Clin. Lab. Haem.
1999, **21**, 155–160

REVIEW
Quality control of flow cytometric immunophenotyping of haematological malignancies

J.W. GRATAMA*
R.L.H. BOLHUIS*
M.B. VAN 'T VEER†

**Department of Clinical and Tumor Immunology and †Hematology, University Hospital, Daniel den Hoed Cancer Center, Rotterdam, The Netherlands*

Summary Immunophenotyping of haematological malignancies has developed as a clinically valuable but technically complicated diagnostic procedure. It involves a variety of methodological features, in-process strategic judgements and an extensive knowledge of clinical, morphological and other laboratory features of the disease processes under study. We discuss the various internal quality control steps necessary to guarantee reliable results with respect to instrument set-up and calibration; sample preparation; selection and validation of monoclonal antibody panels; and flow cytometric data acquisition, analysis and interpretation of results. The quality of the entire procedure is documented by the analysis of representative specimens in the setting of an external quality assurance programme.

Keywords Flow cytometry, immunophenotyping, leukaemia, lymphoproliferation, quality control

Introduction

Immunophenotyping is a widely accepted diagnostic tool for many haematological malignancies in the context of clinical, morphological and cytogenetic data. Flow cytometry is regarded as the method of choice for immunophenotyping because it is fast, objective, quantitative and amenable to standardization (Jennings & Foon 1997). Flow cytometric immunophenotyping is an essential part of the diagnostic procedure in acute lymphoblastic and myeloid leukaemias (ALL and AML) and chronic lymphoproliferative disorders, and its results constitute useful information for therapeutic decision-making in these diseases. Also, flow cytometric monitoring during residual disease may have diagnostic and therapeutic utility in patients with acute leukaemia (Brisco *et al.* 1996; Davis *et al.* 1997; San Miguel *et al.* 1997).

Importantly, flow cytometric immunophenotyping of haematological malignancies is not a simple and standardized assay which provides precise numerical results carrying direct diagnostic information. The various methodological steps, the interpretation of the data and the clinical significance of the results require substantial participation and critical judgement by the laboratory professionals involved. Adequate quality control (QC) procedures are therefore essential. We discuss the consecutive steps during the entire flow cytometric immunophenotyping assay with special emphasis on the various *internal* QC procedures to be performed. We refer the reader for an extensive discussion of these procedures to the consensus recommendations that have been formulated by experts in Europe (Rothe *et al.* 1996) and in the US and Canada (Borowitz *et al.* 1997; Braylan *et al.* 1997a,b; Davis *et al.* 1997; Stelzer *et al.* 1997; Stewart *et al.* 1997).

In addition to performing internal QC, participation in an *external* quality assessment (EQA) programme is useful for each laboratory involved in immunophenotyping haematological malignancies to document the overall quality of its output. The UK National External Quality Assessment Schemes (NEQAS) provide send-outs of stabilized whole blood specimens from patients with haematological malignancies to laboratories in the UK and abroad, followed up by written evaluations (Barnett, Granger & Reilly 1994). The Dutch Foundation for the Immunophenotyping of Hematological Malignancies (SIHON) organizes biannual send-outs of cryopreserved mononuclear cell suspensions to laboratories in the Netherlands and Belgium, followed up by plenary educational meetings (Kluin-Nelemans *et al.* 1996). Although participation in an EQA programme is currently on a voluntary basis, it is expected that it will become compulsory in the near future as part of laboratory

Accepted for publication 18 January 1999
Correspondence: Dr J.W. Gratama, Department of Clinical and Tumor Immunology, Daniel den Hoed Cancer Center, PO Box 5201, 3008 AE Rotterdam, The Netherlands.

accreditation for the immunophenotyping of leukaemias and lymphoproliferations.

Instrument set-up and calibration

To ensure that the flow cytometer performs adequately for any given application, the instrument must be set up appropriately and its performance in measuring fluorescence (FL) intensity must be verified. The design and manufacture of a range of microbead (reviewed by Schwartz *et al.* 1998) and cellular control materials has been pivotal in reaching this goal. To assure optimal instrument performance, two groups of procedures must be performed.

The first group of procedures is carried out at relatively infrequent intervals (e.g. once every 6 months) by qualified service personnel and includes an examination of the efficiency and performance of the laser tube, optical filters, log and linear amplifiers, and photomultiplier tubes (PMT) of all types of flow cytometers. These procedures also include a calibration of the optical alignment in flow cytometers of the 'stream-in-cuvette' type, which constitutes the majority of instruments in use for clinical diagnostic procedures. However, the optical alignment of flow cytometers with sorting facilities, which are equipped with a nozzle ('stream-in-air' type), must be calibrated at each 'cold start' by the instrument operators owing to their relative instability.

The second group of procedures consists of frequent (i.e. at each 'cold start' of the instrument for a given application) monitoring of instrument set-up and performance by the operators to identify both immediate and potential problems (see Schwartz *et al.* 1996 and Stelzer *et al.* 1997 for detailed protocols). Optimal instrument set-up implies that all cellular populations are visible on each FL or light scatter scale. This optimization is performed with a representative stained cell specimen. For forward (FSC) and sideward (SSC) light scatter analysis, linear amplification is generally used, although logarithmic amplification is more useful for the simultaneous visualization of normal and aberrant cell populations with high SSC signals, such as hairy leukaemia cells, plasma cells or cells derived from solid tumours. For FL analysis, logarithmic amplification is generally used because of the wide dynamic range of FL intensities in most biological samples.

Having done so, the correction must be set for the variable levels of overlap between the emission spectra of routinely used dyes such as the green-fluorescent dye fluorescein isothiocyanate (FITC), the orange-fluorescent dye phycoerythrin (PE), and the red-fluorescent dyes peridinin chlorophyll (PerCP), PE-Cy5 or allophycocyanin (APC). The establishment of appropriate colour compensation for any number of dyes requires an equal number of cell suspensions single-labelled with monoclonal antibodies conjugated to those dyes, plus one suspension stained with all dyes. Each cell population should span the FL intensities ranging from weakly positive to a representative high for that assay. Electronic compensation for spectral overlap can be set manually or automatically through the use of certain software packages.

After establishing appropriate application-specific instrument settings, the target channels for the relevant FL parameters are recorded using fluorescent reference beads. The use of these target channels allows, at subsequent occasions, a rapid and reliable verification of the instrument settings. For light scatter parameters, it suffices to run a representative specimen and to verify that all major cell populations are in their typical positions in the FSC vs. SSC dotplot.

The performance of the instrument in quantitative FL measurements is monitored by computing calibration plots for each FL parameter from data generated by the measurement of calibration beads labelled with the relevant dyes in different pre-defined intensities. Trends and variations are documented using Levey–Jennings style charts (Levey & Jennings 1950) that allow the visual inspection of longitudinal data for monitoring instrument performance for precision and trends. If any value falls outside the tolerance limits, the monitoring should be repeated and the instrument settings adjusted if the problems persist.

Sample preparation

The range of specimens submitted for the immunophenotyping of haematological malignancies includes not only venous blood, but frequently bone marrow aspirates and also visceral fluid samples. Collection of the latter types of specimens involves more invasive procedures with more patient discomfort than venepuncture. Therefore, a more liberal policy for specimen rejection is warranted than would be usual, e.g. for enumeration of lymphocyte subsets. However, acceptance of compromised specimens for processing implies that strict and detailed flagging criteria must be used, such as the clotting or haemolysis of bone marrow aspirates or the presence of erythrocytes in liquor samples.

Flow cytometry requires that a single cell suspension is prepared for analysis. Clots and bone spicules may be disrupted using a filter mesh. Most importantly, the risk that the cells of interest, i.e. the abnormal cells, are lost during sample preparation should be minimized. The availability of multiple (i.e. $\geqslant 5$) parameter flow cytometry

allows the analysis of highly complex cellular mixtures. Owing to this, the removal of neutrophils and dead cells by density gradient centrifugation is no longer necessary. Still, current flow cytometric technology requires the physical removal of erythrocytes from blood and bone marrow aspirates, in which they typically outnumber leucocytes by approximately a thousand-fold. Hypotonic erythrocyte lysis carries a smaller risk of loosing abnormal cells than density gradient separation: the latter method is optimized for lymphocyte isolation and abnormal cells that do not have a similar buoyant density as lymphocytes will not be found where expected in the gradient. Therefore, erythrocyte lysis is recommended over gradient separation for immunophenotyping haematological malignancies (Stelzer *et al.* 1997). Importantly, the cell suspensions prepared for flow cytometry should be representative of the fraction of the sample that may contain the abnormal cells. This control can be obtained by comparing the flow cytometric results with cytochemically stained smear preparations of peripheral blood and bone marrow aspirates, and cytospin preparations from visceral fluid samples.

The policy of making every attempt to obtain useful information from each specimen, regardless of its condition, requires that viable and dead nucleated cells be distinguished. Dead cells can severely compromise an immunophenotypic analysis because the damaged surface membrane of such cells may allow the nonspecific uptake of monoclonal antibodies, leading to nonspecific FL signals. The addition of a dye identifying dead cells to all staining cocktails of such samples allows the exclusion of dead cells during data acquisition or during analysis. Useful dyes are 7-amino-actinomycin D (7-AAD) for unfixed samples (Schmid *et al.* 1992) or ethidium monoazide (EMA) for samples that are to be fixed (Riedy *et al.* 1991).

The integrity of the cellular surface membrane is thus a prerequisite for the reliable detection of cell surface antigens. The intracellular or surface membrane localization of some markers is crucial for the interpretation of staining results (e.g. cytoplasmatic CD3 → immature and surface membrane CD3 → mature T lineage cells). Fixation and permeabilization for intracellular antigen detection should therefore be performed after the completion of surface staining. Any method used for fixation and permeabilization must preserve the expression and antigenicity of the markers to be evaluated.

Finally, the number of lymphocyte subsets per unit volume of peripheral blood contributes useful information for monitoring chronic lymphoproliferative diseases. Such absolute cell counts can reliably be obtained by flow cytometry of a single immunophenotyping staining through the addition of a known number of brightly fluorescent counting beads (Gratama *et al.* 1998a).

Selection and validation of monoclonal antibody panels

Any choice of monoclonal antibody combinations for the initial investigation of haematological malignancies must allow the distinction of neoplastic cells from their normal counterparts as well as the enumeration and further characterization of the neoplastic cells. From the 166 clusters of differentiation (CD) defined in 1996 (Kishimoto *et al.* 1997) and the numerous unclustered antigens, some 40–50 bear direct relevance for this purpose (Rothe & Schmitz 1996; Stewart *et al.* 1997). Neoplastic haematopoietic cells not only reflect, to a certain extent, normal cellular maturation, but they also show frequently aberrant phenotypes that distinguish them from their normal counterparts. Such 'malignant' phenotypes have been reported in 90% of ALL, 75% of AML and 90% of monoclonal plasma cell proliferations (Jennings & Foon 1997). The use of at least 5 parameters (forward (FSC) and sideward (SSC) light scatter plus 3 FL parameters) is essential to resolve complex, aberrant immunophenotypes in the many clinical specimens that also contain significant proportions (i.e. $> 80\%$) of normal cells.

Given the large number of monoclonal antibody combinations possible in triple and even quadruple marker stainings only general recommendations can be given for the selection of monoclonal antibody panels (Stewart *et al.* 1997). Depending on the organization of each laboratory and the (available information on the) submitted specimens, one-step diagnostic strategies may consist either of a single, comprehensive monoclonal antibody panel, or of a choice between a limited number of targeted panels. The latter should contain a few 'safeguard' stainings in case the clinical information, on which the panel choice has been based, is wrong (van 't Veer *et al.* 1992). The alternative two-step approach consists of a small screening panel followed by a larger, targeted panel. Advantages of the single-step comprehensive panel are extensive information and minimal requirement of in-process strategic judgements; disadvantages are high reagent costs and inapplicability to samples with low cell counts. The targeted approaches imply more in-process strategic judgements and may be more time-consuming, but save on reagent costs. In short, there is no single 'golden rule' for defining monoclonal antibody panels.

For the follow-up of a given haematological malignancy, a targeted approach tailored to the 'malignant' phenotype is appropriate. In this context, the judicious selection of a few three or four-colour monoclonal antibody cocktails is critical to obtain an appropriate sensitivity and specificity for the detection of minimal residual disease. Still, sufficient safeguards must be built in to enable the detection of emerg-

ing malignant subpopulations with an immunophenotype that differs from the original (San Miguel *et al.* 1997).

Whatever the strategy of the laboratory, the chosen panel(s) of monoclonal antibody combinations must be validated. Firstly, the monoclonal antibody conjugates should be chosen such that spectral overlap can be adequately controlled. PE is preferred as dye for monoclonal antibody-detecting antigens expressed at a low intensity because of its high quantum yield and lack of interference by cellular autofluorescence in comparison to FITC. Secondly, no steric hindrance should occur between the monoclonal antibodies in each cocktail. Therefore, the monoclonal antibodies labelled in combination should yield FL signals of the same intensity as that obtained in single-colour stainings. Thirdly, the amount of monoclonal antibody used for staining must be verified for the staining procedure and type of samples used. The manufacturers' recommendations are often only based on staining with normal cells. Nonspecific monoclonal antibody-binding should be minimized whilst at the same time retaining maximum discrimination between positive and negative cell populations. The chosen amount of monoclonal antibody for staining must be verified whenever a new batch of monoclonal antibody is put into use. This control is conveniently performed by comparing the performance of the old and new batches on an informative cell suspension.

We agree with the US–Canadian consensus recommendations (Stelzer *et al.* 1997) that a positive procedure control (i.e. cells of a healthy donor) and isotype control stainings are redundant. Even specimens dominated by an abnormal population will contain at least a few residual normal cells which serve as benchmark for an appropriate staining technique. Any fluorochrome and isotype control monoclonal antibody is by definition not representative for the great variety of used monoclonal antibodies with their different protein concentrations and fluorochrome to protein ratios. Instead, any monoclonal antibody panel will yield at least a few negative populations for each fluorochrome that can serve as control for nonspecific monoclonal antibodies binding in comparison to unstained cells as an autofluorescence control. (Cellular autofluorescence is increased in some AML cases and during cytoreductive treatment with anthracyclins.)

Flow cytometric data acquisition, analysis, and interpretation of results

Different strategies to acquire and analyse flow cytometric data are used for the initial investigation of haematological malignancies and for follow-up studies. As stated above, the objective of the initial investigation is twofold, to resolve the abnormal population from the normal population, and to characterize the abnormal cells in more detail. For that purpose, a sufficient number of viable cells (typically 10 000–20 000 per staining) should be acquired. The abnormal cells can be identified on the basis of qualitative (i.e. an aberrant phenotype) and/or quantitative data (i.e. the abnormal population cannot be resolved phenotypically from its normal counterparts other than by its exceedingly high frequency in the sample). An appreciation of the normal ranges is essential for the correct interpretation of such data. Non-malignant blood and bone marrow samples from some tens of individuals of the relevant age range, analysed with the monoclonal antibody cocktails for an initial diagnosis in the relevant technique, may constitute a reference source in this respect.

Further characterization of the abnormal population is typically performed by 'gating', i.e. the selection of a specific group of cells on the basis of a (combination of) parameters. These parameters should be detectable in subsequent stainings if all necessary information cannot be retrieved from a single staining. A discussion of detailed gating techniques is outside the scope of this overview. There are two major strategies. In the first, termed sequential gating, events are selected on the initial dotplot or histogram and subsequent gates are placed on the selected events in a cumulative manner (Gratama *et al.* 1998a). An alternative strategy is to keep all dots on screen and to select the relevant cells by combining different regions in different dotplots based on simultaneous gating. According to this technique, cells fulfilling the criteria of two or more different regions are pseudo-colour coded and separately analysed without removing non-relevant cells from the display. This approach was first used in the Paint-A-GATE™ software (Becton Dickinson Immunocytometry Systems, San Jose, CA, USA) and is currently available within a variety of software packages from different manufacturers.

With the abolition of the isotype control stainings, the traditional practice of setting a threshold on the negative population to subsequently calculate the 'percentage positives' by applying this threshold to a single-colour histogram of a stained population, is now considered inappropriate for immunophenotyping haematological malignancies (Borowitz *et al.* 1997). Information conveyed in this way is only accurate if the gated population is pure and the FL distribution is bimodal with well-separated peaks (e.g. $CD4^+$ cells within a population of $CD3^+$ T lymphocytes). Thus, a qualitative description of the abnormal cell population with respect to additional markers is informative and sufficient.

Importantly, the resolution of flow cytometric immunophenotyping of haematological malignancies can be fur-

ther improved by taking FL intensity (approximate density of antigen expression) into account whilst interpreting the data. Useful examples are the low to absent expression of CD45 by AML and ALL blasts (Borowitz *et al.* 1993) and the weak expression of CD20 by B-chronic lymphatic leukaemia cells (Marti et al. 1992). Research methodologies are now available to express this information in a quantitative manner (reviewed by Gratama *et al.* 1998b). However, it remains to be established whether or not such quantification increases the diagnostic power of the assay as compared to the simple, qualitative description of aberrant levels of FL intensity relative to normal reference populations in the sample.

Follow-up studies of haematological malignancies serve to monitor the extent of disease, whilst being vigilant for emerging malignant subpopulations with an immunophenotype differing from the original. Low-level minimal residual disease can only be detected if the detected phenotype of the malignant cells is absent or extremely infrequent in normal samples. Reliable detection of minimal residual disease requires that the aberrant cells can be detected as a cluster (i.e. at least 50 events) in a bivariate dotplot of gated events. This approach often requires the collection of very large list mode data files (e.g. up to 10^6 events). As the residual malignant cells are 'rare events' in such samples, their number will follow a Poisson distribution. According to such a distribution, their CV (in percentage) is $100 \times \text{sqrt}(n) \div n$, in which n = number of events fulfilling the criteria of the malignant cells. Hence, their CV will vary proportionally to the square root of the number of residual malignant cells acquired. For example, acquiring 50 of such cells yields a CV of $100 \times \text{sqrt}(50) \div 50 = 15\%$.

The final control of each immunophenotypic investigation occurs with formulating the summary conclusion. The immunological information should then be combined with appropriate clinical, morphological and other laboratory information. Therefore, the professionals responsible for the final interpretation and signature of the immunophenotyping report must possess a combined knowledge of the laboratory and clinical manifestations of haematological malignancies, as well as practical experience in analytical flow cytometry. The final interpretation of a disease condition that influences patient management can only be given by the physician in charge (Braylan *et al.* 1997b).

The future: consequences for external quality assurance surveys

During the past 10 years, flow cytometric immunophenotyping of haematological malignancies has evolved from single-colour analysis of surface membrane antigens expressed by gradient-isolated mononuclear cells to simultaneous four-colour analyses of surface membrane and intracellular antigens in cell suspensions that resemble the native specimen as closely as possible. This technical progress has greatly enhanced the diagnostic power of the technique in resolving minor populations of aberrant cells in otherwise normal samples. The emerging clinical relevance of the detection of minimal residual disease (Brisco *et al.* 1996; Davis *et al.* 1997; San Miguel *et al.* 1997) requires that EQA programmes not only survey the capability of laboratories to immunophenotype and interpret specimens dominated by malignant cells, but also document their ability to resolve low-frequency aberrant populations. In this context it is important that the distributed test specimens resemble the original ones as much as possible without artefacts incurred by storage and transport. The performance of EQA programmes such as those organized by UK NEQAS and SIHON are pivotal in approaching these goals.

References

Barnett D., Granger V. & Reilly J.T. (1994) The United Kingdom external quality assurance schemes (UK NEQAS) for leukaemia phenotyping. *British Journal of Haematology* **86,** 83.

Borowitz M.J., Bray R., Gascoyne R. *et al.* (1997) US–Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry: data analysis and interpretation. *Cytometry* **30,** 236–244.

Borowitz M.J., Guenther K.L., Shults K.E. & Stelzer G.T. (1993) Immunophenotyping of acute leukemia by flow cytometric analysis: use of CD45 and right-angle light scatter to gate on leukemic blasts in three-color analysis. *American Journal of Clinical Pathology* **100,** 534–540.

Braylan R.C., Borowitz M.J., Davis B.H., Stelzer G.T. & Stewart C.C. (1997a) US–Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry. *Cytometry* **30,** 213.

Braylan R.C., Atwater S.K., Diamond L. *et al.* (1997b) US–Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry: data reporting. *Cytometry* **30,** 245–248.

Brisco M.J., Hughes E., Neoh S.H. *et al.* (1996) Relationship between minimal residual disease and outcome in adult acute lymphoblastic leukemia. *Blood* **87,** 5251–5256.

Davis B.H., Foucar K., Szczarkowski W. *et al.* (1997) US–Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry: medical indications. *Cytometry* **30,** 249–263.

Gratama J.W., Orfao A., Barnett D. *et al.* for the European Working Group on Clinical Cell Analysis (1998a) Flow cytometric enumeration of CD34+ hematopoietic stem and progenitor cells. *Cytometry* **34,** 128–142.

Gratama J.W., D'hautcourt J.L., Mandy F. *et al.* for the European Working Group on Clinical Cell Analysis (1998b) Flow cyto-

metric quantitation of immunofluorescence intensity: problems and perspectives. *Cytometry* **33,** 166–178.

Jennings C.D. & Foon K.A. (1997) Recent advances in flow cytometry: application to the diagnosis of hematologic malignancy. *Blood* **90,** 2863–2892.

Kishimoto T., Von Kikutani H., dem Borne A.E.G. *et al.* eds. (1997) *Leukocyte Typing VI.* Garland Press, London.

Kluin-Nelemans J.C., Van Wering E.R., van 't Veer M.B. *et al.* on behalf of the Dutch Cooperative Study Group on Immunophenotyping of Hematologic Malignancies (SIHON) (1996) Pitfalls in immunophenotyping of leukaemia and leukaemic lymphomas. *British Journal of Haematology* **95,** 692–699.

Levey S. & Jennings E.R. (1950) The use of control charts in the clinical laboratory. *Am J Clin Pathol* 20: 1059–1066.

Marti G., Faguet G., Bertin P. *et al.* (1992). CD20 and CD5 expression in B-chronic lymphocytic leukemia. *Annals of the New York Academy of Sciences* **261,** 480– 483.

Riedy M.C., Muirhead K.A., Jensen C.P. & Stewart C.C. (1991) Use of a photolabeling technique to identify nonviable cells in fixed homologous or heterologous cell populations. *Cytometry* **12,** 133–139.

Rothe G. & Schmitz G. for the Working Group on Flow Cytometry and Image Analysis (1996) Consensus protocol for the flow cytometric immunophenotyping of hematopoietic malignancies. *Leukemia* **10,** 877–895.

San Miguel J.F., Martínez A., Macedo A. *et al.* (1997) Immunophenotyping investigation of minimal residual disease is a useful approach for predicting relapse in acute myeloid leukemia patients. *Blood* **90,** 2465–2470.

Schmid I., Krall W., Uittenbogaart C., Braun J. & Giorgi J. (1992) Dead cell discrimination with 7-amino-actinomycin D in combination with dual color immunofluorescence in single laser flow cytometry. *Cytometry* **13,** 204–208.

Schwartz A., Fernández-Repollet E., Vogt R. & Gratama J.W. (1996) Standardizing flow cytometry: construction of a standardized fluorescence calibration plot using matching spectral calibrators. *Cytometry* **26,** 22–31.

Schwartz A., Marti G.E., Poon R., Gratama J.W. & Fernandez-Repollet E. (1998) Standardizing flow cytometry: a classification system of fluorescent standards used for flow cytometry. *Cytometry* **33,** 106–114.

Stelzer G.T., Marti G., Hurley A., McCoy P. Jr, Lovett E.J. & Schwartz A. (1997) US–Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry: standardization and validation of laboratory procedures. *Cytometry* **30,** 214–230.

Stewart C.C., Behm F.G., Carey J.L. *et al.* (1997) US–Canadian consensus recommendations on the immunophenotypic analysis of hematologic neoplasia by flow cytometry: selection of antibody combinations. *Cytometry* **30,** 231–235.

Van 't Veer M.B., Kluin-Nelemans J.C., Van der Schoot C.E., Van Putten W.L.J., Adriaansen H.J. & Van Wering E.R. (1992) Quality assessment of immunological marker analysis and the immunological diagnosis in leukaemia and lymphoma: a multicentre study. *British Journal of Haematology* **80,** 458–465.