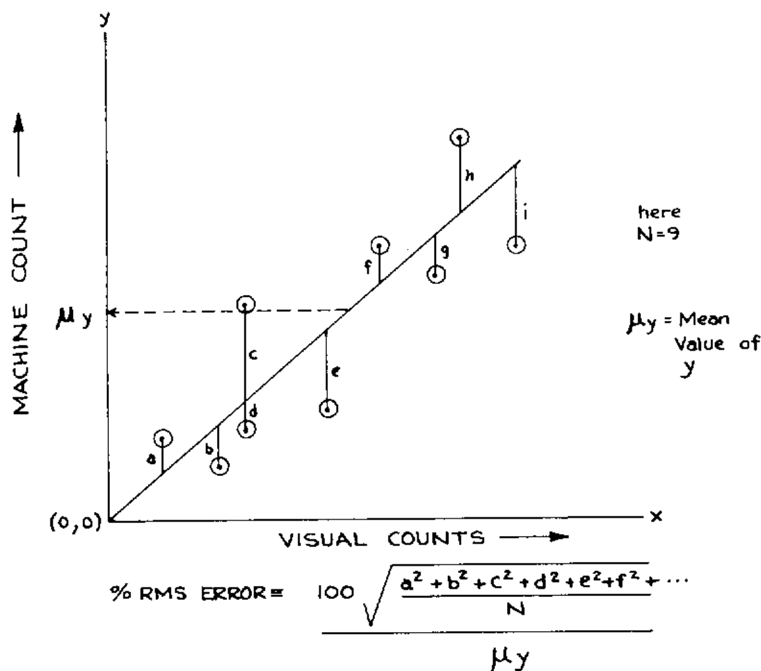


Fig. 4. Diagrammatic plot to demonstrate the meaning of "% RMS Error" using machine versus visual counts.

TABLE II. Survival of TCFU After Drug Exposure: Visual Counts vs FAS-II Counts*

Tumor type	Drug	Visual count (% survival)	FAS-II count (% survival)
Squamous carcinoma of the lung	Methotrexate	85	87
	Adriamycin	82	85
	Chlorambucil	80	75
	Dihydroxyanthracenedione	52	48
Adenocarcinoma of the lung	Methotrexate	100	100
	Adriamycin	10	15
	Dihydroxyanthracenedione	30	20
Ovarian carcinoma	Methotrexate	100	100
	Adriamycin	100	100
	Chlorambucil	100	100
	Cis-platinum	42	30
	Hexamethylmelamine	100	100

*A comparison of percent survival predicted by both visual and FAS-II analyses for three separate experiments.

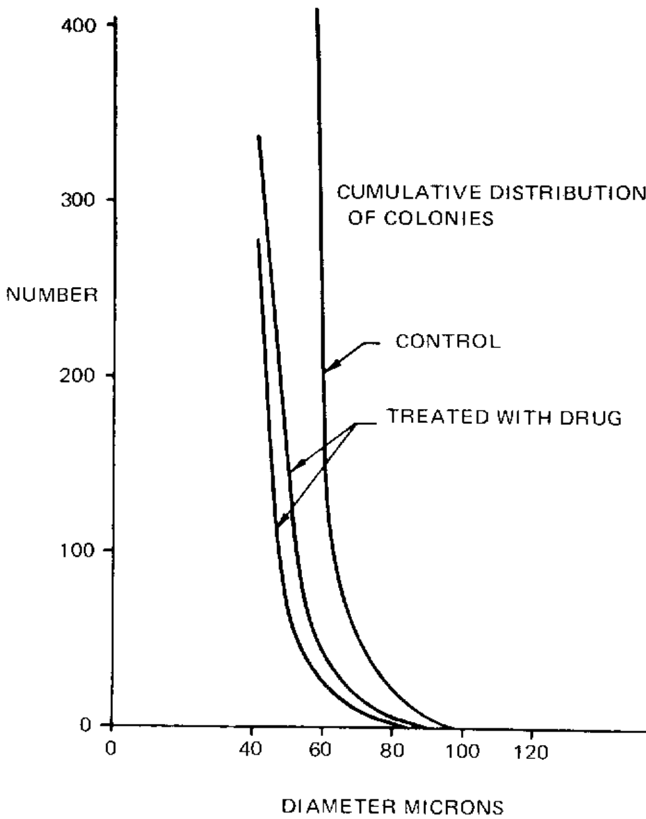


Fig. 5. Plot of cumulative number of colonies versus colony diameter for a control plate and two separate drug experiments.

4) The need for automatic recognition of colonies is demonstrated in Figure 6, which shows the results of counting a typical stem-cell control plate of an ovarian cancer. Without automatic recognition the number of objects (colonies plus extraneous objects or background) is much greater than the actual number of colonies counted using automatic recognition to select the colonies of interest, while excluding from the result all other objects.

Since the Omnicon Alpha lacks adequate ability to differentiate objects based on their shape, the RMS error, as contrasted to that of the FAS-II, was 38% for similar experiments.

5) The average time for the FAS-II to perform an analysis is about three minutes per dish without the use of special programming or the incorporation of high-speed hardware options. With a high-speed system configuration and software, this time will be reduced to less than one minute per dish.

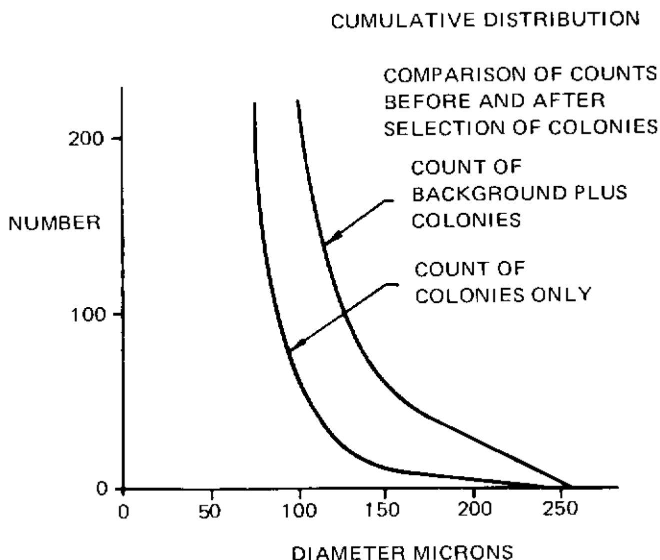


Fig. 6. Plot of cumulative number of colonies versus colony diameter from the same dish with and without use of selection criteria.

DISCUSSION

The extensive experimental evidence obtained to date by personnel from three independent laboratories (University of Arizona, University of Texas, and Bausch & Lomb) indicates that automatic selective colony counting using image analysis technology is not only feasible but also produces results that are in several important ways (speed, repeatability, uniformity of criteria) superior to visual counts.

The development of new instrument capabilities for performing quantitative biological observations on human tumor-colony formation could have major impact on fundamental cancer research. Such capabilities could provide an essential function of selective colony counting for drug screening and routine clinical testing of drug sensitivity on cultured biopsy specimens. Because analytical procedures with the FAS-II system are nondestructive and do not require fixation or staining of the cultures, serial observations can be readily made on multiple samples. Investigations carried out in this fashion could clarify the growth kinetics of clonogenic cells that give rise to clusters and colonies in the presence or absence of cytotoxic drugs or hormone exposure. Such data are of significant importance, as we lack information about what proportion of cells that give rise to clusters are also the progenitors of colonies, and whether spontaneous human tumors contain a hierarchy of progenitors with differing degrees of proliferative and differentiating capability. By examining serial histograms of the size distributions, as well as other features of clusters and colonies, this question may be clarified substantially.

At a more practical level, scanning of the dishes shortly after initial plating would permit enumeration of cellular clumps that might be the result of sample preparation, and would permit, if required, subsequent use of these data for a type of "background subtraction." Alternatively, for routine purposes, several control plates can be fixed in glutaraldehyde (Chapter 12) shortly after they have been plated and stored until final counting to provide a simultaneous evaluation of sample background. Such considerations may be of only limited importance, however, as the common artifacts present in such plates are easily recognized as such by the FAS-II and rejected automatically during the selective counting of colonies.

While initially we had hoped that the less sophisticated Omnicon Alpha would prove adequate to quantify tumor colony formation based merely on size and optical density criteria, these two functions proved to be inadequate because they fail to ensure adequate rejection of artifacts in the plates (bone marrow spicules, clumps, agar holes, or bubbles, etc). Consequently, additional criteria relating to shape, as well as other descriptors, available with the image analysis and detection of the FAS-II proved essential for dealing with some of the most important specimens. For example, in drug assay experiments, wherein reduction of colony-forming units to less than 10% of control was achieved with a drug, the failure to reject extraneous objects had an adverse impact on the accuracy of the data. This is particularly important in the tumor stem cell assay system, since often fewer than 100 colonies are present on the control plates. Thus the usual simple video colony counters that cannot discriminate between colonies and artifacts fail to compute accurately the most important experimental results. Video colony counting might, for instance, report 40–50 surviving colonies on a plate, whereas in reality there were fewer than ten. Since the reduction in survival of colony-forming cells to 40--50% of control is not indicative of clinical response or substantial drug activity (Chapters 16–18), this limitation of simpler instruments drastically reduces their usefulness. Plating of increasing numbers of cells in serial dishes in control experiments did not detect this limitation, and it was only through extensive empirical studies with clinical samples from actual drug assays that we were able to conclude that the image-analysis capabilities of the FAS-II were essential to the enumeration process.

The tumor stem cell assay system is undergoing transition from a research procedure to a clinical procedure. In that context we are considering various options that can be retrofitted to the FAS-II to increase its speed and specimen throughput capability. For example, one logical way to optimize the scanning procedure is to introduce square or rectangular Petri dishes or sample wells having an area similar to that of 35 mm round Petri dishes. The flat area of a 31 × 31 mm square dish could be completely covered by thirty-six 3.3 × 4.4 mm fields. The compatibility of such Petri dish shape with the tumor stem cell assay would have to be, of course, experimentally confirmed. Another alternative is the use of multi-chamber

plates or groups of plates that could be readily handled with a large mechanical stage especially designed for that purpose. Further refinements in the reporting of assay results are also in progress.

The widespread application of automated tumor colony counting may well provide the basis for major advances in new drug screening as well as in clinical oncology. In both circumstances it is essential that standardized criteria for tumor colony counting be available that have identical meaning from laboratory to laboratory. As is underscored by the percent error in manual counts from experienced investigators (as shown by our data), even greater error could be anticipated in manual counting from laboratory to laboratory, since the criteria for manually counting colonies are not easy to teach, and the process of serial counting is subject to substantial problems of fatigue.

We view our efforts as having been successful in solving the major problems of automated tumor colony counting for the stem cell assay system and look forward to the application of this new technology to basic and clinical cancer research and patient management.

REFERENCES

1. Salmon SE, Hamburger AW, Soehnlen B, Durie BGM, Alberts DS, Moon TE: Quantitation of differential sensitivity of human-tumor stem cells to anti-cancer drugs. N Engl J Med 298:1321-1327, 1978.
2. Hamburger AW, Salmon SE: Primary bioassay of human tumor stem cells. Science 197: 461–463, 1977.
3. Hamburger AW, Salmon SE, Kim MB, Trent JM, Soehnlen BJ, Alberts DS, Schmidt HJ: Direct cloning of human ovarian carcinoma cells in agar. Cancer Res 38:3438–3444, 1978.
4. McCarthy CJ, Stevens RE: Dimensions in image analysis. Am Lab, vol 2, 1978.