Minimal residual disease analysis by eight-color flow cytometry in relapsed childhood acute lymphoblastic leukemia

Leonid Karawajew,[1] Michael Dworzak,[2] Richard Ratei,[3] Peter Rhein,[1] Giuseppe Gaipa,[4] Barbara Buldini,[5] Giuseppe Basso,[5] Ondrej Hrusak,[6] Wolf-Dieter Ludwig,[3] Günter Henze,[1] Karl Seeger,[1] Arend von Stackelberg,[1] Ester Mejstrikova,[6] and Cornelia Eckert[1]

[1]Department of Pediatric Oncology/Hematology, Charité Universitätsmedizin, Berlin, Germany; [2]St. Anna Children's Hospital and Children's Cancer Research Institute, Department of Pediatrics, Medical University of Vienna, Austria; [3]Robert-Roessle-Clinic in the HELIOS Klinikum Berlin, Germany; [4]Tettamanti Research Center, Department of Pediatrics, University of Milano-Bicocca, Ospedale San Gerardo, Monza, Italy; [5]Laboratory of Pediatric Onco-Hematology, Department of Pediatrics, University Hospital of Padova, Italy; [6]Department of Pediatric Hematology and Oncology, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic

ABSTRACT

Multiparametric flow cytometry is an alternative approach to the polymerase chain reaction method for evaluating minimal residual disease in treatment protocols for primary acute lymphoblastic leukemia. Given considerable differences between primary and relapsed acute lymphoblastic leukemia treatment regimens, flow cytometric assessment of minimal residual disease in relapsed leukemia requires an independent comprehensive investigation. In the present study we addressed evaluation of minimal residual disease by flow cytometry in the clinical trial for childhood relapsed acute lymphoblastic leukemia using eight-color flow cytometry. The major challenge of the study was to reliably identify low amounts of residual leukemic cells against the complex background of regeneration, characteristic of follow-up samples during relapse treatment. In a prospective study of 263 follow-up bone marrow samples from 122 patients with B-cell precursor acute lymphoblastic leukemia, we tested various B-cell markers, adapted the antibody panel to the treatment protocol, and evaluated its performance by a blinded parallel comparison with the polymerase chain reaction data. The resulting eight-color single-tube panel showed a consistently high overall concordance ($P<0.001$) and, under optimal conditions, sensitivity similar to that of the reference polymerase chain reaction method. Overall, evaluation of minimal residual disease by flow cytometry can be successfully integrated into the clinical management of relapsed childhood acute lymphoblastic leukemia either as complementary to the polymerase chain reaction or as an independent risk stratification tool. *ALL-REZ BFM 2002 clinical trial information: NCT00114348*

Introduction

Minimal residual disease (MRD) in bone marrow is a powerful predictor of clinical outcome in childhood acute lymphoblastic leukemia (ALL).[1-4] MRD evaluation is included in most treatment regimens for newly diagnosed ALL in which it yields an important risk group stratification parameter.[5-9] Multiple studies evaluating MRD continue to demonstrate the prognostic value in newly diagnosed pediatric ALL patients,[10-12] and there is increasing evidence that MRD is also of prognostic importance in relapsed ALL.[13-18] Initial reports on a relatively small numbers of relapsed patients[13-15] have been confirmed in a representative number of relapsed intermediate-risk ALL patients in the ALL-REZ P95/96 trial.[19] Eckert *et al.*[19] reported that MRD after the second induction course was the only parameter independently predicting the occurrence of subsequent relapse. The subsequent ALL-REZ BFM 2002 trial successfully integrated MRD status into the decision to allocate MRD good responders to chemotherapy and poor responders to hematopoietic stem cell transplantation.[16]

In the study by Eckert *et al.*[19] as well as in the majority of previous studies on front-line protocols from the ALL-BFM study group,[1,7,20-22] MRD was assessed using a method based on amplification of clonal antigen-receptor gene rearrangements by polymerase chain reaction (PCR), which has become the reference method in the ALL-BFM protocols.[7,22] The flow cytometric (FCM) analysis of bone marrow follow-up samples is an alternative approach to detect MRD.[23-26] Due to certain advantages/benefits of FCM-MRD monitoring (monitoring of higher numbers of patients, a faster turn-around time, less time- and labor-intensive, less expensive) this method has become increasingly important as a complementary technique to PCR or even a preferred approach in ALL protocols.[27] Detection of blasts by FCM is based on antigen expression differences between normal and malignant B-lineage cells. While normal mature B lymphocytes and B-cell progenitors (hematogones) reveal an ordered pattern of antigen expression on the cell surface, malignant cells may be identified by their leukemia-associated immunophenotypes, which are usually defined at diagnosis.[24,25,27] During therapy, however, the regeneration pattern of normal hematogones[28] and the initial leukemia-associated immunophenotypes of leukemic cells[29-33] may be considerably distorted. The major challenge of FCM-MRD analysis is, therefore, to distinguish

 doi:10.3324/haematol.2014.116707
The online version of this article has a Supplementary Appendix.
Manuscript received on September 5, 2014. Manuscript accepted on April 30, 2015.
Correspondence: leonid.karawajew@charite.de

leukemic cells from normal B-lineage subpopulations in bone marrow during treatment. While this can be more easily accomplished during times of continued chemotherapy and cytopenia, it is decisively more challenging in periods of hematopoietic regeneration. The performance of the FCM-MRD method depends, therefore, on the time points of application within a given treatment protocol, and the results of the FCM-MRD evaluation obtained within one study cannot be extrapolated directly to another trial if different treatment protocols are used.

In the front-line ALL protocols, extensive analyses of FCM-MRD and comparisons with PCR-MRD have been performed.[12,14,25,26,34-38] However, given considerable differences between front-line and relapsed ALL treatment regimens, the FCM-MRD in relapsed ALL and its relationship to PCR-MRD requires an independent comprehensive investigation. Moreover, FCM-MRD, being largely the methodology of detection of rare cells, is a technically challenging and rapidly developing field of flow cytometry.[25,27] While the majority of FCM-MRD studies have been performed using three- and four-color techniques, the recent reports on MRD data within front-line AIEOP-BFM ALL 2000 and DCOG-ALL10 protocols demonstrated improvement of the MRD analysis using six- and seven-color FCM.[35,39]

In the present study we addressed FCM-MRD in the ALL-REZ BFM 2002 relapse trial for childhood ALL using eight-color FCM and compared it with PCR-MRD in a prospective blinded study. We analyzed MRD levels by both methods in a total of 263 follow-up bone marrow samples from 122 patients with B-cell precursor (BCP)-ALL. During the study, we tested various antibody combinations, defined the protocol-adjusted antibody panel, and evaluated qualitative and quantitative concordance between FCM-MRD and PCR-MRD.

Methods

Patients and treatment

Patients in the presented study cohort had a first BCP-ALL relapse and were enrolled in the international trial ALL-REZ BFM 2002. Patients' samples were obtained in accordance with the informed consent guidelines of the local medical ethics committees. The research protocol for the assessment of MRD and the treatment protocol were approved by the local medical ethics committees. The clinical characteristics of the study cohort are presented in *Online Supplementary Table S1*. Standard-, intermediate- and high-risk groups and their treatment strategies are defined in the *Online Supplementary Material and Methods* and *Online Supplementary Table S2*.

Minimal residual disease assessment by polymerase chain reaction and flow cytometry

In order to analyze and optimize the FCM-MRD methodology, 263 follow-up bone marrow samples from 122 patients were assessed by FCM in parallel to PCR analysis at multiple time points during the treatment. Samples were collected and initially prepared uniformly as described in the *Online Supplementary Material and Methods*. PCR-MRD measurements were performed as described previously.[15,19,40]

FCM-MRD was assessed on the basis of standard protocols described for four-color FCM[36,41] and modified in order to fit the eight-color FCM. All tubes contained a cell-permeant nucleic acid dye Syto41, which allowed identification of nucleated cells. MRD was quantified as the percentage of leukemic cells within the nucleated cell gate. Antigen expression was quantified as mean fluorescence intensity (MFI).[42,43] MFI values were corrected for background staining using antigen-negative lymphocyte/normoblast subpopulations. In the absence of appropriate negative subpopulations (for the antigens Bcl-2, CD58, CD44), isotype- and fluorochrome-matched irrelevant monoclonal antibodies were used as negative controls. The eight-color FCM was performed using a BD CANTO II flow cytometer (Becton Dickinson, San Jose, CA, USA).

Protocol-specific optimization of flow cytometry assessment of minimal residual disease by antibody testing and comparison with polymerase chain reaction data

In order to determine the panel of monoclonal antibodies which would correspond to the specific features of the ALL-REZ BFM 2002 protocol, a series of measurements with differing monoclonal antibody combinations was performed in parallel with PCR-MRD in a blinded manner. During this testing phase, three retrospective comparisons of the FCM and PCR data were carried out, in order to get feedback on the quality of the FCM-MRD analysis. After the testing phase, the final panel was determined and validated in a subsequent series of measurements. The antibody tubes tested are shown in *Online Supplementary Table S3*. A list of antibodies, antibody clones, fluorochromes, and manufacturers is provided in *Online Supplementary Table S4*.

Statistical analysis

In the qualitative analysis of concordance between FCM and PCR results, a binary classification test was performed at different MRD cut-off levels. The analysis provided statistical measures of the performance of the FCM method in comparison to PCR as the reference method: sensitivity (% of concordant positive cases in relation to the total number of positive PCR-MRD cases), specificity (% of concordant negative cases in relation to the total number of negative PCR-MRD cases) and overall concordance (% of all concordant cases).

In the quantitative comparison, general associations of FCM and PCR data as continuous variables were evaluated using bivariate non-parametric Spearman correlation statistics. To compare the FCM-MRD and PCR-MRD methods more specifically, Bland-Altman analysis[44] was used.

All calculations, including box plot graphics and the basic operations for the Bland-Altman analysis (one-sample t-test and linear regression analysis), were performed with IBM SPSS Statistics 21 software (IBM Corporation, USA).

Results

Protocol-specific optimization of the eight-color flow cytometry panel

The distinctive feature of the follow-up MRD samples was the presence of normal hematogones (Figure 1), with a predominance of early B-cell precursors (regeneration burst, *Online Supplementary Figure S1*). Hematogones were present in the majority of bone marrow samples (73%) at various levels (0.001% - 49%/nucleated cells, mean 1.1%). In particular low amounts of $CD10^+$ leukemic cells could not be reliably identified against this complex background of regeneration using only the basic panel consisting of CD19, CD10, CD20, CD34, and CD45 antibodies (Figure 1), which have been proven to be indispensable in the

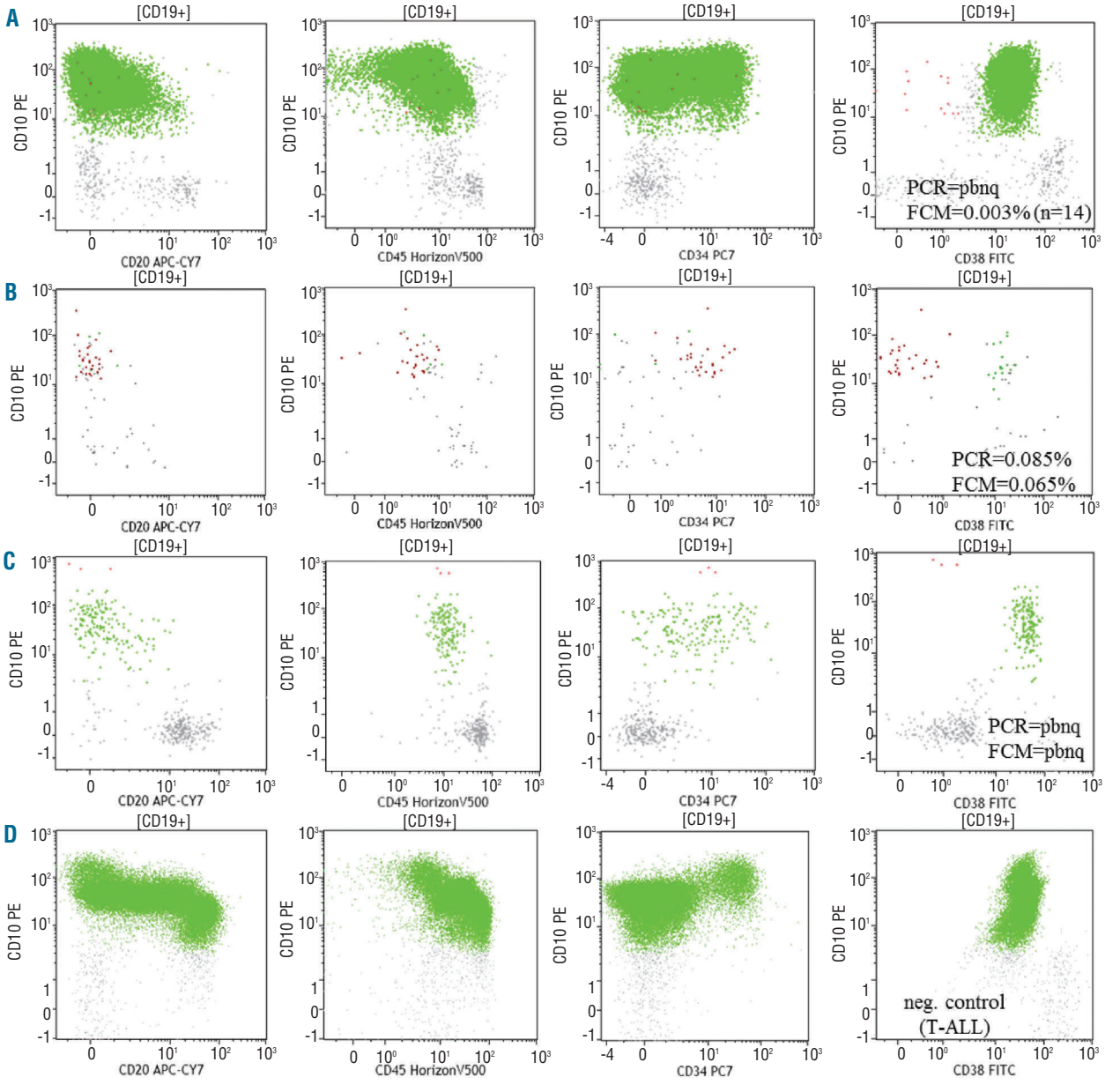

Figure 1. FCM analysis of MRD samples using the protocol-adjusted CD38-tube. All dot plots are pre-gated for Syto41- and CD19-positivities. (A-C) BCP-ALL samples with different levels of B-cell regeneration (green dots) and MRD (red dots). (D) T-ALL sample with a high level of B-cell regeneration (green dots) as cross-lineage negative control staining.

front-line FCM-MRD studies.[24,27,41,45] In order to optimize the monoclonal antibody panel, we tested several cell markers in addition to the basic combination of antibodies. The following surface antigens were tested: CD58 (the useful marker of blasts in the front-line protocol[46]), CD22 (the earliest B progenitor surface marker[42]), CD24 (B lineage marker), CD38 (a widely expressed hematopoietic antigen), and CD44, CD72, and CD86 (antigens with a potential to discriminate normal and leukemic progenitors, described by Coustan-Smith *et al.*[43]). In addition, cytoplasmic expression of TdT (normal and leukemic progenitors) and Bcl-2 (anti-apoptotic protein with an often increased expression in malignant cells[43,47]) were tested. The testing was performed using at least two antibody combinations (tubes). One tube, which contained the most established antibody combination of CD10, CD19, CD20, CD22, CD34, CD58, and CD45 in the ALL-BFM front-line protocol, was used in all cases. The second, experimental tube contained variable combinations of antibodies (*Online Supplementary Table S3*). Each antibody combination was tested in a series of 10 to 50 follow-up samples containing $CD10^+$ leukemic and/or hematogone subpopulations. Mean expression values are presented in *Online Supplementary Table S5* and the distributions of the measured values within sample series are depicted in the Figure 2. Expression of CD38, CD22, CD58, CD72 and Bcl-2 was statistically different between hematogones and leukemic cells (*Online Supplementary Table S5*). In particu-

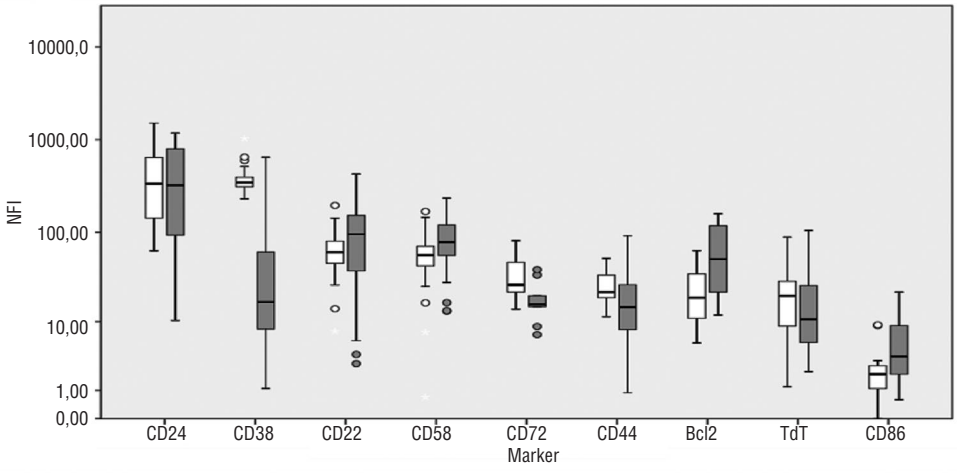

Figure 2. Box plot presentation of marker expression in hematogones (white) and BCP-ALL cells (grey). The markers are ordered in descending order of mean expression values in hematogones.

lar, there was a 6-fold difference of expression of CD38. Importantly, there was a strikingly low variability of CD38 expression in hematogones (Figure 2), which has been characteristic of all CD10-positive stages of B-cell regeneration. In the hematogone fraction, the "noise" background of the CD38 signal was extremely low and free from randomly stained cells (dot plots CD10 *versus* CD38 in Figure 1D). CD38 expression in leukemic cells (total MFI range: 647-1.1) was below the lowest value in hematogones (total MFI range: 1030-231) in the majority of samples (46 of 50, MFI range: 229-1.1). This analysis indicates that CD38 is the most promising marker for discriminating normal and leukemic $CD10^+$ B-cell progenitors in a single antibody tube setting. For the other antigens with statistically significant differences, the absolute fold difference varied between 1.5 and 2.9 (*Online Supplementary Table S5*) with largely overlapping distributions in normal and leukemic precursors (Figure 2). The usefulness of these markers, in particular Bcl-2 and CD72, in selected cases could not be excluded; however, a more detailed investigation would require a considerably higher number of cases, which would be difficult to achieve given that relapsed BCP-ALL is a rare disease.

For the $CD10^{-/low}$ BCP-ALL cases, which are considerably less frequent (27/263, 10.2%, in our MRD series), the discrimination of leukemic cells and $CD19^+CD10^-$ plasma cells is most important. Plasma cells demonstrated consistently high levels of CD38 expression (mean=1600 MFI; range, MFI 2511-906; n=20), significantly exceeding those of hematogones and leukemic cells (*Online Supplementary Table S5* and Figure 2). Together with the negativity of plasma cells for expression of CD22 and CD34 (*data not shown*), the CD38-tube panel was able to effectively discriminate $CD10^{-/low}$ BCP-ALL and plasma cells.

The CD22 antibody was included as the B-lineage commitment marker expressed at all stages of B-cell regeneration, in particular at the stage preceding CD19 expression.[42] These very early $CD19^-CD34^+CD22^+$ B-cell progenitors had a $CD10^+TdT^+$ immunophenotype (*Online Supplementary Figure S2*), and their presence in bone marrow correlated with the percentage of normal $CD19^+$ hematogones (*Online Supplementary Figure S3*). Already small amounts of these cells were indicative of the initial process of $CD19^+$ B-cell regeneration (*Online Supplementary Figure S3*).

With regard to the basic combination of CD10, CD19, CD20, CD34 and CD45 antibodies, we did not observe significant changes in expression during therapy, in contrast to what was reported with the front-line ALL-BFM protocol.[29-33] In particular, the aberrantly high expression of CD10 and CD34, if present at diagnosis, was helpful in detecting MRD (Figure 1C and *data not shown*).

As the final result we established an eight-color single-tube panel containing CD10, CD19, CD20, CD22, CD34, CD38, and CD45 antibodies and a Syto41 cell-permeant nucleic acid dye. This panel is subsequently referred to as the "CD38-tube". The preceding measurements will be considered together and referred to as the "CD58-tube".

CD38- and CD58-tube minimal residual disease series

The testing phase was performed using 159 samples and the CD38-tube was validated using 104 samples. The nucleated cell counts were $>10^5$ in the majority of cases (96%), and in 67% of cases the nucleated cell count was $>3x10^5$. In 11 cases (4%) the number of nucleated cells was below 10^5; nonetheless, we did not exclude these cases from the subsequent comparison with PCR, in order to keep the analysis close to routine conditions. Values of 10^6 and more were present in 21 cases (8%).

With regard to the most frequent aberrations, which were particularly instructive to distinguish leukemic and normal precursors, the expression of CD38 was lower on blasts than on hematogones in 67% of ALL cases. An aberrantly high expression of CD10 and CD34 was present in 38% and 17% of cases, respectively. Together, these aberrations were helpful in FCM-MRD assessement in 96% of cases. In the remaining cases an aberrantly low CD45 expression was helpful. We investigated the combination of CD58 and CD38 in a series of MRD samples (n=30) in parallel to the CD38-tube validation series; however this combination did not reveal any synergy in the identification of leukemic cells (*data not shown*).

In order to investigate potential detection limits of FCM-MRD in terms of absolute MRD cell counts, we set the criterion for quantifiable MRD positivity at the level of 10 cells or more. As an additional tool, for the cases with <10 blast cells, we introduced the term of "positive but not quantifiable" (pbnq) MRD in analogy to PCR. The distribution of MRD levels within the tested series is shown in Figure 3. There were considerably fewer MRD-negative

samples (35.6% *versus* 52.8%) and more positive samples at pbnq levels (14.4% *versus* 3.1%) using the CD38-tube in comparison with the CD58-tube (Figure 3). In order to exclude cohort-specific variability, we estimated the contributions of CD58 and CD38 in a retrospective analysis by omitting these antibodies in the interpretation of MRD data ("leave-one-out" analysis) (Figure 3). This analysis demonstrated that there were no cases in which CD58 contributed decisively to the identification of leukemic cells. In contrast, CD38 was indispensable in 15 out of 67 positive cases. The "leave CD38 out" MRD histogram shifted toward negative cases and, in particular at low MRD levels, the frequency of positive cases decreased (Figure 3).

Concordance analysis of flow cytometry and polymerase chain reaction minimal residual disease assessment

Using the PCR data as the reference, we first investigated the performance of FCM-MRD near to the detection limit of 10 cells (*Online Supplementary Table S6*). FCM-MRD was highly concordant with PCR-MRD down to the detection limit of 10 cells, while at the pbnq level the number of discordant cases increased significantly. This analysis, therefore, indicated the eligibility of the 10-cell detection limit.

Qualitative concordance between FCM and PCR data was analyzed at different MRD cut-off levels in terms of sensitivity, specificity and overall concordance of FCM in relation to PCR as the reference method (Table 1). At the 0.1% cut-off (the hematopoietic stem cell transplantation-relevant MRD level in the ALL-REZ BFM 2002 trial, *Online Supplementary Materials and Methods*), the FCM-MRD sensitivity was 74%, independently of the antibody panel (Table 1). At the 0.01% cut-off, which has been the usual criterion of MRD-positivity in the majority of reports,[12,35-38,41] the sensitivity increased, particularly for the CD38-tube (88% *versus* 80% for the CD58-tube). The difference in sensitivities between the antibody panels became consequently higher when samples were classified as MRD positive or negative with no percentage cut-off (90.8% *versus* 69.7% for the CD38- and CD58-tubes, respectively).

There was a considerable variance in the ratio of discordant FCM^-PCR^+ and FCM^+PCR^- samples between the CD58- and CD38-tube series when no cut-off was applied. In the CD58-tube series, the number of false negative (n=30) and false positive (n=6) cases was markedly shifted toward FCM^-/PCR^+, in accordance with the lower relative sensitivity of the FCM-MRD method reported previously.[35-38,41] In contrast, for the CD38-tube the number of FCM^-PCR^+ and FCM^+PCR^- cases was approximately equal (n=6 and n=8, respectively, Table 1).

The quantitative comparison of FCM and MRD is shown in Figure 4. The bivariate correlation of FCM and PCR was highly significant ($P<0.001$). The correlation coefficients were higher for the CD38-tube series than for the CD58-tube one (0.92 and 0.82, respectively). Similarly to the analysis of qualitative concordance at positive/negative MRD level (Table 1), the CD38-tube but not the CD58-tube series revealed similar numbers of FCM^{pbnq}/PCR^-, FCM^{pbnq}/PCR^{pbnq} and FCM^-/PCR^{pbnq} cases (Figure 4).

With regard to the specificity (Table 1), this parameter decreased in the CD38-tube series when no cut-off was applied, due to the increased number of the discordant PCR^-FCM^+ cases at a pbnq level. Although seemingly in contradiction with the improved performance of the CD38-tube panel, the decrease of specificity is in accor-

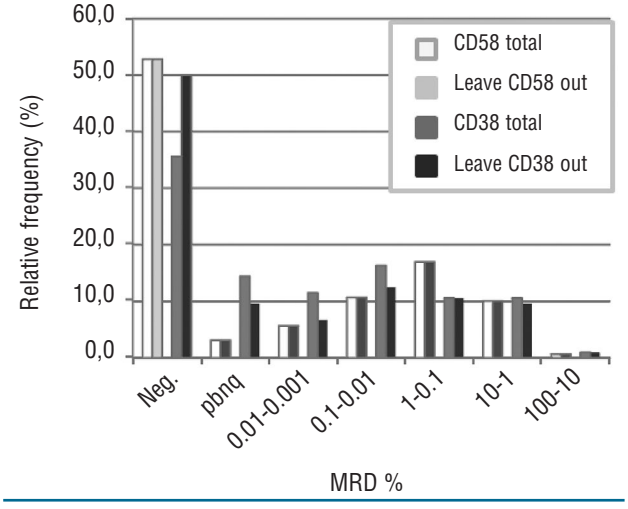

Figure 3. Histograms displaying the distribution of different MRD levels within testing series. The height of the bar (y-axis) corresponds to the relative frequency of the samples falling within the indicated MRD interval (x-axis). The series using experimental CD58-tubes comprised 159 samples, the series using the CD38-tube comprised 104 samples.

Table 1. Qualitative concordance of FCM-MRD in relation to PCR-MRD at different MRD cut-off levels and with no percentage cut-off.

FCM PCR	+ +	- +	+ -	- -	Sensitivity %	Specificity %	Concordance %
Cut-off: 0.1%							
"CD58-tube"	41	14	3	102	74.5	97.1	89.4
"CD38-tube"	23	8	0	72	74.2	100	92.2
Cut-off: 0.01%							
"CD58-tube"	56	14	5	85	80.0	94.4	88.1
"CD38-tube"	37	5	3	58	88.1	95.1	91.3
Positive / negative							
"CD58-tube"	69	30	6	54	69.7	90.0	77.4
"CD38-tube"	59	6	8	31	90.8	79.5	86.5

dance with similar detection limits of the FCM and PCR methods. In fact, as MRD levels in the samples approach the detection limit, the probability increases that the sample measured by one method will be a true-negative and by the other method a true-positive. In the *Online Supplementary Discussion* we consider this aspect in more detail.

Analysis of quantitative differences between flow cytometry and polymerase chain reaction minimal residual disease assessment

In spite of the high bivariate correlation, the majority of the points lay either on or below the 1:1 identity line, indicating lower FCM *versus* PCR values (Figure 4). This observation was statistically confirmed by Bland-Altman analysis, the algorithm used to compare two different methods of measurement.[44] After logarithmic transformation (applicable to the double positive quantifiable cases, n=104), the estimated differences of PCR and FCM values (LDIFF) were significantly different from zero in a one-sample t-test. The mean LDIFF was 0.41±0.05 (95% confidence interval 0.32–0.51), equivalent to a linear factor of 2.5 (Figure 5). The linear regression analysis demonstrated that LDIFF was not significantly dependent on the mean of logarithmic values of PCR-MRD and FCM-MRD (LMEAN), i.e. FCM-MRD and PCR-MRD differed over the whole range of MRD levels (Figure 5).

Since the background of hematogones may potentially interfere with FCM-MRD and PCR-MRD analyses, we performed Bland-Altman analysis for cohorts with different levels of the MRD and regeneration. If the cohorts were limited to the cases with a reliable number of detected blasts (MRD cells >100) and low regeneration (hematogones <100), the mean value of LDIFF did not change significantly (0.37±0.07, P<0.001, n=45). In the total absence of regeneration (<10 hematogones), the difference again did not change significantly (0.35±0.08, P<0.001, n=36). In accordance, there was no significant correlation of either PCR-MRD or FCM-MRD with hematogone levels in linear regression analysis (*data not shown*).

Furthermore, in order to check the impact of the sample quality reflected by nucleated cell percentage, we investigated the relation between LDIFF and nucleated cell values by linear regression analysis. There was no significant correlation between these parameters (*data not shown*).

One of the methodological differences between the two methods is that the FCM signal in follow-up samples does not depend quantitatively on the initial signal level, while PCR-MRD values are linked to the initial PCR signal in the diagnostic sample (quantitative presence of a major or small subclone). A potential error at diagnosis may, therefore, result in a consistent deviation from the true MRD value, which could be potentially detected in the follow-up series for the same patient (intrapatient MRD series). In the total cohort of 122 patients, there were 38 patients with three or more follow-up measurements. In 15 out of these 38 patients, the intrapatient MRD series consisted of at least three positive time points, thus allowing statistical analysis by one-sample t-test and the estimation of the mean LDIFF values. As a result, the series could be classified into three groups with different mean LDIFF values: highly concordant (LDIFF=0.0±0.1, n=4), highly discordant (≥1.0, n=3) and intermediate (0.20-0.85, n=8). The PCR *versus* FCM data for the highly concordant and highly discordant groups are listed in Table 2. The FCM re-analysis of diagnostic and MRD samples from discordant series did not reveal immunophenotypic heterogeneity of the leukemic cell population or misinterpretation of the data. We also re-analyzed PCR performance in the intrapatient MRD series with highly discordant and concordant FCM-

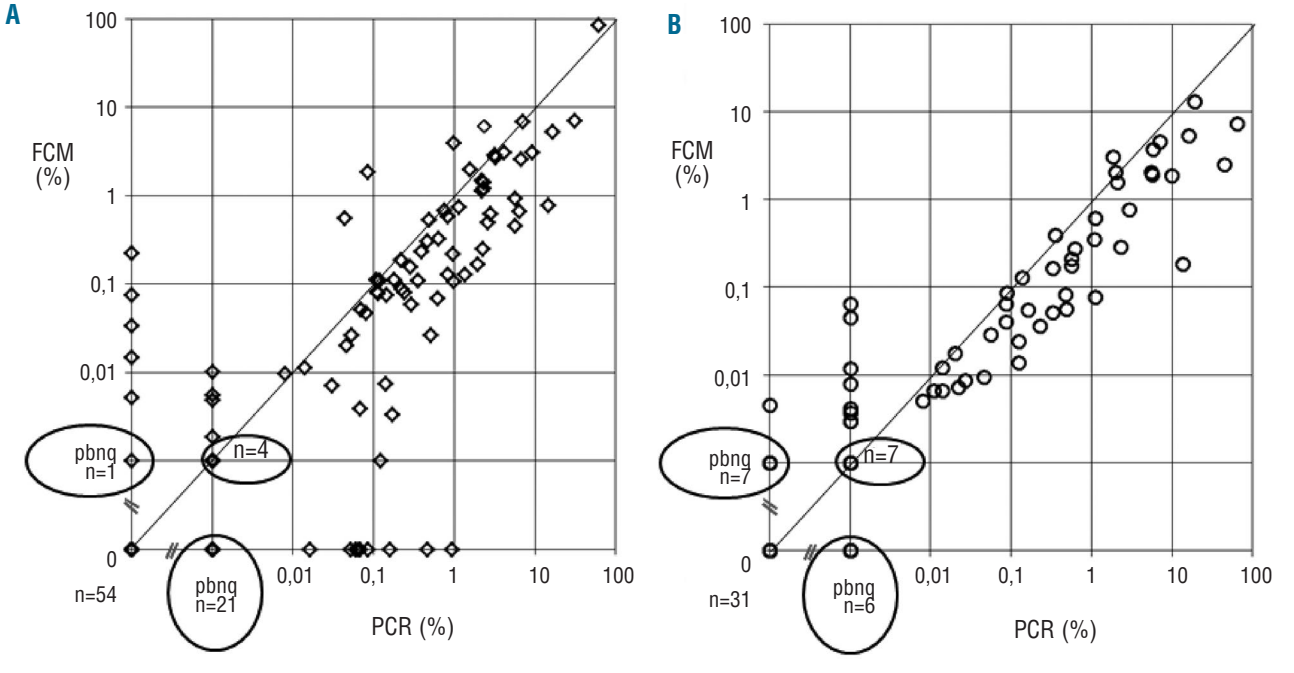

Figure 4. Quantitative comparison of MRD estimates by FCM and PCR using experimental CD58-tubes (A) and the protocol-adjusted CD38-tube (B). The diagonal is the 1:1 identity line.

Table 2. Intrapatient MRD series with low (A) and high (B) levels of difference between PCR-MRD and FCM-MRD (mean LDIFF).

A	Patient # 1		Patient # 2		Patient # 3		Patient # 4	
timepoint	PCR-MRD (%)	FCM-MRD (%)	PCR-MRD (%)	FCM-MRD (%)	PCR-MRD (%)	FCM-MRD (%)	PCR-MRD (%)	FCM-MRD (%)
1	0.11	0.11	1.0	3.9	0.49	0.53	2.0	2.0
2	0.05	0.02	1.5	1.9	0.63	0.32	0.35	0.39
3	0.008	0.009	2.3	1.2	0.12	0.11	1.8	3.0
4	n.a.	n.a.	3.1	2.9	n.a.	n.a.	5.7	3.7
mean LDIFF	-0.1	0.0	0.1	0.0				
t-test*	n.s.	n.s.	n.s.	n.s.				

B	Patient # 13		Patient # 14		Patient # 15	
timepoint	PCR-MRD (%)	FCM-MRD (%)	PCR-MRD (%)	FCM-MRD (%)	PCR-MRD (%)	FCM-MRD (%)
1	0.6	0.06	6.3	0.7	60	7.2
2	1.9	0.17	31	7.0	34	2.5
3	2.2	0.25	17	0.8	1.4	0.20
mean LDIFF	1.0	1.0	1.0			
t-test*	$P<0.05$	$P<0.05$	$P<0.05$			

** one-sample t-test for the difference of log(PCR-MRD) and log(FCM-MRD) / n.s = not significant.*

MRD and PCR-MRD levels (*Online Supplementary Table S7*). In the discordant series, samples from one patient (#13) revealed two clones with different Ct values at diagnosis and different kinetics of response to treatment. Using the smaller subclone as the major one may have resulted in over-interpretation of quantitative results. However, there was no evidence for the presence of subclones which could explain the FCM/PCR differences in the patients #14 and #15. In the concordant series there were no cases with more than one subclone (*Online Supplementary Table S7*).

Flow cytometry minimal residual disease assessment at the time point after induction treatment

MRD after the completion of induction therapy was used for risk stratification (indication for hematopoietic stem cell transplantation) in the relapse trial ALL-REZ BFM 2002.[19] Of 63 cases assessed at this time point, 40 patients had PCR-MRD $<10^{-3}$ and consequently were not allocated to transplantation after consolidation treatment. In this cohort, there was only one case which showed MRD $>10^{-3}$ by FCM. The number of cases which could, therefore, potentially provide information on an independent clinical impact of FCM-MRD was too low for statistical analysis. Methodologically, the low number of the PCR-MRD $<10^{-3}$ / FCM-MRD $\geq 10^{-3}$ cases is in accordance with the overall shift of the LDIFF toward positive values (Figure 5).

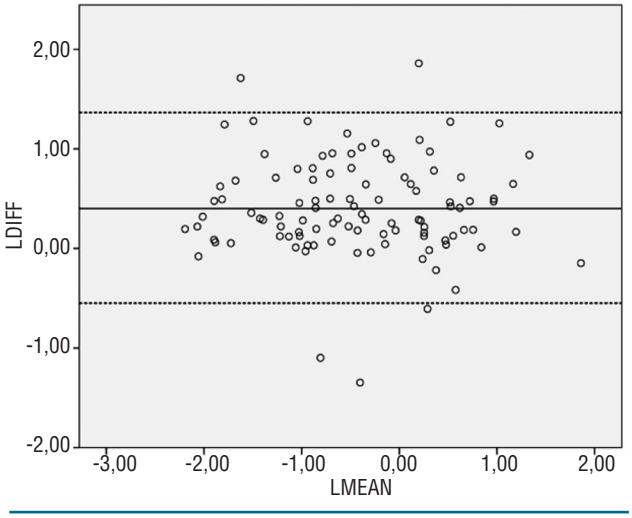

Figure 5. Bland-Altman plot of the difference of PCR-MRD and FCM-MRD (LDIFF) against the mean of the PCR-MRD and FCM-MRD (LMEAN) after logarithmic transformation. The lines indicate the estimated mean LDIFF (continuous line) and the upper and lower limits of agreement, mean LDIFF ± 1.96 SD (dotted lines).

Discussion

In the present study, we report on MRD detection of leukemic cells in bone marrow samples from patients with first ALL relapse enrolled in and treated according to the relapse trial ALL-REZ BFM using eight-color FCM and its relation to PCR-MRD detection. During the study a protocol-specific adaptation of the antibody panel was performed, starting with the panel established in the front-line ALL-BFM 2000 protocol. With the latter protocol, considerable changes of leukemia-associated immunophenotypes were frequently observed in MRD cells after prednisolone treatment prephase[29-32] with a genome-wide shift toward an immunophenotype of normal B cells.[33] During induction therapy in the ALL-BFM 2000 protocol hematogones were usually absent in bone marrow, but appeared at day 78 after treatment cessation and the resting phase showing the "classical" B-cell regeneration pattern with precursors distributed over different maturation stages in synchronous proportions.[28,42]

In comparison to the front-line regimen the ALL-REZ BFM 2002 protocol is characterized by more intensive chemotherapy blocks. Consequently, the patients' bone marrow becomes aplastic and needs rest periods after treatment for the recovery of hematopoiesis. In our study,

hematogones were present in the majority of samples, and the appearance of the regeneration pattern was frequently asynchronous with a predominance of early hematogone populations and absence of late ones, in accordance with the previous study on post-chemotherapy B lymphopoiesis.[28] Of the candidate markers tested during the FCM-MRD optimization phase, the CD38 antigen revealed unique expression on normal hematogones at multiple stages of normal regeneration. Although CD38 has been proposed in a larger panel for MRD,[25,41,48] its prominent role in the identification of normal B-cell precursors during therapy has not yet been assessed in detail.

We further included CD22 antibody into the protocol-adjusted CD38 panel. In addition to providing an independent parameter indicative of early B-cell regeneration (*Online Supplementary Figures S2 and S3*), CD22 monoclonal antibody may potentially substitute the CD19 antibody as a pan-B-cell marker for monitoring the effect of therapeutic interventions targeting the CD19 antigen (e.g. blinatumomab).[49] Moreover, given that the CD22 antigen is a promising target of antibody-based treatment strategies,[49,50] which has been included in the international trial for relapsed ALL that follows the ALL-REZ BFM 2002 (IntReALL-2010, NCT01802814, http://clinicaltrials.gov/show/NCT01802814), the presence of CD22 in the antibody panel would enable permanent expression monitoring of this target in MRD samples. The CD22 antibody, therefore, provides a reasonable enhancement of the range of application of the single CD38-tube panel.

The minimum detectable signal level in FCM depends on the ratio between the levels of specific and unspecific (noise) signals. In the three- to four-color FCM-MRD studies,[34-38] the clustering of a sufficient number (n=30-50) of positive signals was required as an additional parameter to distinguish specific and randomly stained cells. With increasing numbers of simultaneously registered FCM parameters, the unspecific background can be more efficiently reduced since the signals from non-leukemic cells can be gated out better. In the present study, the application of the optimized eight-color tube allowed the criterion for positivity to be reduced down to 10 cells. In addition, this panel allowed the number of tubes to be reduced to a single one. Correspondingly, the number of acquired events could be increased, making the acquisition of $1x10^6$ nucleated cells realistic in the routine diagnostic setting. The minimal number of positive events (n=10) and the number of acquired nucleated cells ($n=10^6$) would, therefore, result in a potential detection limit of 10^{-5}. Given that the measurements in PCR-MRD assessment are usually made on 10^5 cells and that the minimum detectable PCR signal level is, at best, 1 cell/well, the relative sensitivities of both methods under optimal conditions would become comparable at the level of 10^{-5}. It should be noted that optimal conditions imply, in addition to the sufficient number of acquired events, the reliable immunophenotypic discrimination of normal and leukemic progenitors. In this regard, the applicability of every antibody panel is limited, including the antibody panel used in the present study, and additional potential antigen candidates do, therefore, warrant further investigation.

Over the whole MRD range, FCM and PCR values showed high levels of bivariate correlation as in other studies.[12,14,34-39] In absolute values, however, FCM-MRD levels were statistically lower than PCR-MRD values by a factor of 2.5. Of interest, Gaipa *et al.*[35] reported a 2.9-fold higher mean MRD level by PCR as compared to seven-color FCM-MRD in 266 follow-up bone marrow samples in the front-line ALL-BFM study. Overall, differences between quantitative FCM-MRD and PCR-MRD have been reported in the majority of publications addressing this issue.[12,14,34-39] In addition, van der Velden *et al.*[22] reported an intra-assay variability of PCR-MRD of about 3-fold, i.e. in a similar range as the observed variance between PCR and FCM.

Nonetheless, the consistent quantitative shift towards lower FCM levels remains to be explained. Several sources of quantitative PCR-MRD discordance have been proposed, such as variability of biological material, lower FCM sensitivity, high regeneration background, and presence of dead cells.[12,14,25,35-39] We addressed these potential sources of discordance experimentally. Given the uniform pretreatment and the use of sample splitting, the influence of sample variability could be largely excluded. Our data indicate that despite improved sensitivity the quantitative discrepancy remains significant. Importantly, the statistically significant LDIFF level was also independent of the level of regeneration background and of the percentage of nucleated cells. In addition to these commonly discussed aspects, our data allowed the analysis of intra-patient follow-up series with a sufficient number of positive MRD time points. These data suggested that the quantitative discrepancy is not an inherent feature of the methodologies valid for all MRD samples, but there are cases with a consistently high concordance as well as those with a consistently high discordance of FCM *versus* PCR results. The observed consistency of discordance throughout some MRD series indicates that at least in some cases the source of discrepancy may originate from use of the PCR-MRD marker presenting a smaller subclone at diagnosis which might result in an over-interpretation of MRD. For discordant cases with no evidence of subclones which could explain the FCM/PCR differences, the picture might become clearer with a more comprehensive analysis of clonal T-cell receptor/immunoglobulin gene rearrangements using targeted next-generation sequencing.

In conclusion, this is the first report on the application of eight-color FCM for MRD assessment and its relation to PCR-MRD detection in relapsed ALL. The study performed within the ALL-REZ BFM 2002 clinical trial involved a parallel, prospective assessment and blinded comparison of FCM-MRD and PCR-MRD in samples acquired from a representative number of uniformly treated patients. During the study we tested various antibody combinations, defined and validated the protocol-adjusted antibody panel, and evaluated qualitative and quantitative concordance of FCM-MRD and PCR-MRD assessments. The resulting single-tube panel had a sensitivity similar to that of PCR-MRD, although a high number of acquired cells was not reached routinely during the study, due to the preferential application of PCR. The separate acquisition of bone marrow samples for FCM-MRD in the *IntReALL-2010* clinical trial will overcome this shortage. Overall, our data show that FCM-MRD assessment can be successfully integrated into the clinical management of relapsed childhood ALL either independently or as a complement to PCR-MRD risk stratification.

Acknowledgments

We thank Grit Czerwony, Marianne Dunken, Karin Berlin, Wilhelmine Keune, Marion Fillies and the technicians of the molecular genetics laboratory of the ALL-REZ BFM trials for excellent technical assistance. We also thank Deutsche Kinderkrebsstiftung for its funding support of the ALL-REZ BFM 2002 clinical trial. This research has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 278514 – IntReALL

Funding

This work was supported by the Deutsche José Carreras Leukämie-Stiftung, grants DJCLS R09/19 and R11/21; Alfred & Angelika Gutermuth-Stiftung; Deutsche Kinderkrebsstiftung, grant DKS 2013.12; Ministry of Health of the Czech Republic, grant NT 13462-4 (to EM).

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Biondi A, Valsecchi MG, Seriu T, 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(11):1939-1943.
2. Borowitz MJ, Devidas M, Hunger SP, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. Blood. 2008;111(12):5477-5485.
3. Cave H, van der Werff ten Bosch J, Suciu S, 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(9):591-598.
4. Zhou J, Goldwasser MA, Li A, et al. Quantitative analysis of minimal residual disease predicts relapse in children with B-lineage acute lymphoblastic leukemia in DFCI ALL Consortium Protocol 95-01. Blood. 2007;110(5):1607-1611.
5. Basso G, Veltroni M, Valsecchi MG, et al. Risk of relapse of childhood acute lymphoblastic leukemia is predicted by flow cytometric measurement of residual disease on day 15 bone marrow. J Clin Oncol. 2009;27(31):5168-5174.
6. Borowitz MJ, Pullen DJ, Shuster JJ, et al. Minimal residual disease detection in childhood precursor-B-cell acute lymphoblastic leukemia: relation to other risk factors. A Children's Oncology Group study. Leukemia. 2003;17(8):1566-1572.
7. Flohr T, Schrauder A, Cazzaniga G, et al. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. Leukemia. 2008;22(4):771-782.
8. Van der Velden VH, Corral L, Valsecchi MG, et al. Prognostic significance of minimal residual disease in infants with acute lymphoblastic leukemia treated within the Interfant-99 protocol. Leukemia. 2009;23(6):1073-1079.
9. Yamaji K, Okamoto T, Yokota S, et al. Minimal residual disease-based augmented therapy in childhood acute lymphoblastic leukemia: a report from the Japanese Childhood Cancer and Leukemia Study Group. Pediatr Blood Cancer. 2010;55(7):1287-1295.
10. Conter V, Bartram CR, Valsecchi MG, et al. Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. Blood. 2010;115(16):3206-3214.
11. Mussolin L, Pillon M, Conter V, et al. Prognostic role of minimal residual disease in mature B-cell acute lymphoblastic leukemia of childhood. J Clin Oncol. 2007;25(33):5254-5261.
12. Garand R, Beldjord K, Cave H, et al. Flow cytometry and IG/TCR quantitative PCR for minimal residual disease quantitation in acute lymphoblastic leukemia: a French multicenter prospective study on behalf of the FRALLE, EORTC and GRAALL. Leukemia. 2013;27(2):370-376.
13. Raetz EA, Borowitz MJ, Devidas M, et al. Reinduction platform for children with first marrow relapse of acute lymphoblastic Leukemia: a Children's Oncology Group Study[corrected]. J Clin Oncol. 2008;26(24):3971-3978.
14. Coustan-Smith E, Gajjar A, Hijiya N, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. Leukemia. 2004;18(3):499-504.
15. Eckert C, Biondi A, Seeger K, et al. Prognostic value of minimal residual disease in relapsed childhood acute lymphoblastic leukaemia. Lancet. 2001;358(9289):1239-1241.
16. Eckert C, Henze G, Seeger K, et al. Use of allogeneic hematopoietic stem-cell transplantation based on minimal residual disease response improves outcomes for children with relapsed acute lymphoblastic leukemia in the intermediate-risk group. J Clin Oncol. 2013;31(21):2736-2742.
17. Paganin M, Zecca M, Fabbri G, et al. Minimal residual disease is an important predictive factor of outcome in children with relapsed 'high-risk' acute lymphoblastic leukemia. Leukemia. 2008;22(12):2193-2200.
18. von Stackelberg A, Seeger K, Henze G, Eckert C. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. Leukemia. 2004;18(10):1727-1728; author reply 1728-1729.
19. Eckert C, von Stackelberg A, Seeger K, et al. Minimal residual disease after induction is the strongest predictor of prognosis in intermediate risk relapsed acute lymphoblastic leukaemia - long-term results of trial ALL-REZ BFM P95/96. Eur J Cancer. 2013;49(6):1346-1355.
20. Attarbaschi A, Mann G, Panzer-Grumayer R, et al. Minimal residual disease values discriminate between low and high relapse risk in children with B-cell precursor acute lymphoblastic leukemia and an intrachromosomal amplification of chromosome 21: the Austrian and German Acute Lymphoblastic Leukemia Berlin-Frankfurt-Munster (ALL-BFM) trials. J Clin Oncol. 2008;26(18):3046-3050.
21. Steinemann D, Cario G, Stanulla M, et al. Copy number alterations in childhood acute lymphoblastic leukemia and their association with minimal residual disease. Genes Chromosomes Cancer. 2008;47(6):471-480.
22. van der Velden VH, Panzer-Grumayer ER, Cazzaniga G, et al. Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. Leukemia. 2007;21(4):706-713.
23. Campana D. Progress of minimal residual disease studies in childhood acute leukemia. Curr Hematol Malig Rep. 2010;5(3):169-176.
24. Campana D. Minimal residual disease monitoring in childhood acute lymphoblastic leukemia. Curr Opin Hematol. 2012;19(4):313-318.
25. Gaipa G, Basso G, Biondi A, Campana D. Detection of minimal residual disease in pediatric acute lymphoblastic leukemia. Cytometry B Clin Cytom. 2013;84(6):359-369.
26. Dworzak MN, Froschl G, Printz D, et al. Prognostic significance and modalities of flow cytometric minimal residual disease detection in childhood acute lymphoblastic leukemia. Blood. 2002;99(6):1952-1958.
27. Wood BL. Flow cytometric monitoring of residual disease in acute leukemia. Methods Mol Biol. 2013;999:123-136.
28. Dworzak MN, Fritsch G, Fleischer C, et al. Multiparameter phenotype mapping of normal and post-chemotherapy B lymphopoiesis in pediatric bone marrow. Leukemia. 1997;11(8):1266-1273.
29. Dworzak MN, Gaipa G, Schumich A, et al. Modulation of antigen expression in B-cell precursor acute lymphoblastic leukemia during induction therapy is partly transient: evidence for a drug-induced regulatory phenomenon. Results of the AIEOP-BFM-ALL-FLOW-MRD-Study Group. Cytometry B Clin Cytom. 2010;78(3):147-153.
30. Dworzak MN, Schumich A, Printz D, 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(10):3982-3988.
31. Gaipa G, Basso G, Maglia O, et al. Drug-induced immunophenotypic modulation in childhood ALL: implications for minimal residual disease detection. Leukemia.

2005;19(1):49-56.
32. Rhein P, Mitlohner R, Basso G, et al. CD11b is a therapy resistance- and minimal residual disease-specific marker in precursor B-cell acute lymphoblastic leukemia. Blood. 2010;115(18):3763-3771.
33. Rhein P, Scheid S, Ratei R, et al. Gene expression shift towards normal B cells, decreased proliferative capacity and distinct surface receptors characterize leukemic blasts persisting during induction therapy in childhood acute lymphoblastic leukemia. Leukemia. 2007;21(5):897-905.
34. Campana D, Neale GA, Coustan-Smith E, Pui CH. Detection of minimal residual disease in acute lymphoblastic leukemia: the St Jude experience. Leukemia. 2001;15(2):278-279.
35. Gaipa G, Cazzaniga G, Valsecchi MG, et al. Time point-dependent concordance of flow cytometry and real-time quantitative polymerase chain reaction for minimal residual disease detection in childhood acute lymphoblastic leukemia. Haematologica. 2012; 97(10):1582-1593.
36. Irving J, Jesson J, Virgo P, et al. 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. Haematologica. 2009;94(6):870-874.
37. Malec M, van der Velden VH, Bjorklund E, 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(10):1630-1636.
38. Neale GA, Coustan-Smith E, Stow P, 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(5):934-938.
39. Denys B, van der Sluijs-Gelling AJ, Homburg C, et al. Improved flow cytometric detection of minimal residual disease in childhood acute lymphoblastic leukemia. Leukemia. 2013;27(3):635-641.
40. Eckert C, Landt O, Taube T, et al. Potential of LightCycler technology for quantification of minimal residual disease in childhood acute lymphoblastic leukemia. Leukemia. 2000;14(2):316-323.
41. Dworzak MN, Gaipa G, Ratei R, et al. Standardization of flow cytometric minimal residual disease evaluation in acute lymphoblastic leukemia: multicentric assessment is feasible. Cytometry B Clin Cytom. 2008;74(6):331-340.
42. Lucio P, Parreira A, van den Beemd MW, et al. Flow cytometric analysis of normal B cell differentiation: a frame of reference for the detection of minimal residual disease in precursor-B-ALL. Leukemia. 1999;13(3):419-427.
43. Coustan-Smith E, Song G, Clark C, et al. New markers for minimal residual disease detection in acute lymphoblastic leukemia. Blood. 2011;117(23):6267-6276.
44. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-310.
45. Dworzak MN. Immunological detection of minimal residual disease in acute lymphoblastic leukemia. Onkologie. 2001;24(5): 442-448.
46. Veltroni M, De Zen L, Sanzari MC, et al. Expression of CD58 in normal, regenerating and leukemic bone marrow B cells: implications for the detection of minimal residual disease in acute lymphocytic leukemia. Haematologica. 2003;88(11):1245-1252.
47. Hartung L, Bahler DW. Flow cytometric analysis of BCL-2 can distinguish small numbers of acute lymphoblastic leukaemia cells from B-cell precursors. Br J Haematol. 2004;127(1):50-58.
48. Lucio P, Gaipa G, van Lochem EG, et al. BIOMED-I concerted action report: flow cytometric immunophenotyping of precursor B-ALL with standardized triple-stainings. BIOMED-1 Concerted Action Investigation of Minimal Residual Disease in Acute Leukemia: International Standardization and Clinical Evaluation. Leukemia. 2001;15(8): 1185-1192.
49. Hoelzer D. Targeted therapy with monoclonal antibodies in acute lymphoblastic leukemia. Curr Opin Oncol. 2013;25(6):701-706.
50. Raetz EA, Cairo MS, Borowitz MJ, et al. Chemoimmunotherapy reinduction with epratuzumab in children with acute lymphoblastic leukemia in marrow relapse: a Children's Oncology Group pilot study. J Clin Oncol. 2008;26(22):3756-3762.