Loughlin, P.M. and Cooke, T.G. and George, W.D. and Gray, A.J. and Stott, D.I. and Going, J.J. (2007) Quantifying tumour-infiltrating lymphocyte subsets: a practical immuno-histochemical method. *Journal of Immunological Methods 321*(1-2):pp. 32-40.

Quantification of tumour-infiltrating lymphocyte subsets: a practical immuno-histochemical method

Paula M. Loughlin [1], Timothy G Cooke [1], W. David George [1], Alison J. Gray [2], David I. Stott [3] and James J. Going [4]

[1] Section of Surgical Sciences and Translational Research, University of Glasgow; [2] Department of Statistics and Modelling Science, University of Strathclyde; [3] Division of Immunology, Infection and Inflammation and [4] Section of Gene Regulation and Mechanisms of Disease, University of Glasgow.

Running title Quantifying tumour-infiltrating lymphocytes

Address for correspondence

Dr James J Going, Department of Pathology, Glasgow Royal Infirmary, 84 Castle Street, Glasgow G4 0SF

Telephone   +141 211 4046

Facsimile   +141 211 4224

going@udcf.gla.ac.uk

Abstract

Background A need for efficient quantification of tumour-infiltrating T and B lymphocytes (TIL) arose in designing a case-control study of breast cancer outcomes.

Methods Ten adjacent x 40 digital images of 4μ sections immunostained for CD3, CD4, CD8, and CD20 were acquired for each of 16 randomly chosen carcinomas (640 images total). 'Color range' and 'Histogram' tools in Adobe Photoshop 7 were used to count the number of pixels in each image matching a partition of Lab colour representing the diaminobenzidine reaction product. This number was converted to a cell count per square mm calibrated from one, two, three or all 10 available images for each case and antibody.

Results Variation in the number of pixels per immunostained cell necessitated individual calibration for each antibody / case combination. Calibration based on two fields containing the most labelled pixels gave a cell count minimally higher (+5.3%) than the count based on 10-field calibration, with 95% confidence limits -14.7 to +25.3%. Median and interquartile range for cell counts per mm^2 were 6280 (490-14,200) for CD3; 5160 (1960-20,000) for CD4; 4280 (2997-9126) for CD8; and 1860 (54-10031) for CD20. As TIL density varied ~100-fold between different antibodies and cases, this accuracy and precision are adequate for the purpose.

Conclusion The methodology described offers sufficient accuracy, precision and efficiency to quantify the density of TIL sub-populations in breast cancer using commonly available software, and lends itself to batch processing.

Abbreviations

DAB	Diaminobenzidine
TIL	Tumour infiltrating lymphocytes
P_i	Number of labelled pixels in image i ($1 \leq i \leq 10$)
ΣP_i	Number of labelled pixels in a set of 10 images
p_i	Number of labelled pixels in a calibration area in image i
$p_m, p_{m'}, p_{m''}$	Number of labelled pixels in calibration areas of first three of a set of 10 images ranked by the number of labelled pixels
c_i	Number of cells present in a calibration area (counted) in image i
$c_m, c_{m'}, c_{m''}$	Number of labelled cells in calibration areas of first three of a set of 10 images ranked by the number of labelled pixels
C_i	Number of cells in image i (counted or estimated)
$\Sigma C_i,$	Number of cells in a set of 10 images (estimated)
F_n	Calibration factor calculated from n images (n = 1,2,3 or 10)
RGB	Colour space defined by 8 bit red/green/blue colour axes
Lab	Colour space defined by luminance (L), green/red and yellow/blue colour axes

Keywords

Breast cancer, Tumour infiltrating lymphocytes, Measurement, Colour image histogramming, Photoshop, Lab colour space, Semi-automated cell counting

Introduction

Cytotoxic T cells, complement-mediated cell lysis and antibody-directed cellular cytotoxicity all have the potential to destroy neoplastic cells. As metastatic progression is responsible for most cancer deaths, immunological killing of cancer cells shed into blood or lymph could be important, but the extent to which such mechanisms do actually control growth and dissemination of spontaneous human cancers remains unclear (1,2).

It is plausible, though, that the intensity and composition of the host immune response to a primary tumour could indicate the potential of immune surveillance to prevent metastatic cancer. To test this hypothesis requires measurement of tumour-infiltrating lymphocyte populations. In designing a retrospective case-control study of disease-free survival in breast cancer, we sought an efficient and objective method for measuring the density of lymphoid infiltrates in immunostained sections of formalin-fixed paraffin-embedded breast cancer tissue. Immunohistochemistry can identify different cell populations (T cell subsets, B cells etc) but qualitative or semi-quantitative estimates of lymphocyte subsets may be poorly reproducible. We wished to employ a quantitative approach but manual counting is prohibitively laborious. This paper describes approaches to the automation of this task using digital images.

What do we actually wish to measure? As cells are discrete, countable entities then an obvious measurement would be cells per mm^3. However, this measurement requires 3D stereological tools such as the 'optical disector' or 'unbiased brick' which may not be practicable for a large project, and may have their own biases (3,4). Histological sections, being quasi-two dimensional, suggest that a more accessible measure might be cells per mm^2. Other possibilities include the volume fraction V_v occupied by immune cells, which, by the principle of Delesse (5), could be estimated from the area fraction A_a of an infinitesimally thin section (approximated in reality by a histological section of conventional thickness).

Point counting estimates the area fraction and therefore also the volume fraction, but if the volume fraction is small, the number of points which must be counted to estimate it with a specified degree of precision is relatively large. A subjective element may also influence the decision as to whether a sampling point falls on the object to be measured.

Binary thresholding of a digital image resembles 'point counting'. Every pixel is treated as a sampling point, and its colour properties determine whether it is to be counted. This sounds easy but determining appropriate thresholds is not straightforward. For an immunoperoxidase signal visualised with diaminobenzidine and counterstained with haematoxylin, the problem is to decide which pixels are 'brown' enough to represent signal to be counted and which are either blue (counterstained nuclei) or unstained, unlabelled background. A variety of approaches have been proposed, from simple thresholding of colour channels to complex image deconvolutions (6-8).

The approach adopted was based on the idea that relevant pixels would resemble each other in colour, and would therefore be grouped in an appropriate 3D colour space (9). In 24-bit RGB colour space 3 separate 8-bit (one byte) numbers plotted on mutually perpendicular red, green and blue colour axes define a colour cube composed of 256 x 256 x 256 individual elements specified by each possible RGB number triple. In Lab colour space, luminance (L) is plotted against mutually perpendicular chromaticity axes (green/red, a and yellow/blue, b). Proximity of points within a perceptually uniform colour space(10) such as Lab space implies similarity of colour, and it appeared therefore that this proximity could be used to identify pixels representing the signal to be counted. We also wished to employ readily available software, so to begin with tools provided in Adobe Photoshop 7 were used.

Materials and methods

Cases and immunostaining. Sixteen invasive breast cancers were chosen at random from archives at Glasgow Royal Infirmary. For each case four micron sections of one representative block were immunostained as a single batch using a standard immunoperoxidase methodology and primary antibodies against CD3, CD4, CD8, CD20 and CD35. [Antibody, procedure details required] For each immunostained slide ten consecutive adjacent digital images of fields located at the infiltrative edge of the carcinoma were acquired using a Fuji HC300Z digital camera and a Nikon Eclipse E600 microscope with a x40 apochromatic objective, and saved as uncompressed 24-bit RGB TIFF files. Field size (measured by stage graticule) was 218 x 170 microns = $0.03706 \text{ mm}^2 = 1/26.98 \text{ mm}^2 \approx 1/27 \text{ mm}^2$, so cell counts multiplied by 27 equal cell counts per mm^2.

Measuring the immunostaining signal. Adobe Photoshop (version 7) allows the colour of any individual pixel to be sampled and set as the 'foreground' colour. The 'Select/Colour Range' tools allow all pixels in the image having a colour similar to the foreground colour to be selected, which can then be counted using the 'Histogram' tool. How closely colours in the image must resemble the foreground colour is determined by setting a property known as 'fuzziness': a low value ensures that only pixels closely similar in colour will be selected, while a higher value broadens the selected colour range, making it possible to count the number of image pixels of a particular colour or colour range, narrowly or broadly defined.

Even if the open image file is in RGB mode, the Photoshop 'Colour Range' tool operates in Lab colour space. This can be demonstrated using the 'LabMeter' colour measurement software tool, available as a free download (www.curvemeister.com). This provides a square image representing all values on the green-red (a) and yellow-blue (b) chromaticity axes of Lab colour space (i.e. the a,b colour plane), at a user-specified luminance value. Setting the foreground colour in Photoshop to RGB 160, 67, 23 (Lab 41,38,44) and 'fuzziness' = 100, the 'Select/Colour Range' command chooses a square portion of the LabMeter image for which a = 38±15 and b = 44±15, inclusive. Varying the luminance (L value) of the test image or a gradient image showed that a similar but slightly wider selection obtains on this axis, with L = 41±19 inclusive being chosen. Lab 41,38,44 is a brown corresponding to DAB staining of medium intensity. Using fuzziness=100 includes weaker and stronger staining in the selected area. This value and range were chosen to select approximately 90% of the area of the section in which the DAB signal could be identified visually. One hundred per cent selection was not sought, to avoid identifying areas of the section lacking specific staining. Figure 1 illustrates an example of pixel selection in Photoshop with these parameters for an image of an immunostained section containing many labelled cells, and one containing none. In all cases there was visual control of the selected area. It is also instructive to examine a 3D histogram of the pixel distributions of these images in Lab colour space (Figure 2). The stained section includes pixels corresponding to the DAB signal while the image without labelled cells shows pixels corresponding to the background and nuclei only. In essence, we are counting a representative subset of these pixels.

Associating the measured signal with cell counts. The next step was to examine the relationship between the number of cells present in an image and signal strength measured by the number of pixels falling into a particular Lab colour range. A total of 16 cases x 5 antibodies x 10 images = 800 images were collected on the same microscope, camera, and light setting, of slides stained in a single batch on an automated staining machine. These precautions were observed to maximise consistency of analysis without extraordinary measures. CD35 positive cells were present in significant numbers in very few of these images, and images of CD35 staining were excluded from subsequent analysis.

Each image was opened in Adobe Photoshop 7 with foreground colour set to R = 160, G = 67, B = 23 and 'fuzziness' set to 100. The 'Select/Colour Range' tool was used to select the labelled pixels; these highlighted on screen so that labelled cells can be identified. The number of selected pixels was recorded from the Image/Histogram dialogue. The file was then closed and the process repeated file by file until all files had been processed.

If the number of labelled cells in an image was not too great, all were counted to calibrate for that particular image the relationship (labelled pixels per cell) between total signal and cell number. If there were too many cells to be counted easily, the 'Rectangular Marquee' or 'Lasso' tool were used to define a representative subregion of the image within which all labelled cells could be counted and within which colour selection and pixel counting allowed a calibration (labelled pixels per cell) value to be calculated for that particular image. A complete field or field subset cell count with the corresponding pixel count was made for every image. To be certain that labelled cells were neither missed nor counted twice, the Eraser tool was used to place a spot of colour on each labelled cell as it was counted. This was quick and efficient.

The expectation was that the number of labelled pixels, for a particular combination of case and antibody, would be proportional to the number of cells present in a field; and that once the system had been calibrated, it would not be necessary to count individual cells, but only the pixels using the semi-automated methodology described above. Particular interest was attached to variations in calibration from field to field, case to case and antibody to antibody. Batch-to-batch variation was not addressed, because if case-to-case variation in the relationship between pixel counts and cell

number is significant, then calibration will have to be undertaken in every case, which will control for batch-to-batch variation also.

For any individual field i the data available are the number, P_i, of labelled pixels in the whole image; the number, p_i, of labelled pixels and the number, c_i, of cells present in the calibration area. The estimated number of cells in a field is $C_i = P_i/(p_i/c_i)$; in fields containing few enough cells to count them all, the formula is $C_i = P_i / (P_i / C_i)$, which cancels to $C_i = C_i$, as expected. The best estimate of the number of cells, ΣC_i, in all 10 fields is

$$\Sigma C_i = P_1(p_1/c_1)+...+ P_i(p_i/c_i)+...+ P_{10}(p_{10}/c_{10}),$$

and from this figure an estimate of the average calibration factor, F_{10}, weighted in proportion to the number of cells present in individual fields can be derived as

$$F_{10} = \Sigma P_i/\Sigma C_i.$$

We expected that it would be necessary to calibrate each case and antibody combination individually, on account of differences in fixation and processing between cases, and differing epitope robustness and antibody binding affinity. But is calibration for every *field* necessary, given that all ten fields were adjacent to each other on the same section, and had been exposed to identical handling, prosection, fixation, processing, storage and staining? We sought, therefore, to find the maximum number of fields which would have to be calibrated by cell counting to allow an acceptable estimate of the number of cells present, in comparison with the number of cells estimated by calibrating *every* field.

Our data allow us to examine variations in the pixels/cell calibration factor on a field-to field, case-by case and antibody-by antibody basis. Plotting all individual p_i/c_i measurements against P_i for all cases, antibodies and fields (CD3, CD4, CD8, CD20) allows one to see relationships between the calibration factors and the total labelling (figure 3). Where few labelled pixels are present in a field, there is wide variation in the calibration, but the estimates of the calibration factor derived from fields in which more labelled pixels are present fall within a narrower range. It appeared appropriate therefore to base the calibration on the fields containing the largest number of labelled pixels, to reduce noise associated with smaller cell and pixel counts.

Accordingly, three different estimates were made for each case/antibody combination: the calibration factor (over all 10 fields) was calculated as described

above (F_{10}); and estimated using calibration factors calculated from the field *m* containing the greatest number of labelled pixels as $F_1 = (p_m/c_m)$; from the sum of that field and the field *m'* with the next largest number of labelled pixels as $F_2 = (p_m+p_{m'})/(c_m+c_{m'})$ and from the sum of the three fields with the largest, second and third largest number of pixels as $F_3 = (p_m+p_{m'}+p_{m''})/(c_m+c_{m'}+c_{m''})$.

In expressing actual cell counts, the number of cells per field has been multiplied by 27 to give the results in cells per mm².

Results

Comparison of F_{10} **against** F_1, F_2 **and** F_3. A scatter plot (figure 4) shows a greater degree of scatter for F_1 against F_{10} than for F_2 or F_3 against F_{10}, as might have been expected. We can look more closely at the degree of agreement (following Bland and Altman (11)) by computing $100(F_1 - F_{10})/0.5(F_1+F_{10})$, $100(F_2-F_{10})/0.5(F_2+F_{10})$ and $100(F_3 - F_{10})/0.5(F_3+F_{10})$ to express the degree of agreement between the different calibration factors as a percentage of their means. The mean difference and its standard deviation are: F_1 v F_{10}, 8.3% (17.4%); F_2 v F_{10}, 5.3% (10.2%) and F_3 v F_{10}, 4.4 % (7.8%). Clearly, F_3 agrees best with F_{10}, but F_2 is nearly as good and is less work to derive, requiring only two calibration measurements.

The differences are normally distributed (judged by normal probability plots and Shapiro-Wilk W test), so these figures tell us that we can be 95% confident that a cell count using F_1 will not be more than 42.5% greater and not more than 25.9% less than a cell count derived using F_{10}. The 95% confidence limits for F_2 are +25.3% and -14.7% and for F_3 are +19.6% and -10.9%. For many purposes these will be offer adequate accuracy and precision.

Differences in calibration between antibodies and between cases. This section loooks at whether it is necessary to calibrate cell counting for different cases or different antibodies. Figure 5 plots the raw calibration data for CD3, CD4, CD8 and CD20. The scatter of points is very similar for CD3 and CD8; for CD4 it tends to be lower, and higher for CD20. Another way of looking at this is to take mean F_{10} values for all cases for each antibody. For CD3 the mean F_{10} is 770 pixels/cell ± SEM 61; for CD4 it is 603 ± 53; for CD8 it is 896 ± 67; and for CD20, 1161 ± 103. Were one to take the mean of these values (857.5 pixels/cell) to represent them all, the number of CD20+ B cells and CD8+ T cells would be systematically overestimated by 35% and 4.5% and CD3

and CD4+ T cells would be underestimated by 10.2 % and 29.3%. These represent non-trivial biases which for many purposes would not be acceptable.

Figure 6 looks at case to case variation. It plots normalised factors F_{10} for CD3, CD4, CD8 and CD20 and shows that that generally the calibration factors F_{10} lie in a range between about 75% and 140% of the average for the series. One case is clearly an outlier and may have been subjected to unusually lengthy fixation. Again, the differences in the relationship between the pixel counts and the cell counts which they imply suggest that this variation must be taken into account.

Finally, it is worth taking a preliminary look at the range of densities of different TIL populations in breast cancer. This gives some indication of the kind of precision which must be achieved to detect biologically significant differences; if the range is very wide then greater precision may not be required in comparison to the situation which would obtain if the range of observed densities was small.

Figure 7 and table 1 present this data in units of cells per square millimetre.

	CD3	CD4	CD8	CD20
Mean	12,900	6,016	7,010	4,750
SD	15,700	5,570	6,600	5,850
Median	6,280	5,160	4,280	1,860
Total Range	490-59,700	54-20,000	270-22,800	0-16,092
IQ range	3,969-14,158	1,958-8,330	2,997-9,126	54-10,031

Table 1

Cell count summary table

The range of TIL densities is very wide; there are major differences between different cases in the intensity of the TIL response evoked, in keeping with subjective impressions of the situation. This is a pre-condition for significant differences in tumour behaviour to be related to TIL density, and makes the task of analysis easier.

Discussion

Immunostaining pixel counts made using a widely available tool (Adobe Photoshop) can be converted into cell counts per square mm for tumour infiltrating lymphocyte

subtypes. Likely errors associated with this conversion are small in comparison with the range of TIL densities in breast carcinomas in this preliminary study.

A purpose of this study was to develop a methodology employable in larger-scale studies. Many studies published in the pathology literature are statistically underpowered. Our ongoing case/control study of outcomes in breast cancer is designed to examine 111 breast cancer patients ('cases') in which metastatic relapse occurred and 222 cancer patients ('controls') without relapse. Carcinomas are matched for size, grade, estrogen receptor and lymph node status, and the women are matched by age as a surrogate for menopausal status. Examining a core set of TIL subtypes (CD3, CD4, CD8 and CD20) will require measurement of 333 x 4 x 10 = 13320 digital images. High-throughput approaches to the analysis of ≈ 40 gigabytes of image data are required. Manual processing of all images in Photoshop is a non-starter. We view this enabling, preliminary study as a stepping stone to a more streamlined approach.

Calibration of conversion from pixel to cell counts is required for each case/antibody combination. For a set of histological sections immunostained as a batch on automated staining equipment, identical colour selection parameters can be used on all sections in that batch; visual inspection showed satisfactory selection of areas identified visually as immunostained with DAB from field to field, case to case and antibody to antibody, all in the same batch.

We studied 640 images in this pilot study. Pixel counts were made manually for all these images. We plan to automate this step and have begun work on the development of a program to do so, using batch processing of image files. The intention is to write the program in a programming language (C) for which compilers are readily available for different computer platforms (Windows, Apple, Unix/Linux...). We anticipate that the public domain image processing software 'ImageMagick' (available at http://www.imagemagick.org) will be called by the program to convert different file types into a form suitable for the necessary image arithmetic for counting image pixels which meet the user-specified criteria for the signal.

These criteria may be established within Photoshop. Photoshop does not offer a tool revealing the Lab ranges implied by specific 'Colour Range' and 'Fuzziness' settings; indeed there is no published algorithm detailing precisely how Photoshop performs

the image arithmetic behind these tools. Howevcr, these parameters can be established using the LabMeter tool (free from www.Curvemeister.com) to measure the a,b range and a suitable gradient image to define the L range being quantified. Other proprietary and public domain image analysis programs may also allow appropriate Lab parameters to be defined.

Even in its present form the method could be used for studies in a small to medium scale. We do not think it will be possible to avoid the need to calibrate images manually, although histogram specification may be worth exploring to see how it performs in this context. This transforms each colour plane in a new image to have the same colour histogram as a reference image. This may be most relevant for staining/batch to batch variation.

For some purposes calibration based on a single image from a set may define a conversion factor with adequate precision to be satisfactory in a particular application. For the application we have described we consider that two fields are sufficient. This represents an 80% reduction in the labour of counting, compared to 10-field calibration. This brings the method within the range of what is practicable and will facilitate the use of objective histological cell counting in clinical and experimental tumour immunology.
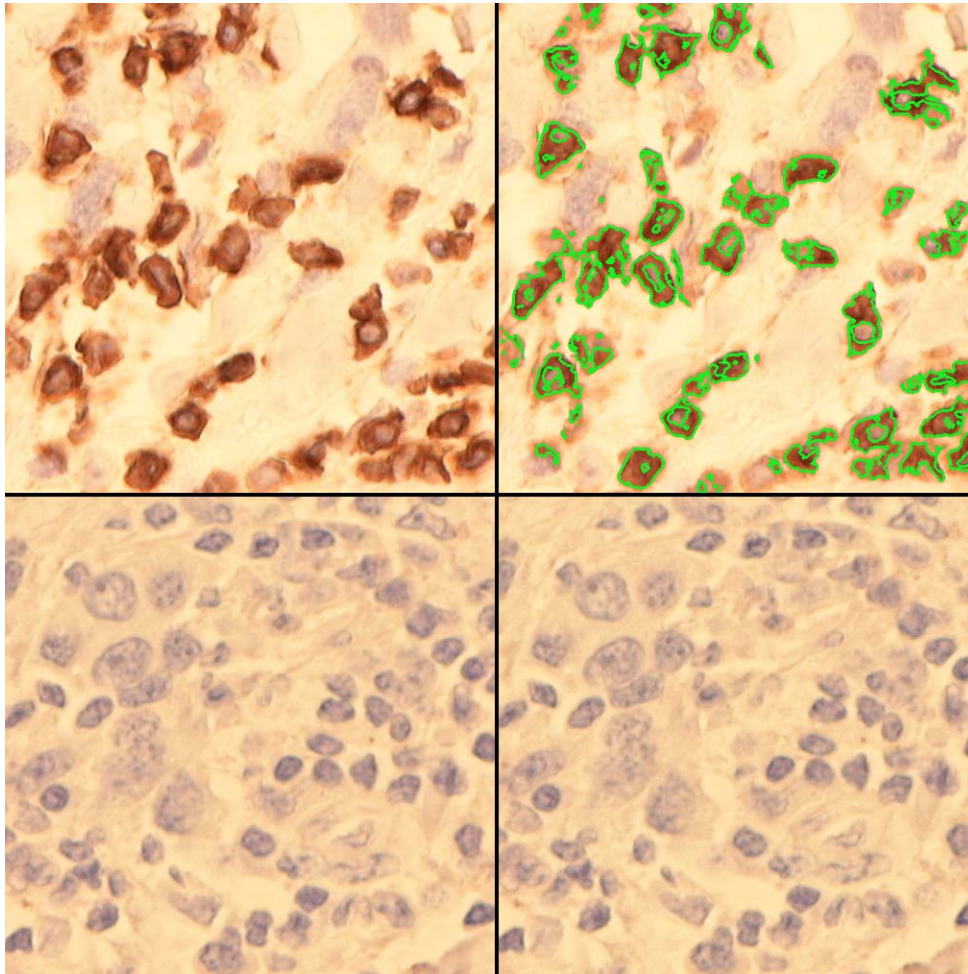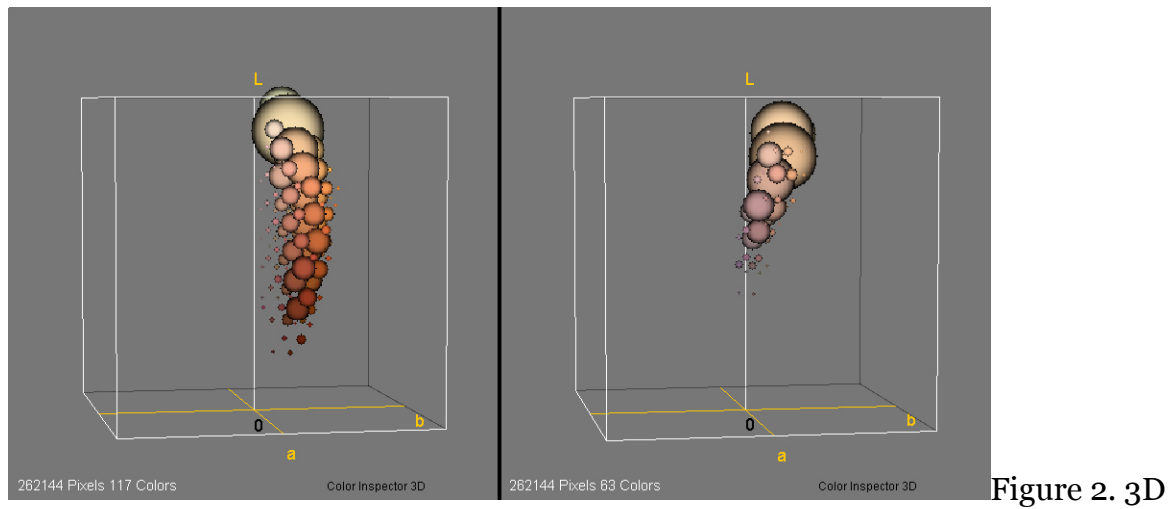
Titles and legends to figures

Figure 1. Top left: Immunoperoxidase staining shows many CD3+ tumour-infiltrating lymphocytes in this breast carcinoma. Top right: outlined areas selected in Photoshop using 'Select/Colour Range' as described in materials and methods. Bottom left and right: No CD35+ cells are present in this field (left), and no pixels are selected by Photoshop (right)
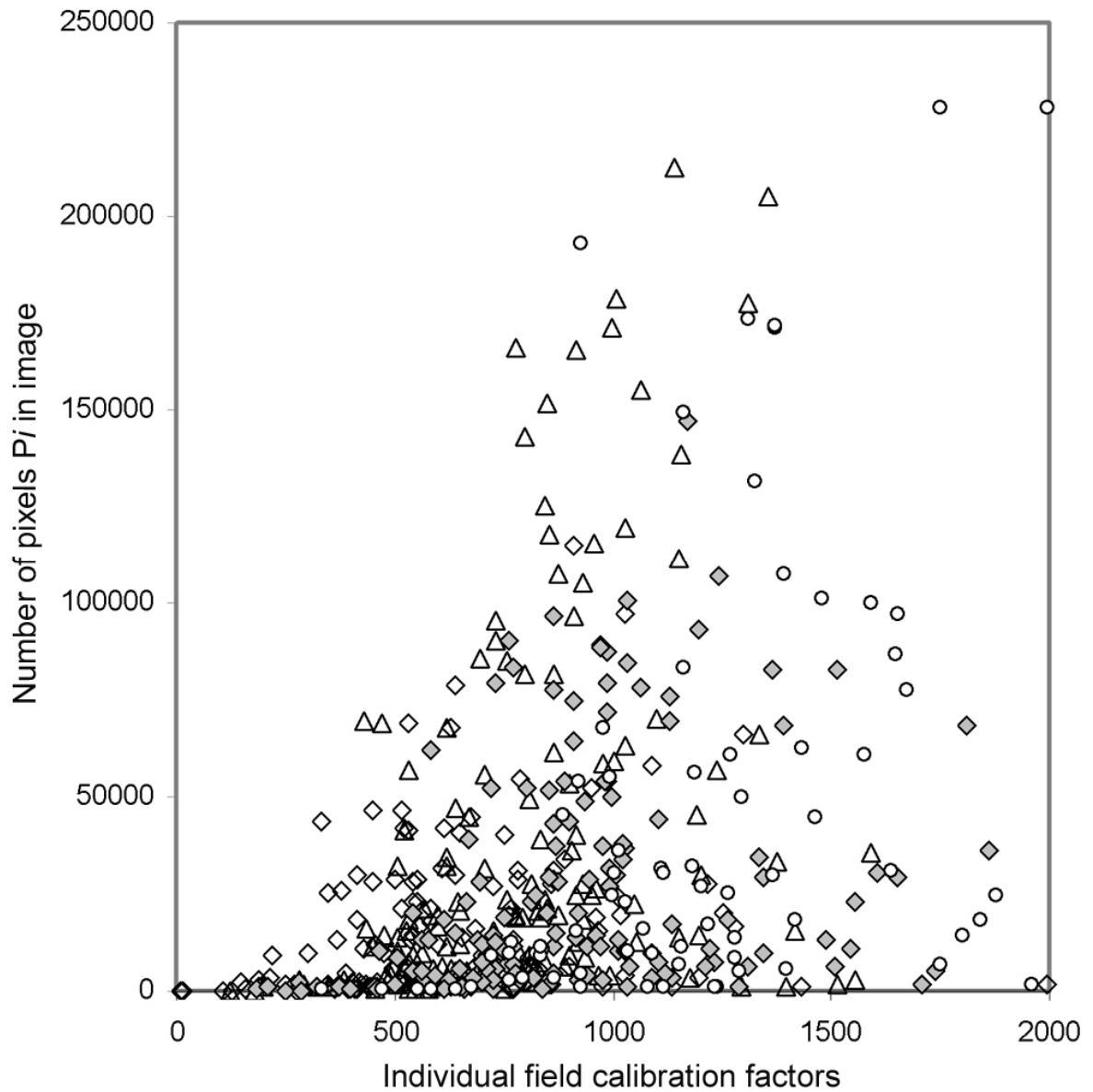
Figure 2. 3D histograms of the images in figure 1. Left: CD3. Right: CD35. The difference in the colour distribution of coloured pixels between the two images as a consequence of the numerous CD3+ cells and the absence of CD35+ cells is obvious. These 3D histograms were generated using the '3D Color Inspector/Color Histogram' plugin [Kai Uwe Barthel] for the public-domain image processing software Image J (available at http://rsb.info.nih.gov/ij/). Luminance is on the vertical axis and the chromaticity axes a and b are indicated on the base of the Lab colour cube.

Figure 3. Scatter plot of calibration factors p_i/c_i for all evaluable fields. The calibration factor is plotted on the x axis against the number of labelled pixels P_i in the image which yielded that calibration factor for CD3 (triangle), CD4 (diamond), CD8 (grey diamond) and CD20 (circle).
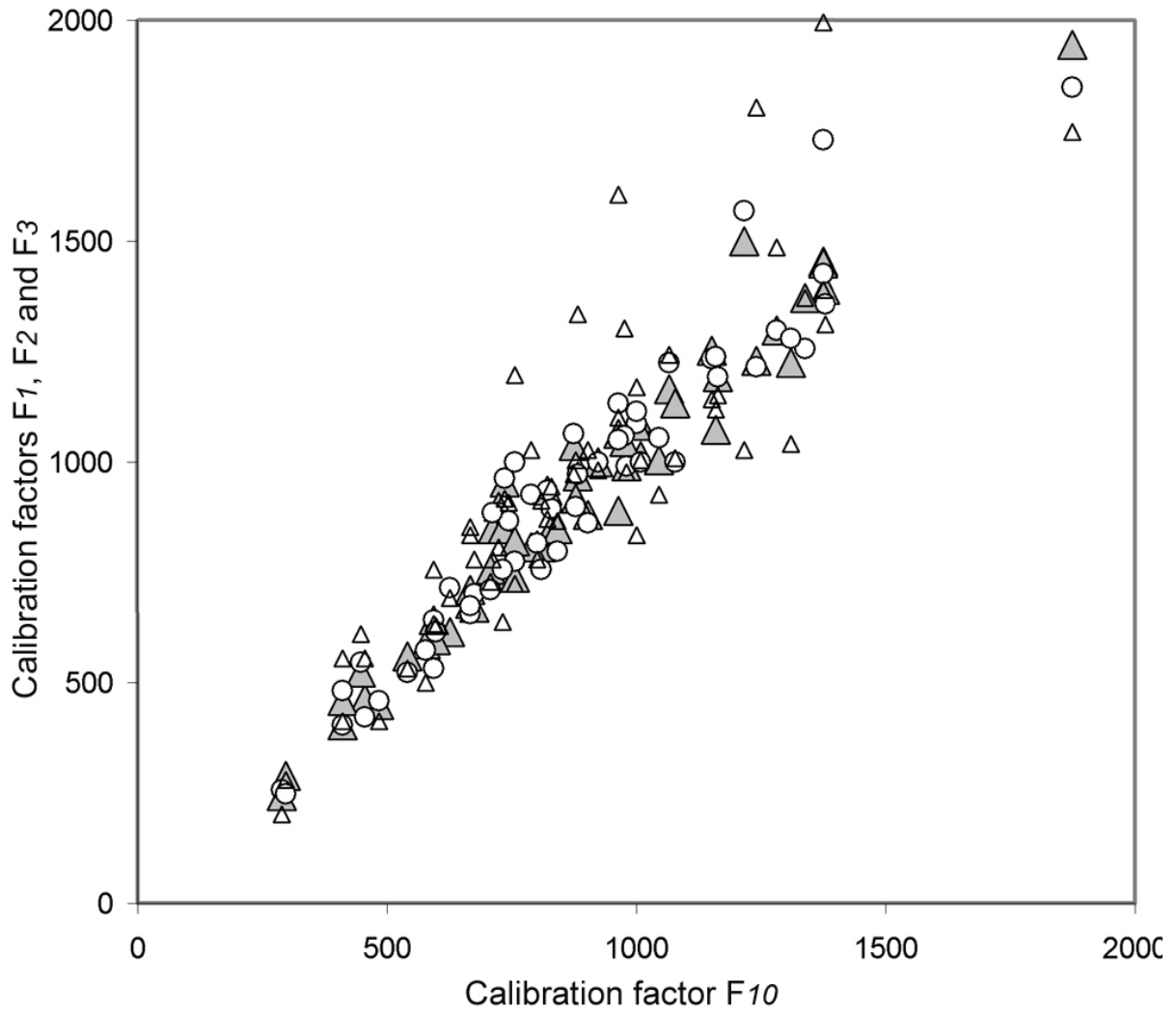
Figure 4. Scatter plot of calibration factors (pixels per cell) derived from all evaluable fields. The calibration factor F_{10} is plotted on the x axis against the three calibration factors derived from available measurements for the single field with the most labelled pixels (F_1; small $\triangle$), the two fields with the most and second most labelled pixels (F_2; o), and the three fields with the three most labelled pixels (F_3; larger grey $\triangle$). One-field calibration shows considerably more scatter than two- or three-field calibration.
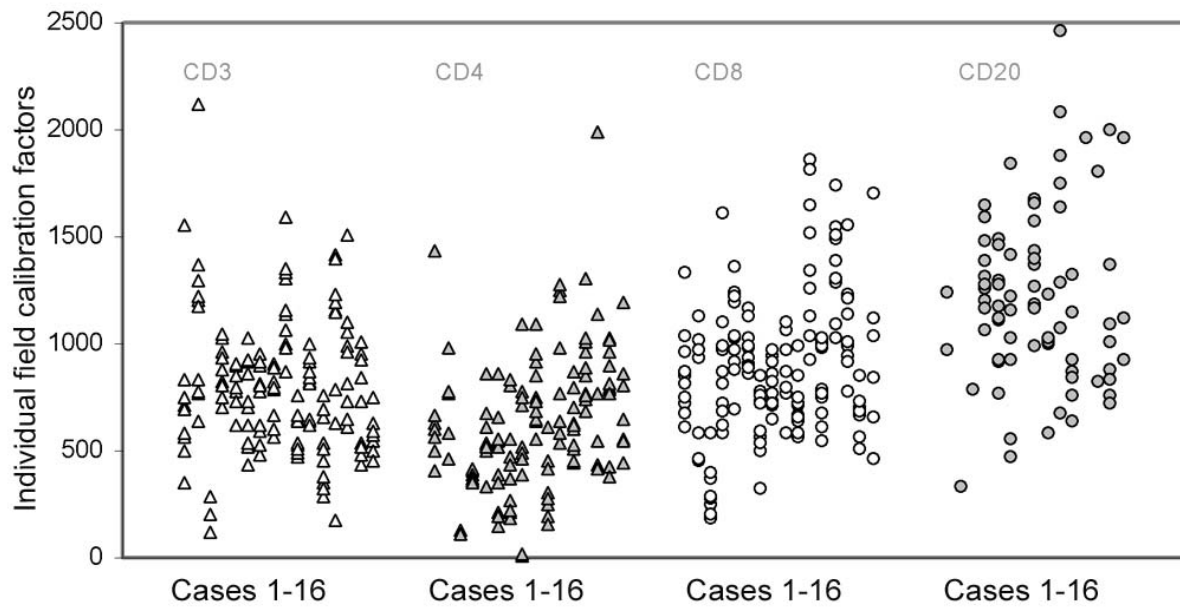
Figure 5. Scatter plot of calibration factors (pixels per cell) derived from all evaluable fields arranged by case and by antibody. Each vertical column of data points represents one case from 1 to 16 and the order of cases is the same for each antibody. Calibration factors are comparable for CD3 and CD8; those for CD4 tend to be lower and for CD 20, higher.
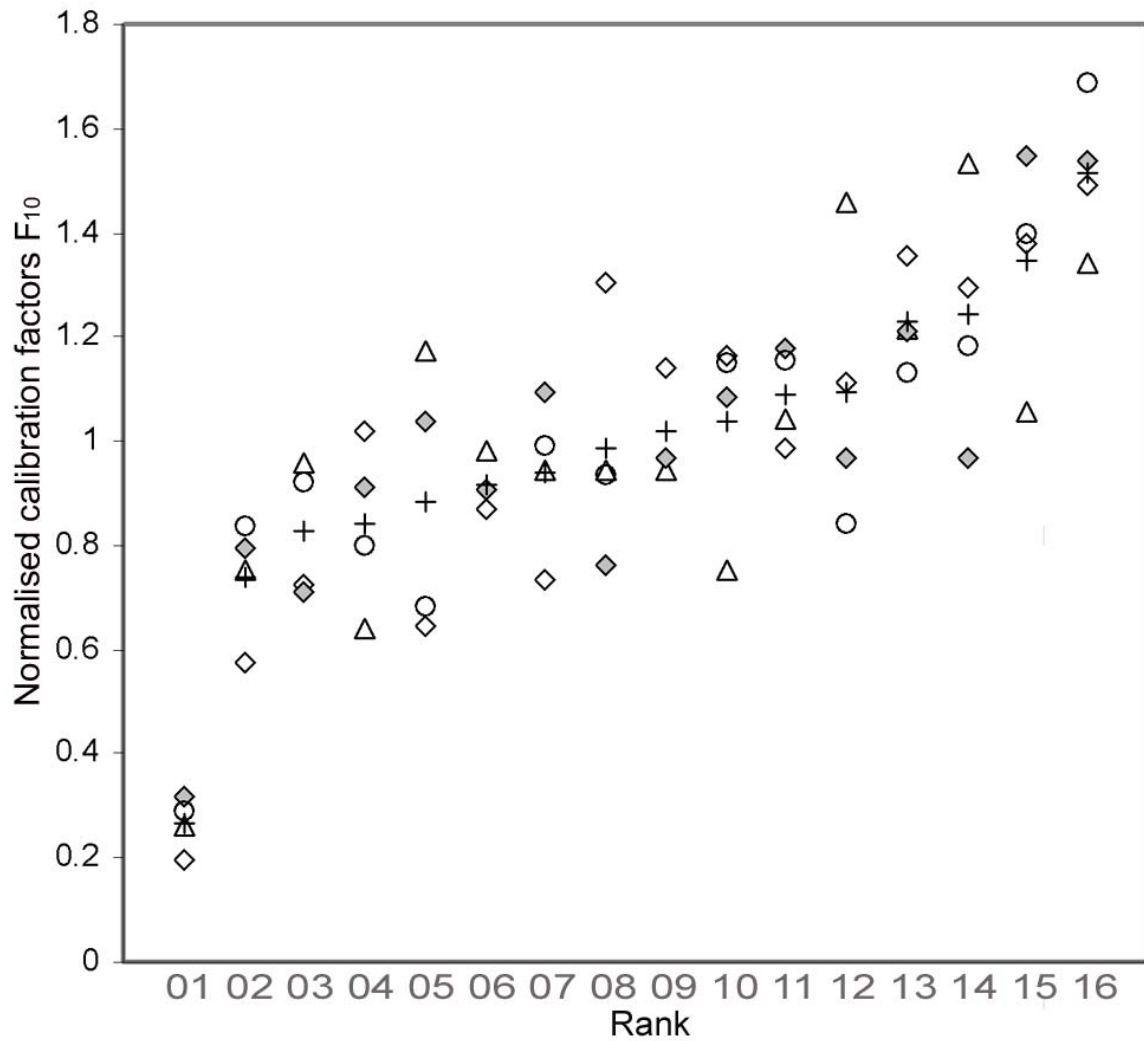
Figure 6. Scatter plot of calibration factors F_{10} for CD3, CD4, CD8 and CD20: CD3 (triangle), CD4 (diamond), CD8 (grey diamond) and CD20 (circle). Data are normalised against mean F_{10} and ranked by mean normalised F_{10} value. This plot indicates that for the most part calibration factors F_{10} lie in a range between about 75% and 140% of the average. The first case is clearly an outlier and may have been subjected to unusually lengthy fixation. '+' indicates mean normalised F_{10} values.
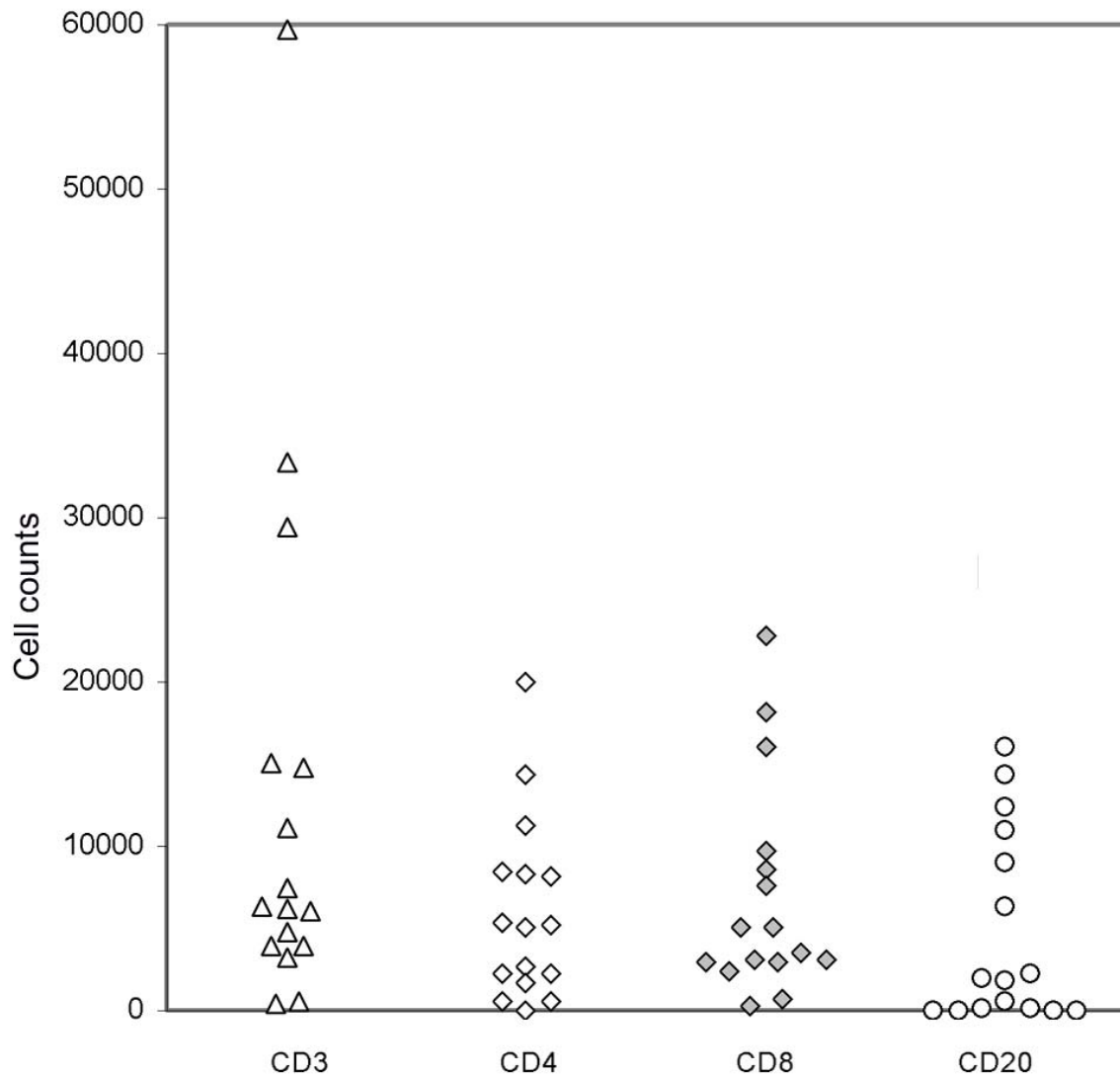
Figure 7. Scatter plot of cell counts for CD3, CD4, CD8 and CD20: CD3 (triangle), CD4 (diamond), CD8 (grey diamond) and CD20 (circle). There is a wide range of values for each lymphocyte sub-population.

References

1. Mufson RA. Tumor antigen targets and tumor immunotherapy. Front Biosci 2006;11:337-43.
2. Willimsky G, Blankenstein T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. Nature 2005;437(7055):141-6.
3. Guillery RW, August BK. Doubt and certainty in counting. Prog Brain Res 2002;135:25-42.
4. von Bartheld C. Counting particles in tissue sections: choices of methods and importance of calibration to minimize biases. Histol Histopathol 2002;17(2):639-48.
5. Royet JP. Stereology: a method for analyzing images. Prog Neurobiol 1991;37(5):433-74.
6. Brey EM, Lalani Z, Johnston C, Wong M, McIntire LV, Duke PJ, Patrick CW, Jr. Automated selection of DAB-labeled tissue for immunohistochemical quantification. J Histochem Cytochem 2003;51(5):575-84.
7. Matkowskyj KA, Schonfeld D, Benya RV. Quantitative immunohistochemistry by measuring cumulative signal strength using commercially available software photoshop and matlab. J Histochem Cytochem 2000;48(2):303-12.
8. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 2001;23(4):291-9.
9. Poynton C. A guided tour of color space. 1995; San Francisco. p 167-180.
10. Sangwine SJ,Horne REN (Eds) . The colour image processing handbook. London: Chapman and Hall; 1998.
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-10.