Establishment and validation of a standard protocol for the detection of minimal residual disease in B lineage childhood acute lymphoblastic leukemia by flow cytometry in a multi-center setting;

Julie Irving,[1] Jenny Jesson,[2] Paul Virgo,[3] Marian Case,[1] Lynne Minto,[1] Lisa Eyre,[2] Nigel Noel,[3] Ulrika Johansson,[4] Marion Macey,[4] Linda Knotts,[5] Margaret Helliwell,[6] Paul Davies,[2] Liam Whitby,[7] David Barnett,[7] Jeremy Hancock,[3] Nick Goulden[8] and Sarah Lawson[2] on behalf of the UKALL Flow MRD group and UK MRD steering group

[1]Northern Institute for Cancer Research, Newcastle upon Tyne; [2]Birmingham Children's Hospital, Birmingham; [3]Southmead Hospital, Bristol; [4]The Royal London Hospital, London; [5]Yorkhill Children's Hospital, Glasgow; [6]Sheffield Children's Hospital, Sheffield; [7]UK NEQAS for Leucocyte Immunophenotyping, Royal Hallamshire Hospital, Sheffield, and [8]Great Ormond Street Children's Hospital, London, UK

ABSTRACT

Minimal residual disease detection, used for clinical management of children with acute lymphoblastic leukemia, can be performed by molecular analysis of antigen-receptor gene rearrangements or by flow cytometric analysis of aberrant immunophenotypes. For flow minimal residual disease to be incorporated into larger national and international trials, a quality assured, standardized method is needed which can be performed in a multi-center setting. We report a four color, flow cytometric protocol established and validated by the UK acute lymphoblastic leukemia Flow minimal residual disease group. Quality assurance testing gave high inter-laboratory agreement with no values differing from a median consensus value by more than one point on a logarithmic scale. Prospective screening of B-ALL patients (n=206) showed the method was applicable to 88.3% of patients. The minimal residual disease in bone marrow aspirates was quantified and compared to molecular data. The combined risk category concordance (minimal residual disease levels above or below 0.01%) was 86% (n=134). Thus, this standardized protocol is highly reproducible between laboratories, sensitive, applicable, and shows good concordance with molecular-based analysis.

Key words: childhood acute lymphoblastic leukemia, minimal residual disease, flow cytometry.

Introduction

The outcome for childhood acute lymphoblastic leukemia (ALL) has dramatically improved over the last 50 years with current cure rates approaching 90% and this is attributable to the introduction and gradual intensification of combination chemotherapy, with contemporary regimens involving the use of 7-8 drugs, along with improvement of prognostic factors.[1] However, these data suggest that a proportion of children are likely to be over treated with current therapeutic regimens and conversely a proportion may benefit from more intensive therapy. Thus, the challenges now remaining are to further increase cure rates and to achieve this cure with the minimal chemotherapy to avoid unnecessary toxicities. This goal may be achieved by tailoring therapy to each individual patient's risk of relapse.

Several studies have shown that minimal residual disease (MRD) status during the early stages of therapy provides prognostic information which is independent of more classic prognostic markers such as presenting white blood cell count, age, cytogenetic analyses and immunophenotype.[2-9] Data from these studies was sufficiently compelling to warrant the incor-

poration of an MRD assessment into subsequent trials to enable risk-directed therapy, i.e. more intensive therapy for MRD positive, high-risk patients and dose reduction for good responders. At present, most trials are too immature to conclude that MRD risk-directed therapy reduces relapse rates or maintains survival rates after treatment reduction. However, early indications are promising.

The two principal methods for MRD detection in childhood ALL are the molecular analysis of B- and T-cell receptor gene rearrangements and the flow cytometric analysis of aberrant immunophenotypes, both methods being predictive of outcome. As yet, flow MRD analyses have largely been performed in single reference laboratories but to be incorporated into larger national and international trials, a quality assured, standardized method which can be performed in a multi-center setting might be preferable. In this paper, we report data from the UK Flow MRD group, a network of 6 UK laboratories, which have validated a standardized protocol for B-lineage ALL. We show that this protocol has high sensitivity and technical applicability, has good concordance with the *gold standard* molecular-based analysis and importantly, is highly reproducible between laboratories across different instrument platforms.

Design and Methods

Detailed methodologies are described in the *Online Supplementary Appendix*.

Discussion and Results

Internal and external quality assurance testing of Flow minimal residual disease

Quality assurance testing consisted of mock MRD samples being prepared with one or more known LAIPs, posted to all 6 network laboratories, and the samples were then analyzed and reported independently. Most of these were prepared by laboratories, within the network with fresh material (n=15) but more recent samples were provided by the UK National External Quality Assessment Scheme (NEQAS) who have initiated an MRD programme using mock samples prepared with fixed, stabilized material[10] (n=6). In addition, list mode data files of MRD samples acquired in one center were analyzed by all network laboratories to assess gating strategies (n=2). Thus, there were a total of 23 quality assessment exercises consisting of 42 separate LAIP analyses. Figure 1 shows a scatter plot of the individual quality control values of the 6 network laboratories against the median value on logarithmic scales. All values lie within or close to one logarithm of the consensus value though increased variability is apparent at low levels below 0.1%. The inter-laboratory correlation coefficients range from 0.97 to 0.99. For each sample, the consensus risk category found by the laboratories was obtained, i.e. MRD equal or higher than 0.01% was classified as high risk. Discordance with the consensus was obtained in two samples, one where 4 out of 6 centers

agreed and one with 5 out of 6 agreeing. Thus the inter-laboratory agreement on risk category compared to the consensus risk was 100% for 4 laboraties, 90% for one and 80% for another (*Online Supplementary Table S2*). One of these risk discordant examples, along with many of the outliers in Figure 1, occurred early on in the study period and were attributable to inappropriate gating which was subsequently standardized during group workshops. In the second, results of the same MRD sample using 2 LAIPs in duplicate (4 assays per lab) scored consistently positive in 3 laboratories and consistently negative in 2. The MRD percentage for this sample ranged from 0.004-0.02%, with a mean and standard deviation of 0.008%±0.005 and thus was a true borderline result. There were no discordant results from more recent exercises (n=15, 31 individual Flow MRD assessments, performed over a three year period, on average 4-5 per year) or with any of the stabilized NEQAS samples.

Sensitivity and variability of the standardized method

The sensitivity of the assay was assessed by spiking leukemic blasts with a known LAIP into normal bone marrow and then preparing serial dilutions down to below 0.01%. A sensitivity of at least 0.01% was confirmed for all LAIP combinations tested including CD38 (n= 3), CD45 (n= 3), CD58 (n=5) and CD66c (also known as Korsa) (n=3). Figure 2 shows representative dot plots from one of these combinations. To assess interassay variability, mock MRD replicates were labeled with 2 different LAIPs, 45/10/34/19 and 38/10/34/19 and analyzed using 2 different cytometers, a Coulter XL and a Coulter FC500. The coefficient of variation ranged from 2.2 to 4.1%, 3.14 to 5.47% and 10.21 to 13.13% for 10%, 0.5% and 0.05% MRD mocks, respectively (*Online Supplementary Table S3*).
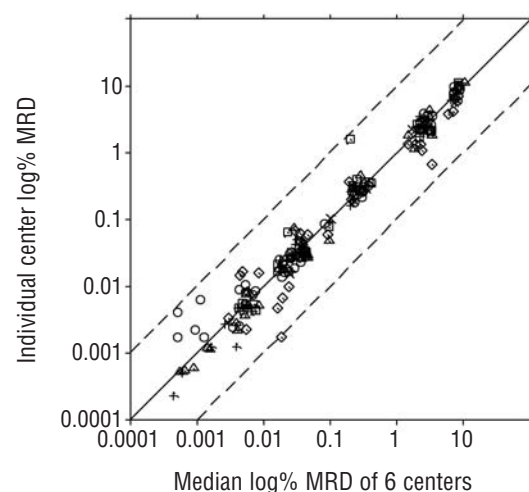
Figure 1. Scatter plot of quality control results showing the individual values for the 6 centers, a different symbol for each center, versus the median value of all 6 on a log scale. The middle line is the line of identity and the dashed outer lines represent one logarithm from the consensus value. Lab 6 has fewer data points due to their later inclusion into the network.
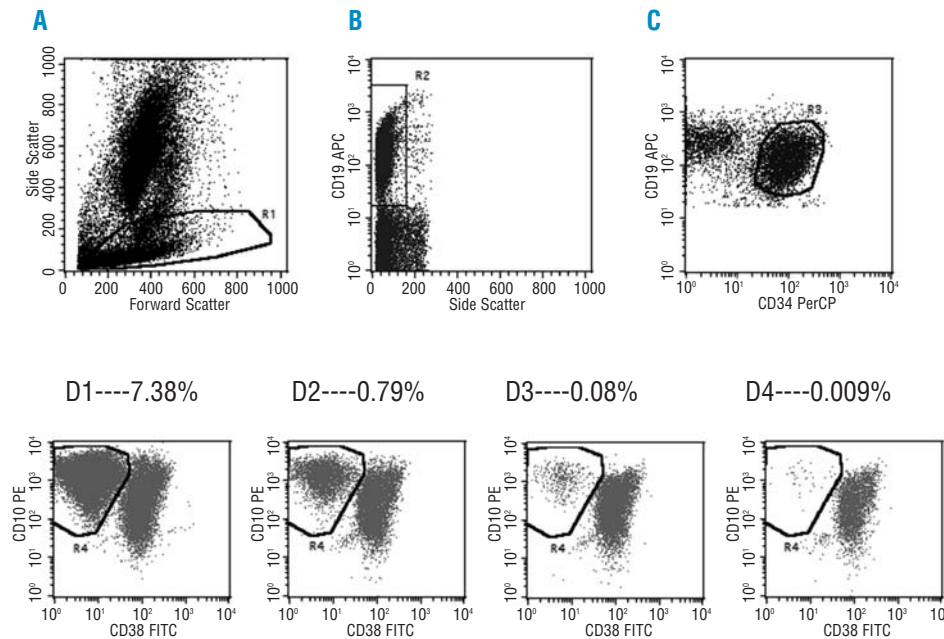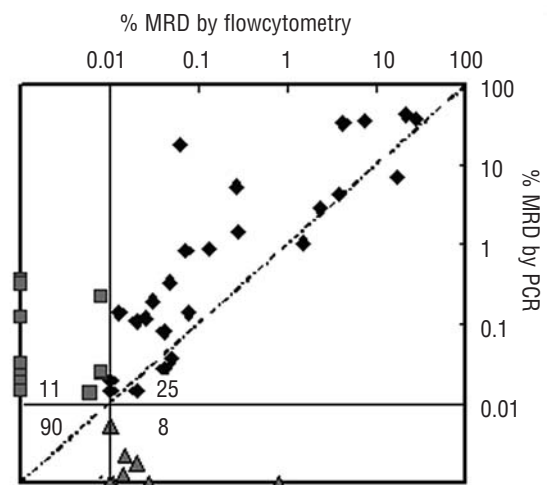
Figure 2. Mock minimal residual disease samples were prepared, serially diluted (10%-0.01%) and analyzed using the standard method. Dot plots A-D represent the sequential gating strategy, i.e. (B) is gated on a lymphoid gate (R1) derived from (A), (C) is gated on R1 and a region 2 (R2) which defines CD19, low side scatter cells. Finally, (D) dot plots are gated for R1 and R2 and a region 3 (R3) containing CD19 and CD34 dual positive cells defined in (C). Dot plots D1-4 show the final analyses plots for the serial dilutions, with the blast region identified as R4.

Applicability of the standardized method in prospective samples

From 206 diagnostic precursor B-lineage ALL patients who were prospectively screened with the standard protocol, 182 had 2 or more sensitive LAIPs and was therefore applicable for 88.3% of patients. The most commonly identified LAIPs included CD45 and CD38, selected due to their underexpression, CD58 for overexpression, and CD66c for aberrant expression, compared to normal B-cell progenitors. A table of the LAIPs and the frequency used in the cohort of patients is shown in *Online Supplementary Table S4*. Of the 182 patients assessed by Flow MRD, 24.7% (n=45) were high-risk at day 28.

Comparison of minimal residual disease as measured by PCR and by flow cytometry

The current UKALL 2003 trial involves randomization based on levels of MRD, as measured by PCR of patient-specific immunoglobulin rearrangements, at two time points during treatment: day 28, at the end of remission induction and week 11, at the completion of consolidation. MRD quantification of bone marrow aspirates at these time points was performed by both PCR and flow cytometry in 134 children. Overall, 90 samples were low-risk by both methods, 25 were high risk by both, 8 were high-risk by flow but low-risk by molecular, and 11 were low-risk by flow and high-risk by molecular (Figure 3). Excluding the 90 cases below the threshold of both methods, the percentage of cases in which log PCR and log Flow MRD were within half a log was 47.6% and within one log was 76.2%. The risk category concordance was 79% at day 28 (n=91; 25 positive and 47 negative by both techniques) and 100% at week 11 (n=43; all negative) giving a combined figure of 86%. For the 25 samples which were high-risk by both methods, the relationship exhibits general propor-

Figure 3. Comparison of minimal residual disease quantification by molecular and flow technologies. (*highlights 2 data points which are overlaid).

tionality with correlation, r= 0.76. Most of the 19 discordant samples were around the threshold level and for 8 of them, MRD was detectable by both techniques but did not attain the critical level of 0.01% by both assays.

Modern management of childhood ALL relies on risk stratification, individualizing treatment according to each child's risk of relapse and MRD assessment is considered to be the most sensitive and specific predictor of relapse. While it is more commonly measured by molecular technologies, MRD analysis by flow cytometry has gained interest in recent years but to date has largely been performed in single reference laboratories. The aim of this UKALL Flow MRD Laboratory Network study was to develop and validate a common protocol for MRD detection, and the data presented

here show that we have achieved a robust, quality controlled, sensitive method for MRD assessment in precursor B-cell ALL which can be replicated in a multi-center setting across different instrument and reagent platforms. It is applicable to almost 90% of children presenting with precursor B-cell ALL and shows good correlation with molecular analyses. The excellent inter-laboratory concordance for quality control exercises is attributable to all network laboratories participating in on-site training and regular workshops, accruing sufficient patient samples to maintain expertise and the availability of central data review. A recent study by Dworzak *et al.* also demonstrates that Flow MRD in childhood ALL can be standardized for reliable multicentric assessment.[11]

There are a number of advantages of quantifying MRD by flow cytometry over molecular methodologies.[12-15] It is quicker, with results available within hours of sampling and cheaper, because of reduced staff costs; analysis in the UK setting, suggests that a flow assay cost may be 70% that of PCR. The potential to increase this assay from 4 to 6 colors may further improve on speed and cost. Another important benefit is that the gating out of high side scatter cells may exclude apoptotic blasts which could potentially yield false positive results by PCR. In addition, in patients from whom bone marrow samples are not available, LAIPs can still be determined in peripheral blood with low blast counts, an option not available to molecular MRD techniques. Disadvantages include possible false-negatives due to immunophenotypic modulation induced by drug therapy or due to the selection of a minor sub-population of blasts.[16,17] To circumvent this, our standardized protocol does not rely on the overexpression of CD10 as a sole marker, avoids rigid gating and recommends the tracking of at least two LAIPs per patient. Flow MRD analysis with four colors cannot routinely reach the same level of sensitivity as molecular methodologies but attains the level of sensitivity required for risk strati-fication and has been shown to be predictive of outcome.

An approach taken by several groups is to use both flow and molecular methods in tandem, to ensure a MRD result for every patient.[18,19] Previous studies comparing MRD as assessed by four-color flow cytometric and PCR- based methodologies have shown high concordance but have been invariably performed in an ideal fashion i.e. on the same mononuclear cell preparation.[18-20] In the largest study to date, Neale *et al.* found a rate of concordance of 97% using a 0.01% MRD level, in a cohort of 1,300 peripheral blood and bone marrow samples.[19] In our study, we also show a good concordance rate of 86%, despite the fact that the samples for the two assays were separate aliquots of a bone marrow aspirate, which

were sent separately to the molecular and flow laboratories and then processed differently; mononuclear preparations for molecular and red cell lysis for flow. For approximately half of the discrepant results, MRD was detectable by both techniques but did not attain the critical level of 0.01% by both assays suggesting that this may be attributable to sample or methodological variability in borderline patients. One important observation is that 3 samples with highly discordant results and in which molecular MRD was significantly higher than flow, were classified as CD10 negative/weak. The reduced CD10 antigen expression can make gating analyses more difficult and further optimization of the standardized method may be necessary for this minor subgroup of patients. However, despite the good concordance between techniques, it does raise a clinical dilemma: for every 100 patients analyzed about 30 will be classified as high-risk by either technique but 10 of these 30 will be different children, depending on the methodology used.

Future plans for the UKALL Flow MRD network include the development of a quality assured protocol for T-lineage ALL prior to its incorporation into the next UKALL trial for newly diagnosed children where it may compliment molecular MRD detection. A valuable role of the flow method may be in the assessment of MRD at very early time points during remission induction therapy, to identify very good responders who may be candidates for early chemotherapeutic dose reduction.[21,22] The method may also be informative for risk stratification in relapsed ALL and pre- and post-stem cell transplant, and may also have a role in the adult setting.[23-25] A real strength of Flow MRD which has not yet been fully exploited is its capacity not only to quantify MRD but also to provide qualitative information about the residual leukemic blasts, e.g. to gain insight into *in vivo* drug resistance mechanisms such as expression/function of ABC transporters within the persisting leukemic cells, and would involve the expansion of this protocol from a 4 colour to a 5 or 6 colour assay. This approach may lead to therapies targeting the MRD blasts, as has been recently described by Dworzak *et al.*, for antiCD20-directed immunotherapy,[26] and may prove beneficial for MRD positive, high-risk children.

Authorship and Disclosures

All authors performed research and/or analyzed data. JI and JJ wrote the first draft of the manuscript and all other authors critically appraised it. SL devised and together with JJ, co-ordinated the study.

The authors reported no potential conflicts of interest.

References

1. Kersey JH. Fifty years of studies of the biology and therapy of childhood leukemia. Blood 1998;92:1838.
2. Cave H, van der Werff ten Bosch J, Suciu S, Guidal C, Waterkeyn C, Otten J, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer--Childhood Leukemia Cooperative Group. N Engl J Med 1998;339:591-8.
3. Ciudad J, San Miguel JF, Lopez-Berges MC, Vidriales B, Valverde B, Ocqueteau M, et al. Prognostic value of immunophenotypic detection of minimal residual disease in acute lymphoblastic leukemia. J Clin Oncol 1998;16:3774-81.
4. Coustan-Smith E, Sancho J, Hancock ML, Boyett JM, Behm FG, Raimondi SC, et al. Clinical importance of minimal residual disease in childhood acute lymphoblastic leukemia. Blood

2000;96:2691-6.

5. Biondi A, Valsecchi MG, Seriu T, D'Aniello E, Willemse MJ, Fasching K, et al. Molecular detection of minimal residual disease is a strong predictive factor of relapse in childhood B-lineage acute lymphoblastic leukemia with medium risk features. A case control study of the International BFM study group. Leukemia 2000;14:1939-43.

6. van Dongen JJ, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet 1998;352:1731-8.

7. Coustan-Smith E, Behm FG, Sanchez J, Boyett JM, Hancock ML, Raimondi SC, et al. Immunological detection of minimal residual disease in children with acute lymphoblastic leukaemia. Lancet 1998; 351:550-4.

8. Nyvold C, Madsen HO, Ryder LP, Seyfarth J, Svejgaard A, Clausen N, et al. Precise quantification of minimal residual disease at day 29 allows identification of children with acute lymphoblastic leukemia and an excellent outcome. Blood 2002;99: 1253-8.

9. Dworzak MN, Froschl G, Printz D, Mann G, Potschger U, Muhlegger N, et al. Prognostic significance and modalities of flow cytometric minimal residual disease detection in childhood acute lymphoblastic leukemia. Blood 2002;99:1952-8.

10. Barnett D, Granger V, Storie I, Peel J, Pollitt R, Smart T, et al. Quality assessment of CD34+ stem cell enumeration: experience of the United Kingdom National External Quality Assessment Scheme (UK NEQAS) using a unique stable whole blood preparation. Br J Haematol 1998; 102:553-65.

11. Dworzak MN, Gaipa G, Ratei R, Veltroni M, Schumich A, Maglia O, et al. Standardization of flow cytometric minimal residual disease evaluation in acute lymphoblastic leukemia: Multicentric assessment is feasible. Cytometry 2008;74:331-40.

12. Szczepanski T. Why and how to quantify minimal residual disease in acute lymphoblastic leukemia? Leukemia 2007;21:622-6.

13. Campana D. Minimal residual disease studies in acute leukemia. Am J Clin Pathol 2004;122 Suppl:S47-57.

14. van der Velden VH, Boeckx N, van Wering ER, van Dongen JJ. Detection of minimal residual disease in acute leukemia. J Biol Reg Homeost Agents 2004;18:146-54.

15. Campana D, Coustan-Smith E. Minimal residual disease studies by flow cytometry in acute leukemia. Acta Haematologica 2004;112:8-15.

16. Borowitz MJ, Pullen DJ, Winick N, Martin PL, Bowman WP, Camitta B. Comparison of diagnostic and relapse flow cytometry phenotypes in childhood acute lymphoblastic leukemia: implications for residual disease detection: a report from the children's oncology group. Cytometry 2005;68:18-24.

17. Gaipa G, Basso G, Maglia O, Leoni V, Faini A, Cazzaniga G, et al. Drug-induced immunophenotypic modulation in childhood ALL: implications for minimal residual disease detection. Leukemia 2005;19:49-56.

18. Kerst G, Kreyenberg H, Roth C, Well C, Dietz K, Coustan-Smith E, et al. Concurrent detection of minimal residual disease (MRD) in childhood acute lymphoblastic leukaemia by flow cytometry and real-time PCR. Br J Haematol 2005; 128:774-82.

19. Neale GA, Coustan-Smith E, Stow P, Pan Q, Chen X, Pui CH, et al. Comparative analysis of flow cytometry and polymerase chain reaction for the detection of minimal residual disease in childhood acute lymphoblastic leukemia. Leukemia 2004;18:934-8.

20. Malec M, van der Velden VH, Bjorklund E, Wijkhuijs JM, Soderhall S, Mazur J, et al. Analysis of minimal residual disease in childhood acute lymphoblastic leukemia: comparison between RQ-PCR analysis of Ig/TcR gene rearrangements and multicolor flow cytometric immunophenotyping. Leukemia 2004;18:1630-6.

21. Panzer-Grumayer ER, Schneider M, Panzer S, Fasching K, Gadner H. Rapid molecular response during early induction chemotherapy predicts a good outcome in childhood acute lymphoblastic leukemia. Blood 2000;95:790-4.

22. Coustan-Smith E, Sancho J, Behm FG, Hancock ML, Razzouk BI, Ribeiro RC, et al. Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic leukemia. Blood 2002;100:52-8.

23. Krampera M, Perbellini O, Vincenzi C, Zampieri F, Pasini A, Scupoli MT, et al. Methodological approach to minimal residual disease detection by flow cytometry in adult B-lineage acute lymphoblastic leukemia. Haematologica 2006;91:1109-12.

24. Coustan-Smith E, Gajjar A, Hijiya N, Razzouk BI, Ribeiro RC, Rivera GK, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. Leukemia 2004; 18:499-504.

25. Knechtli CJ, Goulden NJ, Hancock JP, Grandage VL, Harris EL, Garland RJ, et al. Minimal residual disease status before allogeneic bone marrow transplantation is an important determinant of successful outcome for children and adolescents with acute lymphoblastic leukemia. Blood 1998;92:4072-9.

26. Dworzak MN, Schumich A, Printz D, Potschger U, Husak Z, Attarbaschi A, et al. CD20 up-regulation in pediatric B-cell precursor acute lymphoblastic leukemia during induction treatment: setting the stage for anti-CD20 directed immunotherapy. Blood 2008;112: 3982-8.