Methodology article

A new spreadsheet method for the analysis of bivariate flow cytometric data

George Tzircotis, Rick F Thorne and Clare M Isacke*

Address: Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

Email: George Tzircotis - george.tzircotis@icr.ac.uk; Rick F Thorne - rick.thorne@icr.ac.uk; Clare M Isacke* - clare.isacke@icr.ac.uk

* Corresponding author

Abstract

Background: A useful application of flow cytometry is the investigation of cell receptor-ligand interactions. However such analyses are often compromised due to problems interpreting changes in ligand binding where the receptor expression is not constant. Commonly, problems are encountered due to cell treatments resulting in altered receptor expression levels, or when cell lines expressing a transfected receptor with variable expression are being compared. To overcome this limitation we have developed a Microsoft Excel spreadsheet that aims to automatically and effectively simplify flow cytometric data and perform statistical tests in order to provide a clearer graphical representation of results.

Results: To demonstrate the use and advantages of this new spreadsheet method we have investigated the binding of the transmembrane adhesion receptor CD44 to its ligand hyaluronan. In the first example, phorbol ester treatment of cells results in both increased CD44 expression and increased hyaluronan binding. By applying the spreadsheet method we effectively demonstrate that this increased ligand binding results from receptor activation. In the second example we have compared AKR1 cells transfected either with wild type CD44 (WT CD44) or a mutant with a truncated cytoplasmic domain (CD44-T). These two populations do not have equivalent receptor expression levels but by using the spreadsheet method hyaluronan binding could be compared without the need to generate single cell clones or FACS sorting the cells for matching CD44 expression. By this method it was demonstrated that hyaluronan binding requires a threshold expression of CD44 and that this threshold is higher for CD44-T. However, at high CD44-T expression, binding was equivalent to WT CD44 indicating that the cytoplasmic domain has a role in presenting the receptor at the cell surface in a form required for efficient hyaluronan binding rather than modulating receptor activity.

Conclusion: Using the attached spreadsheets and instructions, a simple post-acquisition method for analysing bivariate flow cytometry data is provided. This method constitutes a straightforward improvement over the standard graphical output of flow cytometric data and has the significant advantage that ligand binding can be compared between cell populations irrespective of receptor expression levels.

Background

The investigation of receptor-ligand interactions by flow cytometry is a technique commonly employed in immunology and cell biology primarily due to the ability to rapidly analyse populations of cells. This, however, results in the generation of large data sets, the further analysis of which is inherently problematic. With existing software, alterations in ligand binding in response to stimuli or as a result of receptor manipulation are difficult to dissect. Particularly problematic is the comparison of different transfected cell populations, which frequently have variable protein expression, or when treatment of cells causes a shift in receptor expression. To date two main approaches have been taken to overcome these issues. First, different populations of cells can be matched for receptor expression levels either by fluorescence activated cell sorting (FACS) (e.g. [1]) or by selecting single cell clones (e.g. [2]). The main disadvantage of this approach is that expression levels in the different populations/clones have to be constantly monitored. This can become costly in terms of FACS usage, tissue culture expenses and time, and impractical when dealing with multiple transfectants especially if multiple clones for each transfectant have to be maintained. The second approach has been to post-analyse flow cytometric data. For this, a series of cell subpopulations are assigned based on the level of receptor expression to a set of fluorescence channel ranges (e.g. [3,4]). The corresponding mean fluorescence intensity for ligand binding is then calculated allowing the data set to be presented as a line graph of receptor expression versus ligand binding. This method has the advantage of allowing receptor:ligand interactions to be studied over a wide range of receptor expression levels.

Consequently, binding of ligand to different transfected cell populations can be compared. The main problem is that the method of data analysis is entirely manual and therefore dividing the population into a large series of data points becomes unmanageable. Building upon this concept, we have developed an automated spreadsheet-based method to post-analyse flow cytometry data. Using commonly available computer software, this spreadsheet enables the analysis of two-colour flow cytometric data by calculating the average fluorescence intensity value of the variable parameter for all cells lying within a single fluorescence channel of a constant parameter. This provides the correlation of data at the highest level of accuracy. To demonstrate the use and advantages of this new method, two worked examples of the interaction of the adhesion receptor CD44 with its ligand hyaluronan are reported here.

Results and discussion

CD44 is a transmembrane adhesion receptor and part of the hyaladherin protein family whose common ligand is the extracellular glycosaminoglycan hyaluronan [5,6]. Two-colour flow cytometry has been widely used to characterise this receptor-ligand interaction using fluorescein isothiocyanate (FITC) conjugated hyaluronan and anti-CD44 antibodies, either directly conjugated or detected with a second layer antibody. By this approach, the binding capacity of CD44 mutants or the activation of hyaluronan-binding activity following various treatments has been investigated. However, as FITC-hyaluronan binding is strictly dependent on CD44 expression levels, the analysis is compromised where expression levels are not matched. As described in the background, two main approaches have been taken to overcome this problem. Cells have been FACS sorted or single cell cloned to generate starting populations with equivalent levels of CD44 expression [1,2]. Alternatively flow cytometry data has been analysed manually to assess levels of hyaluronan binding relative to receptor expression [3,4]. The following examples demonstrate how the spreadsheet method can be used to overcome these problems.

Example 1: Monitoring receptor:ligand interactions after cellular treatment

The mouse T-cell lymphoma cell line BW5147 expresses CD44 and constitutively binds hyaluronan. This binding is known to be increased by long-term phorbol myristate acetate (PMA) treatment [7]. However it has been difficult to determine whether this increase in hyaluronan binding results from an increased binding activity of CD44, that is receptor activation, or increased CD44 expression. To assess whether the spreadsheet method could resolve this issue the following experiment was undertaken.

BW5147 cells were left untreated or treated with 100 ng/ml PMA for 1 h (short-term treatment) or 48 h (long-term treatment). The cells were incubated with FITC-hyaluronan followed by biotinylated anti-CD44 mAb IM7 and PE-streptavidin. Cells were then stained with membrane impermeable DNA binding dye TO-PRO-3 and subjected to three-colour flow cytometry. TO-PRO-3 is a widely used viability dye and only cells which exhibited low TO-PRO-3 fluorescence were used in the analysis. Histogram analysis of the data (Fig. 1A) or displaying the data as two dimensional dot-plots (Fig. 1B) shows that CD44 levels and FITC-hyaluronan binding increased after 48 h PMA treatment but not after 1 h treatment.

The flow cytometry data was exported and analysed using the spreadsheet method as described in the Methods section. Briefly, the mean FITC-hyaluronan fluorescence intensity was calculated for each of the 1024 PE-CD44 channels and the resulting points were plotted (Fig. 1C). It was empirically determined that more than 4 cells were required to provide an adequate average fluorescence intensity value for any particular channel and therefore
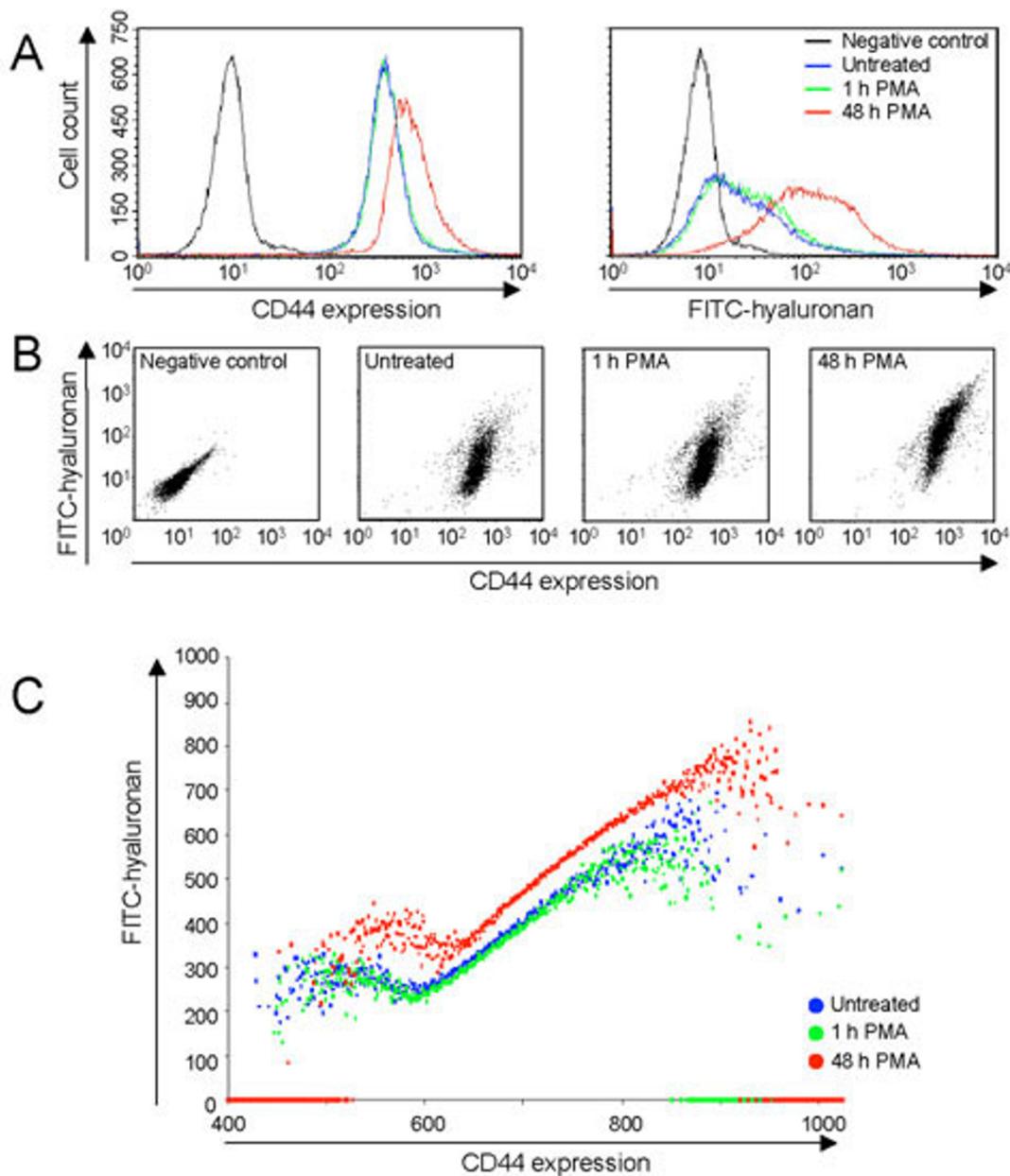
Figure 1
Analysis using the spreadsheet method of the effects of PMA treatment on hyaluronan binding to CD44
BW5147 cells were left untreated or treated with 160 nM PMA for 1 h or 48 h and then stained for three-colour flow cytometry as described in the Methods. For the negative control, untreated cells were stained with FITC-BSA followed by PE-streptavidin. The flow cytometry data is displayed as follows: **A**, histogram plots showing CD44 expression or FITC-hyaluronan binding for each treatment. **B**, two-parameter dot plots correlating FITC-hyaluronan binding against CD44 expression. **C**, the data was analysed using the spreadsheet method (see text for further details). The values given are arbitrary fluorescence units. Where there were 3 cells or fewer for a given PE fluorescence channel, the point is plotted at 0 on the y-axis.

channels with 3 cells or fewer were rejected from the analysis and are shown as a 0 value on the y axis. The plots clearly demonstrate that across the entire range of CD44 expression, no appreciable differences were observed between untreated and 1 h PMA treated cells. In contrast, long term PMA treatment results in increased FITC-hyaluronan binding relative to untreated cells at all CD44 expression levels. Therefore it can be concluded that long-term PMA treatment results in an activation of CD44 which enhances its binding capacity.

Given that this increased receptor activity is only observed after long-term PMA treatment it is likely that this reflects the induction by PMA of a newly synthesised modified CD44 population with altered binding properties. To date, the best characterized post-translational modification of CD44 which might result in altered ligand binding is a change in receptor glycosylation [8].

Example 2: Comparing receptor:ligand interactions in two transfected cell populations

The CD44 negative murine T-lymphoma cell line AKR1 has commonly been used as a transfection model to study CD44 function [9]. Expression of human or mouse wild type CD44 (WT CD44) in these cells confers to them the ability to bind hyaluronan [1,4]. In these studies it was demonstrated that FITC-hyaluronan binding is dependent on the level of CD44 expression and is typically only seen after a threshold level of CD44 expression is reached. These studies also reported that CD44 mutants in which the cytoplasmic domain has been removed (CD44-T) have a hyaluronan binding defect; although this mutant receptor binds hyaluronan, the threshold expression level required for hyaluronan binding is greater than that observed for WT CD44. Here we have investigated whether the spreadsheet method can be used to compare the hyaluronan binding capacity of two CD44 constructs with unmatched expression levels without the need to generate matched populations by FACS sorting or single cell cloning.

AKR1 cells were stably transfected with human WT CD44 or CD44-T and were subjected to the FITC-HA binding assay as described in Fig. 1, except that CD44 expression levels were monitored using the mouse anti-human CD44 mAb E1/2 followed by anti-mouse PE-F(ab)$_2$. Histogram analysis of data (Fig. 2A) shows that although total populations of both transfectants have a similar mean CD44 fluorescence intensity (geometric mean 249 and 253 arbitrary units for WT CD44 and CD44-T respectively), the CD44-T cells display a much wider receptor expression range. The corresponding FITC-hyaluronan binding histograms show that the majority of CD44-T transfected cells exhibit little ligand binding above control levels. The remaining cells, however, have a broad binding profile

that overlaps and extends beyond the maximum FITC-hyaluronan binding by WT CD44. Examining the two dimensional dot-plots (Fig. 2B) indicates that a FITC-hyaluronan binding threshold exists for both WT CD44 and CD44-T and that this threshold is possibly higher for CD44-T. However, the complexity of these profiles does not permit any accurate conclusions to be drawn from this data and highlights the deficiencies in comparisons between the two cell populations using traditional graphical outputs.

Using the spreadsheet, the average hyaluronan binding for cells lying in each PE-CD44 channel is calculated, providing visually simplified dot plots presented as overlays. The plots generated by the spreadsheet clearly demonstrate that the threshold level of hyaluronan binding by WT CD44 is reached at approximately 450 fluorescence units while the threshold for the CD44-T transfected cells is reached at approximately 600 fluorescence units. However, once this threshold has been reached, the CD44-T highest expressing cells reach binding levels similar to those of WT CD44.

In addition, a Student's t-test can be applied to the data to identify regions of the plot where FITC-hyaluronan binding is significantly different between WT and mutant CD44. This statistic is only calculated for CD44 channels where 4 cells or more are counted for both cell lines. If the FITC-hyaluronan binding between two cell lines counted in a particular CD44 channel is found be significantly different at the 99.9% level, a point is plotted at position 980 on the y-axis for that particular channel. If there is no significant difference, a point is not plotted (see Fig. 2C). Using this analysis, the region between 450 and 800 fluorescence units displays the most robust area of statistical significance.

This second example illustrates the problems of comparing ligand binding in two cell populations transfected with different receptor constructs. In the case of hyaluronan binding by WT CD44 and CD44-T, the problem is acute as it is difficult to achieve transfected populations with similar expression profiles possibly because the CD44-T mutant has a significantly reduced half-life compared to WT CD44 [10]. With the spreadsheet method a direct comparison has been made between these two non-identical transfected populations. The demonstration that CD44-T can bind hyaluronan with high efficiency provided it is expressed at sufficiently high levels provides important clues as to how ligand binding by CD44 might be regulated. One explanation for the data presented here is that CD44 needs to be stabilised at the plasma membrane, for example by clustering or association with the cytoskeleton, and that this is only achieved at threshold levels of receptor expression [11]. The higher threshold of
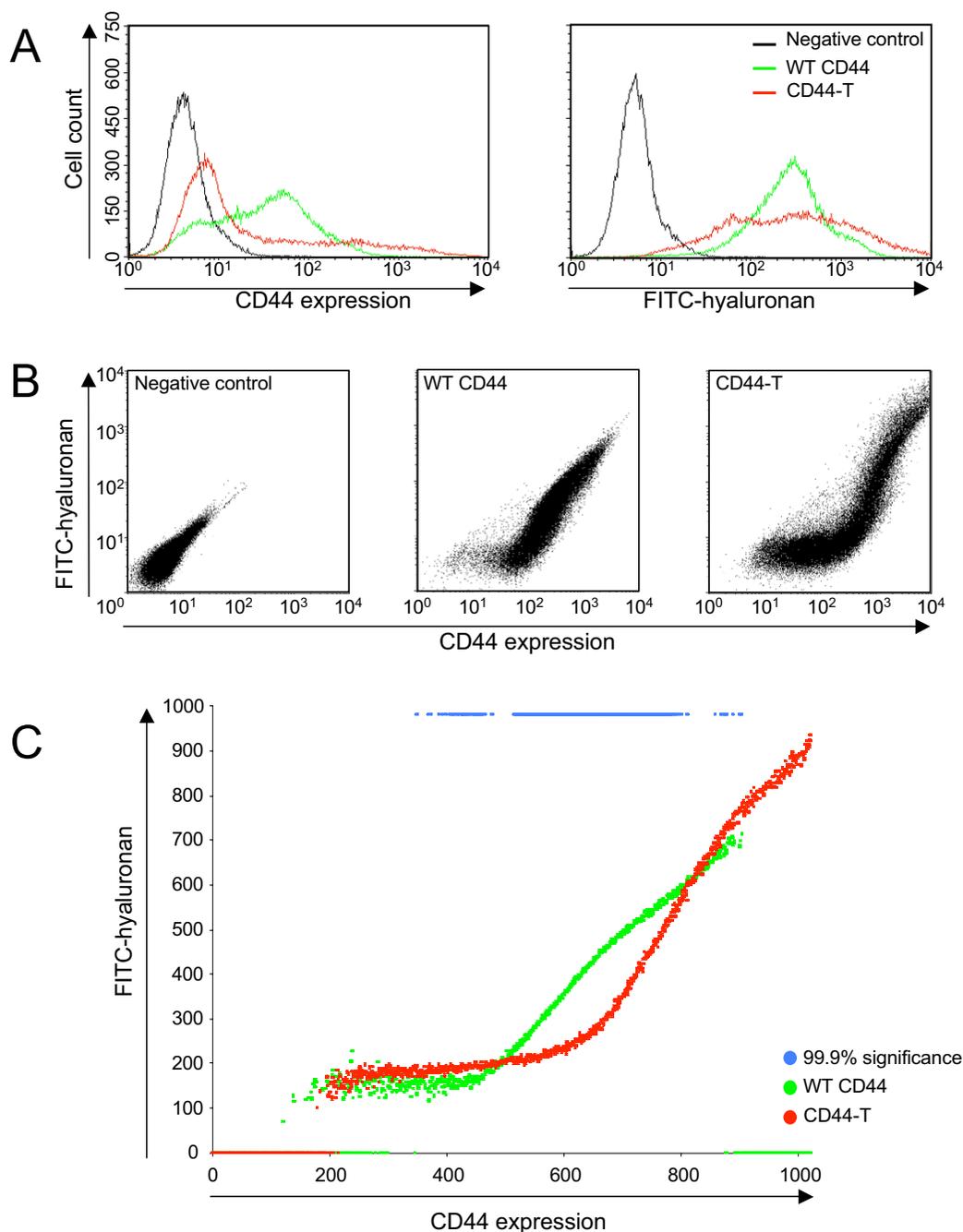
Figure 2
Use of the spreadsheet method incorporating students' t-test analysis to examine hyaluronan binding function of wild type and mutant CD44 receptors expressed in AKR1 cells AKR1 cells transfected with WT CD44 or CD44-T were stained for three-colour flow cytometry as described in the Methods. For the negative control, WT CD44 expressing cells were stained with FITC-BSA and rabbit anti-mouse PE-F(ab)$_2$. The flow cytometry data is displayed as follows: **A**, histogram plots showing CD44 expression or FITC-hyaluronan binding. **B**, two-parameter dot plots correlating FITC-hyaluronan binding against CD44 expression. **C**, the data was analysed using the spreadsheet method (see text for further details). Where there were 3 cells or fewer for a given PE fluorescence channel, the point is plotted at 0 on the y-axis. Values are arbitrary fluorescence units. Analysis used the spreadsheet method including application of a students' t-test. Where there was a significant statistical difference in FITC-hyaluronan binding between WT CD44 and CD44-T at the 99.9% level, a point is plotted at 980 on the y-axis.

CD44-T expression required for ligand binding may reflect a requirement for the cytoplasmic domain in stabilising the receptor at the cell surface but that this requirement can be overcome if sufficiently high levels of the mutant receptor are expressed due to the enforced close proximity of receptors.

Conclusions
The spreadsheet method demonstrated here is applied to the problem of CD44-hyaluronan binding but is also generally applicable to the study of other receptor-ligand interactions or where two dependent parameters are being compared using flow cytometry. The large data sets acquired by flow cytometry are intrinsically complex and problematic to analyse. Previous workers have attempted to mathematically model flow cytometric curves of cell populations [12] but the complex nature of these curves has been a barrier to further analysis. Roederer and colleagues [13] developed a test they have termed 'probability binning analysis' to determine whether a test distribution of flow cytometry data is different from a control distribution. This was done by dividing data into a series of bins each containing an equal number of cells and applying a variant of the chi-squared statistic. This method estimates the probability that the two distributions are significantly different and although powerful, this approach is relatively difficult to implement. The spreadsheet method provides a considerable advantage over previous techniques in that it utilises commonly available programs to simplify flow cytometric data. This constitutes a straightforward improvement upon the standard form of graphical output of flow cytometric data generating a representation of areas of statistical significance. In addition, this method provides the first step for further manipulation of the data, for example to calculate affinity constants or to perform more complex statistical analyses, using advanced mathematical packages.

Methods
Cell lines and flow cytometry
The mouse T-cell lymphoma cell lines BW5147 and AKR1 were maintained as previously described [9,10]. Populations of AKR1 cells transfected with WT CD44 and the cytoplasmic tail truncation mutant CD44-T constructs in the pSRα eukaryotic expression vector were established and selected as previously described [1].

For binding assays, 2×10^6 cells were washed twice, incubated for 1 h with 250 microlitres of 10 micrograms/ml FITC-hyaluronan at 37°C before washing twice more. All dilutions and washes were done in Hanks Balanced Salts Solution (HBSS; Life Technologies) supplemented with 1% foetal calf serum (FCS; Life Technologies). In some experiments, FITC-bovine serum albumin (Molecular Probes) was used at 10 micrograms/ml as a negative con-

trol. BW5147 cells were subsequently stained with 1 micrograms/ml biotinylated anti-CD44 mAb IM7 (Caltag Medsystems) followed by 0.5 micrograms/ml PE-streptavidin (Pharmingen, Becton-Dickinson) and AKR1 cells stained with 1 microgram/ml anti-CD44 mAb E1/2 followed by 25 micrograms/ml PE-conjugated rabbit anti-mouse Ig F(ab)$_2$ (DAKO Cytomation). Cells were washed twice and resuspended in 0.3 micromolar TO-PRO-3 (Molecular Probes) diluted in phosphate buffered saline. A Becton-Dickinson FACSCalibur analyser running CellQuest V3.2 software (Becton-Dickinson) was used to read cell fluorescence values. The population of cells with low TO-PRO-3 fluorescence was selected and the phycoerythrin (PE)-CD44 and FITC-hyaluronan fluorescence values of 60,000 cells were read.

Data was collected with compensation adjusted for FITC (FL-1 = FL-1 – 0.3% FL-2) and PE (FL-2 = FL-2 – 27% FL-1). In our application, we have found that compensation settings have little overall effect on the data provided that the flow cytometer photomultiplier (PMT) voltages are set up so that fluorescence is detected well within the available range (data not shown). This minimises the chance of compensation moving data into an area where the detection range becomes non-linear (usually at the borders of detection) or even outside detection thus giving a skewed result. However, it is strongly suggested that the use of alternative fluorochromes or different instrumentation will require optimisation of the compensation levels.

Analysis of data using the spreadsheet method
Formulae used in the spreadsheet calculation are listed in Table 1. The first step in the analysis sequence is to locate all cells with identical fluorescence channel values for CD44 and average their corresponding FITC-HA fluorescence channel values (formula A). This incorporates formula B where if 3 cells or fewer are found with a particular CD44 value the value "x" is returned, the average is not calculated and a point is plotted at position 0 on the y-axis of the dot plot. If two cell lines or treatments are being compared then a paired Students t-test can be performed. The standard deviation and variance of averaged data (formulas C and D, respectively) are first calculated followed by the confidence interval, which is calculated using the Excel function (formula E). The t-test is performed whenever the spreadsheet detects valid data resulting from formula A (i.e. more than 4 cells of a particular fluorescent intensity are present and an average value has been calculated) in both cell lines or treatments (formula F). The calculation for degrees of freedom is shown in formula G and critical t-values for a range of degrees of freedom are generated using an Excel function (formula H). These are then compared to the calculated t-values from formula F and a point plotted in the chart for significance at the 99.9% level (formula I) at position 980 on the y-axis and

Table 1: Excel Formulae used in spreadsheet analysis

Formula A (average)	{=(IF(D3<4,"x",AVERAGE(IF($A$3:$A$60000=$A2,$B$3:$B$60000)))))}
Formula B (cell count)	=COUNTIF($A$3:$A$60000,$A2)
Formula C (standard deviation)	{=IF(B3="x","x",(STDEV(IF($A$3:$A$60000=$A2,$B$3:$B$60000)))))}
Formula D (variance)	=IF(B3="x","x",C3^2)
Formula E (99.9% confidence interval)	=IF(B3="x",0,CONFIDENCE(0.001,C3,D3))
Formula F (paired students t-test)	=IF((AND(B3<>"x",N3<>"x")),(ABS(B3-N3))/(SQRT((F3/D3)+(R3/P3))),"x")
Formula G (degrees of freedom)	=IF((AND(B3<>"x",N3<>"x")),((D3+P3)-2),"x")
Formula H (critical t value 99.9%)	=TINV(0.001,T3)
Formula I (Significance at 99.9%)	=IF((AND(B3<>"x",N3<>"x")),(IF((H3>(LOOKUP(I3,$T$3:$T$1002,$V$3:$V$1002))),980,-50)),-50

Formulae used in spreadsheet calculations to analyse flow cytometric data. See methods for further details on the function of each.

using the channel value as the x-axis coordinate. If the calculated t-value is greater than the corresponding critical t-value then the null hypothesis that the two fluorescence readings are the same is accepted and a point is not plotted.

Instructions for use

Data was extracted from the Becton-Dickinson format using the program **FCS Assistant version 1.1** http://www.fcspress.com (©R. Hicks, UK) although other programs which are capable of extracting raw data from flow cytometry files (for example WinList (Verity Software House, ME, USA) or FlowJo (Tree Star Inc. CA, USA)) can also be used. Each CellQuest data file to be analysed was opened in FCS Assistant and flow cytometric data exported as raw tabular text from the FILE menu. The resulting raw tabular text files (*.rtt) were opened in Microsoft Excel and the appropriate FL columns for the "Constant" and "Variable" values were selected, copied and pasted into the corresponding calculation sheet columns. The "Constant" column (FL-2/CD44) corresponds to the x-axis coordinate and the "Variable" column (FL-1/FITC-hyaluronan) corresponds to the y-axis coordinate. The data transfer was repeated for each cell line or treatment to be analysed. A macro was prepared to start spreadsheet calculations in Microsoft Excel spreadsheet calculation was initiated by pressing the "Calculate spreadsheet" button in the "DATA" sheet window of the workbook. Spreadsheet files have been prepared and tested on Windows 2000/XP using Microsoft Office 2000 Professional and on Apple Macintosh OS 9.2 using Microsoft Office 2001. The graphical output is automatically generated in the "Chart" worksheet.

Authors' contributions

The experimental work and the development of the spreadsheet was performed by GT towards his PhD studies. CMI and RFT assisted with the preparation of the manuscript and supervised the experimental work.

Additional material

Additional File 1
Apple Macintosh OS 9.2 compatible Microsoft Excel (2001) spreadsheet for comparison of two cell lines only including a students' t-test analysis.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-5-10-S1.xls]

Additional File 2
Apple Macintosh OS 9.2 compatible Microsoft Excel (2001) spreadsheet for comparison of up to 6 cell lines without students' t-test analysis.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-5-10-S2.xls]

Additional File 3
Microsoft Windows 2000 and XP compatible Microsoft Excel (2001) spreadsheet for comparison of two cell lines only including a students' t-test analysis.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-5-10-S3.xls]

Additional File 4
Microsoft Windows 2000 and XP compatible Microsoft Excel (2001) spreadsheet for comparison of two cell lines only including a students' t-test analysis.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2121-5-10-S4.xls]

References
1.   Uff CR, Neame SJ, Isacke CM: **Hyaluronan binding by CD44 is regulated by a phosphorylation- independent mechanism.** *Eur J Immunol* 1995, **25:**1883-1887.

2. Lesley J, English NM, Gal I, Mikecz K, Day AJ, Hyman R: **Hyaluronan binding properties of a CD44 chimera containing the link module of TSG-6.** *J Biol Chem* 2002, **277:**26600-26608.
3. Peck D, Isacke CM: **CD44 phosphorylation regulates melanoma cell and fibroblast migration on, but not attachment to, a hyaluronan substratum.** *Curr Biol* 1996, **6:**884-890.
4. Perschl A, Lesley J, English N, Trowbridge I, Hyman R: **Role of CD44 cytoplasmic domain in hyaluronan binding.** *Eur J Immunol* 1995, **25:**495-501.
5. Day AJ, Prestwich GD: **Hyaluronan-binding proteins: tying up the giant.** *J Biol Chem* 2002, **277:**4585-4588.
6. Ponta H, Sherman L, Herrlich PA: **CD44: from adhesion molecules to signalling regulators.** *Nat Rev Mol Cell Biol* 2003, **4:**33-45.
7. Liu D, Zhang D, Mori H, Sy MS: **Binding of CD44 to hyaluronic acid can be induced by multiple signals and requires the CD44 cytoplasmic domain.** *Cell Immunol* 1996, **174:**73-83.
8. Lesley J, Hyman R, English N, Catterall JB, Turner GA: **CD44 in inflammation and metastasis.** *Glycoconj J* 1997, **14:**611-622.
9. Lesley J, He Q, Miyake K, Hamann A, Hyman R, Kincade PW: **Requirements for hyaluronic acid binding by CD44: a role for the cytoplasmic domain and activation by antibody.** *J Exp Med* 1992, **175:**257-266.
10. Neame SJ, Isacke CM: **The cytoplasmic tail of CD44 is required for basolateral localization in epithelial MDCK cells but does not mediate association with the detergent-insoluble cytoskeleton of fibroblasts.** *J Cell Biol* 1993, **121:**1299-1310.
11. Thorne RF, Legg JW, Isacke CM: **The role of the CD44 transmembrane and cytoplasmic domains in co-ordinating adhesive and signalling events.** *J Cell Sci* 2004, **117:**373-380.
12. Lampariello F, Aiello A: **Complete mathematical modeling method for the analysis of immunofluorescence distributions composed of negative and weakly positive cells.** *Cytometry* 1998, **32:**241-254.
13. Roederer M, Treister A, Moore W, Herzenberg LA: **Probability binning comparison: a metric for quantitating univariate distribution differences.** *Cytometry* 2001, **45:**37-46.