**Blood First Edition Paper, prepublished online November 30, 2016; DOI 10.1182/blood-2016-07-726307**

# Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia

Prisca Theunissen[1#], Ester Mejstrikova[2#], Lukasz Sedek[3], Alita J. van der Sluijs-Gelling [4], Giuseppe Gaipa[5], Marius Bartels[6], Elaine Sobral da Costa[7], Michaela Kotrová[2], Michaela Novakova[2,6], Edwin Sonneveld[4], Chiara Buracchi[5], Paola Bonaccorso[5], Elen Oliveira[7], Jeroen G. te Marvelde[1], Tomasz Szczepanski[3], Ludovic Lhermitte[8], Ondrej Hrusak[2], Quentin Lecrevisse[9], Georgiana Emilia Grigore[10], Eva Froňková[2], Jan Trka[2], Monika Brüggemann[6], Alberto Orfao[9], Jacques J.M. van Dongen[1], and Vincent H.J. van der Velden[1], on behalf of the EuroFlow Consortium

*[1]Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; [2]CLIP - Childhood Leukaemia Investigation Prague, Department of Paediatric Haematology and Oncology, Second Faculty of Medicine, Charles University, University Hospital Motol, Prague, Czech Republic; [3]Department of Pediatric Hematology and Oncology, Zabrze, Medical University of Silesia (SUM), Katowice, Poland; [4]Dutch Childhood Oncology Group, The Hague, Netherlands; [5]Centro Ricerca Tettamanti, Clinica Pediatrica Università di Milano Bicocca, Monza (MB),Italy; [6]Department of Hematology, University of Schleswig-Holstein, Campus Kiel, Kiel, Germany; [7]Department of Pediatrics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil; [8] Department of Hematology, Hôpital Necker-Enfants-Malades (AP-HP) and UMR CNRS 8147, University of Paris Descartes, Paris, France; [9]Cancer Research Center (IBMCC-CSIC), Department of Medicine and Cytometry Service, University of Salamanca (USAL) and Institute of Biomedical Research of Salamanca (IBSAL), Salamanca, Spain; [10]Cytognos SL, Salamanca, Spain.*

[#] PT and EM equally contributed

**Correspondence**: Prof. J.J.M. van Dongen, MD, PhD
Department of Immunology, Erasmus MC
University Medical Center Rotterdam
Wytemaweg 80
3015 CN Rotterdam
The Netherlands
Tel: +31 10 7044094
Fax: +31 10 704 4731
email: J.J.M.van_Dongen@lumc.nl

**Short tile:** flowcytometric MRD analysis in BCP-ALL

**Keywords:** BCP-ALL, MRD, immunophenotyping, PCR, NGS

**Word count text:** 3999 (max 4000)

**Word count abstract:** 227 (max 250)

**Figures:** 5

**Tables:** 3

**Reference count:** 47

**Supplemental Data:** present

**Key points:**

- Standardized flow cytometry allows highly sensitive MRD measurements in virtually all BCP-ALL patients
- If sufficient cells are measured (>4 million), flowcytometric MRD analysis is at least as sensitive as current PCR-based MRD methods

**Abstract**

A fully-standardized EuroFlow 8-color antibody panel and laboratory procedure was stepwise designed to measure minimal residual disease (MRD) in B-cell precursor (BCP) acute lymphoblastic leukemia (ALL) patients with a sensitivity of ≤$10^{-5}$, comparable to real-time quantitative (RQ)-PCR-based MRD detection via antigen-receptor rearrangements.

Leukocyte markers and the corresponding antibodies and fluorochromes were selected based on their contribution in separating BCP-ALL cells from normal/regenerating BCP cells in multidimensional principal component analyses. After five multicenter design-test-evaluate-redesign phases with a total of 319 BCP-ALL patients at diagnosis, two 8-color antibody tubes were selected, which allowed separation between normal and malignant BCP cells in 99% of studied patients. These two tubes were tested with a new erythrocyte bulk-lysis protocol allowing acquisition of high cell numbers in 377 bone marrow follow-up samples of 178 BCP-ALL patients. Comparison with RQ-PCR-based MRD data showed a clear positive relation between the percentage concordant cases and the number of cells acquired. For those samples with >4 million cells acquired, concordant results were obtained in 93% of samples. Most discordances were clarified upon high-throughput sequencing of antigen-receptor rearrangements and blind multicenter re-analysis of flowcytometric data, resulting in an unprecedented concordance of 98% (97% for samples with MRD <0.01%).

In conclusion, the fully-standardized EuroFlow BCP-ALL MRD strategy is applicable in >98% of patients with sensitivities at least similar to RQ-PCR (≤$10^{-5}$), if sufficient cells (>4x$10^{6}$, preferably more) are evaluated.

**Word count:** 227 (max: 250)

## Introduction

Most current treatment protocols for B-cell precursor (BCP) acute lymphoblastic leukemia (ALL) include minimal residual disease (MRD) measurements, generally based on PCR analysis of rearranged antigen receptor genes[1-3]. Although flow cytometry (FCM) can be used for MRD detection as well[4-9], studies so far indicate that the specificity and sensitivity of FCM-MRD diagnostics is inferior to PCR-based MRD diagnostics[10-13]. Nevertheless, we and others have recently shown that the use of 6- or 7-color immunostainings combined with the introduction of new markers and new marker combinations significantly improved FCM-MRD analysis in BCP-ALL patients[10, 12]. These improvements were particularly related to specificity, whereas the sensitivity still appeared to be lower than for the PCR-based methods. To further improve FCM-based MRD diagnostics, more objective and efficient discrimination of BCP-ALL cells from normal BCP cells and improved sample preparation procedures for acquisition of larger numbers of cells are a prerequisite.

Eight-color immunostainings may contribute to improve flowcytometric MRD detection in BCP-ALL patients. Recently, an 8-color antibody tube was developed in the ALL-REZ-BFM 2002 trial[14]. This tube contained seven antibodies (CD10, CD19, CD20, CD22, CD34, CD45, CD38) and the nucleic acid dye Syto41 and gave concordant MRD results with PCR-MRD data in 86.5% of samples. A Chinese study reported an 8-color antibody tube (CD10, CD19, CD20, CD34, CD38, CD45, CD58, plus CD66c or CD13/CD33 or NG2/CD15) with a sensitivity of 0.001% in 81.6% of patients [8]. Shaver et al elegantly analyzed the relative contribution that each marker and/or pair of markers made to detect MRD[15], and concluded that a single 8-color tube consisting of CD9, CD10, CD19, CD20, CD34, CD38, CD45, and CD58, could provide as much diagnostic utility as their existing three-tube panel with 12 markers.

Within the EuroFlow Consortium (EU-FP6, LSHB-CT-2006-018708), we aimed to design standardized 8-color immunophenotyping protocols for multicenter MRD measurement in BCP-ALL and to improve the sensitivity of the assay to $\leq 10^{-5}$ (at least comparable to PCR). First, in order to select the most informative markers in distinguishing BCP-ALL from normal BCP cells, we applied novel software tools and principal component-based analyses[16, 17]. In each cycle of design-test-evaluate-redesign, the antibody tubes were tested on BCP-ALL samples and normal and/or regenerating bone marrow (BM), followed by assessment of the contribution of each antibody, until satisfactory results were obtained after five testing rounds. Second, a flowcytometric protocol for staining and acquisition of large numbers of cells (>4 million) was developed, allowing theoretical sensitivities of at least 0.001% ($\leq 10^{-5}$). Finally, the selected antibody tubes and standardized laboratory procedures were prospectively validated on follow-up samples from BCP-ALL patients, using the EuroMRD PCR-MRD methods in parallel as gold standard[2].

## Materials and methods

***BCP-ALL patients and normal controls***. Data were collected in seven EuroFlow centers. BM samples obtained from healthy donors or patients in whom no hematological malignancy could be detected (e.g. BM samples submitted for lymphoma staging, neuroblastoma staging) were used as control BM for normal/reactive BCP cells. BM samples obtained from pediatric ALL patients after induction therapy (day 78 of therapy) or one year after stop of therapy, proven to be MRD-negative by RQ-PCR analysis, were used as source of regenerating BCP. In the first part of the study (panel design and optimization), 319 BCP-ALL patients which were consecutively received during five design-test-evaluate-redesign phases (initial phase: n=69; phase 1: n=61; phase 2: n=28; phase 3: n=78; phase 4: n=83) were included. In the second part of the study (MRD analysis), 377 follow-up samples obtained from 178 BCP-ALL patients (day 15: n= 111; day 33: n= 139; day 78: n= 107; other time points: n=20) were included. Patient characteristics are summarized in **Table 1**. The institutional review board of each participating center approved this study and informed consent for study participation was obtained from each patient and/or his/her legal guardian.

***Immunophenotyping MRD panel design***. First, BM samples obtained from 69 BCP-ALL patients at diagnosis were stained with the EuroFlow BCP-ALL antibody panel (23 different antibodies in 4 8-color tubes)[18]. The subsequently designed and optimized MRD tubes were tested during phase 1 to 4 on diagnostic BM samples from BCP-ALL patients using the standardized EuroFlow sample preparation and instrument set-up protocols[18 19]. Data were analyzed using Infinicyt software by comparing BCP-ALL cells with the nearest normal/reactive BCP subsets using APS plots (see Supplemental Methods) as illustrated in **Figure 1** and **Supplemental Figure 1 and 2.** Regenerating BCP cells from 6 T-ALL patients were used as an additional negative control **(Supplemental Figure 3)**.

***Immunophenotyping MRD analyses.*** The finally selected BCP-ALL MRD tubes were evaluated on BM samples obtained during follow-up of BCP-ALL patients, using an optimized bulk-lysis protocol (see Results)[20]. BM samples were processed according to this new EuroFlow bulk-lysis protocol and subsequently stained using the regular EuroFlow protocol[19]. MRD analyses and interpretation were performed locally and data was subsequently sent to the BCP-ALL-MRD coordinator for central evaluation. Initial FCM-MRD data analysis was performed using two-dimensional dot plots for sequential gating of BCP-ALL cells, comparable to previous studies using 4-6 color stainings[10, 12]. For this study, we provisionally defined a minimum of 10 clustered events to consider a sample as MRD positive (lower limit of detection, LOD) and a minimum of 40 clustered events for accurate quantitation of the MRD level (lower limit of quantitation, LLOQ)[21]**.** Inter-laboratory variability in data analysis was evaluated as described in the Supplemental Methods and **Supplemental Figure 4**).

***RQ-PCR-based MRD analyses.*** MRD levels were routinely determined by real-time quantitative PCR (RQ-PCR) analysis of rearranged immunoglobulin (IG) and/or T-cell receptor (TR) gene rearrangements in laboratories participating in the quality control rounds of the EuroMRD network (see www.EuroMRD.org)[3, 22-26]. RQ-PCR data, performed in triplicate, were analyzed according to the

EuroMRD guidelines, using the criteria to prevent false-negative MRD results[2]. Since application of these criteria might result in some false-positive RQ-PCR results,[2] we performed next-generation sequencing (NGS) to confirm or to exclude the presence of MRD in discordant samples considered positive by RQ-PCR but negative by FCM.

***NGS-based MRD analyses.*** NGS was generally performed as described previously[27]. Briefly, depending on the *IGH*, *TRG*, and/or *TRD* rearrangements applied as MRD targets in the RQ-PCR analysis, we performed a targeted approach: the follow-up samples were amplified using the multiplex primer set(s) of the relevant IG/TR locus only and data analysis was focussed on the specific junctional region sequence (i.e. the one used for RQ-PCR analysis). The primers for TCRG were newly designed (**Supplemental Table 1**) and individual primer combinations from multiplex PCR were tested for sensitivity using NGS for diluted diagnostic ALL samples from patients with respective V and Jgamma segment combinations, all reaching the sensitivity of $10^{-5}$. All data were finally scored as either MRD-positive or MRD-negative.

## Results

***Design and optimization of 8-color MRD labeling for BCP-ALL:*** In the *initial phase*, five antibodies (CD19, CD45, CD34, CD10 and CD20) were upfront selected as backbone markers since they allow appropriate BCP gating as well as characterization of several BCP subpopulations, and are known to allow discrimination between normal BCP and BCP-ALL cells[10, 28, 29, 30]. To evaluate which other markers could contribute to optimal separation of BCP-ALL cells from normal/reactive BCP cells, the EuroFlow BCP-ALL diagnosis panel [18] was applied to 69 BCP-ALL patients as well as to normal/reactive BM samples. Based on principal component analysis (visualized through APS plots) [16, 17, 19] of the BCP-ALL cells versus normal/reactive BCP cells (analyzed per tube), CD9, CD123, CD66c, CD81, CD24 and CD10 appeared to be markers that were most frequently differentially expressed (see **Supplemental Figure 5**). These markers were combined with the five backbone markers listed above and complemented with TdT and CD58, both previously reported to be of relevance for BCP-ALL MRD analyses [10, 28, 29, 31]. The remaining open position was filled in with surface membrane (Sm) IgKappa/IgLambda, as a potential exclusion maker for more mature BCPs. Fluorochrome positions were primarily determined based on the position of the involved markers in the EuroFlow BCP-ALL panel [18].

The resulting three *Phase-1* MRD tubes (**Table 2)** were subsequently tested on 61 consecutive BCP-ALL patients at diagnosis and the discriminatory power was evaluated by comparing the leukemic BCP with the nearest normal BCP subset in APS plots. Whereas both tube 1 and tube 3 gave good/fair separation in approximately 60% of cases, tube 2 was clearly less informative (fair/good separation in <35% of cases)(**Figure 2**). When the tube providing the best separation for each patient was selected, good/fair separation was observed in 77% of cases. Considering only tube 1 and 3, good/fair separation was still observed in 71% of cases. These data indicate that tube 1 and 3 had complementary value and confirm the limited value of tube 2.

To evaluate the relevance of each individual marker in discriminating BCP-ALL cells from normal/reactive BCP cells, those markers that received a weight over 10% in the first or second prinicipal component in an APS view of the nearest normal BCP cells and the BCP-ALL cells were selected. CD66c (80% of cases), CD9 (63%), and CD123 (55%) contributed most frequently.

Based on these Phase-1 BCP-ALL results, the panel was redesigned: CD58, TdT, SmIgK/L, and CD81 were (at least provisionally) excluded, whereas CD22, which might be important for gating of B-cells in case of CD19-targeting therapies, was included (**Table 2**).

The *Phase-2* BCP-ALL MRD tubes were evaluated on diagnostic samples from 28 consecutive BCP-ALL patients. Good/fair separation between BCP-ALL cells and their nearest normal/reactive BCP counterpart was possible in approximately 75% of cases in both tubes (**Figure 2**), showing significant improvement over the Phase 1 tubes. If the best score of both tubes was used for each case, over 85% of BCP-ALL cases showed good/fair separation from the corresponding normal BCP subset and in only 3 cases (12%) separation was poor.

During Phase-2, additional studies were performed: 1) Because of non-optimal (relatively weak) CD9 (MEM61) staining, another CD9 clone (ML13) was evaluated with much stronger results; 2) Two newly available fluorochromes (APC-C750 and APC-A750) showed lower background than APC-H7 (less binding to apoptotic cells; no binding to monocytes); 3) More detailed evaluation of the usefulness of CD81 versus CD24 (re-evaluation of Phase 1 data) showed that CD81 was more frequently differentially expressed between normal/reactive BCP cells and BCP-ALL cells and showed that CD81 in combination with only the backbone markers resulted in a higher percentage of cases with good separation than CD24 did (31% versus 20%); 4) Because CD66c and CD123 are both virtually negative on normal/reactive BCP cells, we tested whether these markers could be combined in the PE channel and concluded that background levels were not affected by combining these two markers (data not shown). Based on these data, the new CD9 clone was included in the MRD panel, APC-H7 was replaced by APC-C750/A750, CD24 was replaced by CD81, and CD66c and CD123 were combined into one fluorescence channel. The open FITC position was used for further evaluation of CD58. The combined data provided the Phase-3 BCP-ALL MRD tubes (**Table 2)**.

The two *Phase-3* BCP-ALL MRD tubes were evaluated on 78 BCP-ALL patients. Overall, tube 1 resulted in good/fair separation in 90% of cases, whereas this was achieved in 82% of cases for tube 2 (**Figure 2**). In the three cases for which tube 1 did not result in good separation, normal/reactive BCP and BCP-ALL cells could be separated in tube 2, mainly due to differential expression of CD81. Further evaluation showed that CD38 (~35% of cases), CD66c/CD123 (~30%), and CD81 (~19%) improved the separation between normal/reactive and malignant BCP cells as compared to the five backbone markers only, whereas CD9, CD58 and CD22 had no or limited additional value. CD9 in tube 1 was therefore replaced by CD81-FITC (which demonstrated equally good staining patterns as CD81-APC-C750) and tube 2 was discarded.

Since in a few cases the evaluated MRD tubes did still not yet result in sufficient separation between normal/reactive and malignant BCP cells, we evaluated several other markers reported to be of potential interest for MRD analysis (e.g.: CD44-FITC, CD27-PE, CD164-FITC, CD73-PE, CD49f-FITC, CD200-PE, CD86-FITC, and Drebrin-PE) [32-36]. Based on initial testing on diagnostic BCP-ALL samples, CD73 and CD304 appeared to be most promising based on the level and frequency of overexpression (~20% for CD73 and ~40% for CD304) and their stability during follow-up (data not shown). Since it appeared not to be possible to combine these two markers with CD66c and CD123 in a single fluorescence channel (due to too high background levels), a second tube was designed; this tube was identical to tube 1 but with CD73/CD304 instead of CD66c/CD123 in the PE channel (**Table 2**).

The two *Phase-4* BCP-ALL MRD tubes were run on 83 consecutive diagnostic BCP-ALL samples. Overexpression of CD66c/CD123 or CD73/CD304 was observed in 45% and 46% of cases, respectively; 31% of cases did not show overexpression of either CD66c/CD123 or CD73/CD304. Tube 1 resulted in good/fair separation in 89% of cases, whereas this was attained in 82% of cases for tube 2 (**Figure 2**). If the best score of both tubes was used, 99% of cases showed good/fair separation

between the BCP-ALL cells and the nearest normal/reactive BCP subset. Therefore, tube 1 and tube 2 were complementary to each other and one might either decide at diagnosis which tube is best for monitoring the particular patient or use both tubes to have an extra internal control and more precise measurements. These two optimized tubes were considered to be final and ready for further evaluation in follow-up samples of BCP-ALL patients.

***Optimization of the flowcytometric MRD sample preparation protocol.*** We aimed for a sensitivity of $\leq 10^{-5}$, at least comparable to the sensitivity reached in RQ-PCR-based MRD analysis. If a cluster of 10-40 BCP-ALL cells should be present to consider a sample as positive, one should acquire at least 4 million cells in order to reach the required sensitivity. Since the cellularity of BM samples obtained during the early phases of treatment is frequently low[37], staining whole BM samples using the regular EuroFlow protocols would not allow acquisition of millions of cells. We therefore designed and tested a new EuroFlow erythrocyte bulk-lysis procedure: sufficiently large volumes of BM, i.e. containing >10 million cells, are lysed and the leukocytes are subsequently resuspended in a small volume of washing buffer. This new protocol allowed staining of 10 million cells in 100 µl cell suspension per tube (**Supplemental Table 2**). Evaluation of this new protocol showed that the percentage of doublets did not increase, that the number of evaluable leukocytes increased significantly, and that there were no major differences in cellular composition as compared to the regular EuroFlow staining protocol (**Figure 3**). Given the large increase in the number of cells stained with this new approach, all antibody titers were re-evaluated; modifications appeared not to be necessary.

***Evaluation of the EuroFlow BCP-ALL MRD tube.*** To evaluate whether the newly designed high-throughput EuroFlow BCP-ALL MRD strategy performed well, we tested the final MRD tubes on follow-up samples of BCP-ALL patients. Based on the immunophenotype of the BCP-ALL cells at diagnosis, one MRD tube was selected and subsequently used for MRD evaluation. First, flowcytometric MRD data obtained in 178 BCP-ALL patients were compared with routinely obtained PCR-MRD data. As shown in **Figure 4A**, the concordance between the FCM-MRD data and PCR-based MRD data was highly dependent on the number of cells acquired by flow cytometry. In addition, the sensitivity of FCM-MRD (percentage samples positive by both FCM and PCR relative to samples positive by PCR) significantly increased when higher cell numbers were acquired (**Figure 4B**). Therefore, only samples in which MRD could clearly be detected by FCM-MRD or samples which had sufficient cells acquired for reaching a sensitivity of $\leq 10^{-5}$ were included in the subsequent analyses. Based on a LLOQ of 40 events, at least $4x10^6$ cells should be acquired, which was possible in 227 out of 377 samples (60%). FCM-MRD data obtained in these patients was comparable to PCR-based MRD results in 93% of samples (**Figure 5** and **Table 3**). All but one of the 17 discordant samples (seven FCM+/PCR- and ten FCM-/PCR+) had MRD levels $<10^{-4}$ (**Supplemental Table 2**). Bland–Altman analysis showed higher PCR-based MRD values with a mean difference of 0.34 log or factor 2.2 (**Supplemental Figure 6**).

***Detailed evaluation of discordant cases.*** To evaluate the discordant cases, several additional analyses were performed. First, flow cytometry FCS data files were blindly distributed to four laboratories for re-analysis of the FCM-MRD data. Out of the seven cases initially scored positive by FCM-MRD and negative by PCR-MRD, six were interpreted as negative by all four centers upon FCM-MRD re-analysis, while one sample (day 15) was consistently scored positive by all four centers. Second, RQ-PCR-MRD data were checked for cases negative by FCM-MRD and positive by PCR-MRD. In eight of ten cases (all confirmed to be FCM-MRD negative by re-analysis in different centers), PCR-MRD data were considered positive based on a single well in a single target. To further evaluate whether this low-level positivity was potentially caused by nonspecific amplification, NGS was used to confirm the possible presence of the leukemia-specific antigen receptor rearrangement. In seven out of nine available samples, NGS-MRD was negative, whereas MRD-positivity could be confirmed in the remaining two patients. Thus, 6/17 (35%) discordant cases were due to initial misinterpretation of the FCM-MRD data and at least 7/17 (41%) was due to over-interpretation of PCR-MRD data; the remaining 4 cases appeared to be truly discordant cases (**Supplemental Table 3**). After these additional evaluations the actual concordance increased to 98%. If only samples with MRD levels <0.01% were included, 97% gave concordant results.

## Discussion

After five phases of optimization, we finally selected two 8-color antibody tubes which only differed for the markers present in the PE channel (CD66c/CD123 versus CD73/CD304). These two tubes are comparable with 8-color BCP-ALL MRD tubes recently used in other studies, since all proposed panels include CD19, CD10, CD20, CD34, and CD45. In our study, these markers were considered as backbone markers from start onwards, based on our previous experience[10, 12, 18, 28-30]. Also CD38 is present in all proposed panels and was proven to be relevant in our present study, as well as in the studies of Karawajew et al.[14] and Shaver et al[15]. The remaining two positions were completed with different markers: CD9, CD13/CD33, CD15/NG2, CD58 (two studies), CD66c (two studies), CD73, CD81, CD123, CD304, and a nucleic acid dye. In our analysis, CD9, CD58, and CD22 appeared be of limited value and therefore were discarded, while Shaver and colleagues, who also applied mathematical modelling systems, identified these as important MRD markers[15]. However, they did not test CD66c, CD73, CD123, or CD304 and the difference between the contribution of CD9 and CD81 in their study was limited[15]. CD15/NG2 might be relevant in ALL with MLL gene rearrangements, mainly occurring in infants[38]. These cases are rare and frequently present with a pro-B-ALL immunophenotype which can relatively easily be distinguished from normal BCP cells and plasma cells[39]. In our study, we tested and finally selected four markers that are frequently abnormally expressed on BCP-ALL cells: CD66c (associated with BCR-ABL and hyperdiploidy)[40, 41], CD123 (associated with hyperdiploidy[42]), CD73 [36], and CD304 (possibly associated with TEL-AML1)[33]. By combining two of these markers in a single fluorescence channel, abnormal expression could be identified in approximately 70% of BCP-ALL patients. Furthermore, in combination with the backbone markers and CD81, BCP-ALL cells could clearly be distinguished from normal BCP cells in virtually all patients. Thus, after multiple phases of multicenter testing of a wide range of leukocyte markers, antibody clones and fluorochrome-conjugated reagents using objective novel software tools, we were able to select two highly effective BCP-ALL MRD tubes.

It remains to be evaluated whether the designed BCP-ALL MRD tubes can also be used during antibody-based therapies. Especially Blinatumomab and CAR-T-cells (targeting CD19) may hamper the gating of BCP based on CD19. Although alternative gating strategies can be applied (e.g. based on CD10, CD34 and/or CD45), one could also decide to add CD24 and/or CD22 to the current tubes (transforming it into a 10-color tube) in MRD-based trials involving Blinatumomab[43]. Addition of CD24 and CD22 will also have the advantage that the earliest BCP cells, expressing CD24 and/or CD22 but not yet CD19,[44] can be identified; this may be of relevance for the identification of all BCP cells in regenerating BM samples.

In order to obtain MRD data with good sensitivity, acquisition of large numbers of cells appears to be a prerequisite. There is no consensus yet about the number of cells needed for a population. Most studies in BCP-ALL indicate a minimum number between 10 and 50 events, while a recent consensus report on MRD detection in multiple myeloma patients defined 20 and 50 cells as the LOD and LLOQ, respectively[21 14, 45]. Consequently, a sensitivity of $10^{-5}$ (generally reached in PCR-MRD and NGS-MRD

analysis), requires acquisition of ≥$10^6$ cells, preferably ≥5x$10^6$ cells. We therefore developed the new Euroflow bulk lysis protocol, allowing acquisition of such high cell numbers. Although the bulk lysis protocol contains several washing steps, which will likely result in some cell loss, there is no evidence for selective loss of BCP-ALL cells, given the high concordance between the final FCM-MRD results and the PCR-MRD results. To our best knowledge, other FCM-MRD studies so far have not acquired ≥4x$10^6$ events and therefore could not have reached the same sensitivity as shown here, although the study by the ALL-REZ-BFM 2002 trial group comes close[14]. Our data clearly shows that acquisition of large numbers of cells (≥4x$10^6$) is a prerequisite for obtaining good sensitivities and data which are truly comparable to PCR-MRD data.

We finally tested the newly designed and optimized BCP-ALL MRD tubes in combination with the bulk lysis protocol on follow-up samples from BCP-ALL patients and compared the FCM-MRD data with PCR-MRD data. The concordance between both methods (in the absence of any cut-off) was extremely high (93%), and was significantly better than in previous studies (82.3%[10] and 86.5%[14]). As mentioned above, this increased concordance is likely due to the higher sensitivity of the current study, resulting from the higher number of cells analyzed. Detailed evaluation of the discordant samples using NGS-based approaches showed that several discrepant cases were due to over-interpretation of the PCR-MRD data, as the involved leukemia-specific IG/TR rearrangements could not be detected by (qualitative) NGS analysis. Consequently, it is most likely that in these discordant cases the positive RQ-PCR MRD data, interpreted according to the EuroMRD criteria for prevention of false-negative MRD results, are due to non-specific amplification[2, 10, 46, 47] and that these samples actually were MRD-negative.

Next to these “false-positive PCR-MRD” samples, part of the initially discordant cases could be explained by “false-positive FCM-MRD” results: samples were initially scored MRD-positive (generally at very low levels of <0.01%), but these samples were consistently scored MRD-negative upon blind re-analysis at four different centers. Interpretation of FCM-MRD data, especially at MRD levels <0.01%, is still expert-based and depends on the number of events in the suspected population, their distance from normal, and the homogeneity of the suspected population (clustering of suspected cells). Whereas the number of events may easily be defined, distance from normal, and homogeneity of the population are more complex to be objectively defined. Novel software tools, including automated gating approaches and maturation pathway analysis (**Supplemental Figure 7**), are currently being developed within the EuroFlow consortium and will facilitate more standardized and objective FCM-MRD measurements in the near future [37]. Preliminary maturation-based FCM-MRD data showed very good concordance with PCR-MRD data, although further improvements are needed for detection of low levels of MRD (<0.01%).

The NGS-MRD analyses and re-analyses of FCM-MRD data increased the concordance to an unprecedented rate of 98%; the remaining discordant cases are likely due to statistical variation around the detection limits of both assays[14]. Therefore, the here presented EuroFlow FCM-MRD strategy proved to be highly sensitive (at least comparable to PCR-MRD) and fast, and allows

standardized quantification of MRD in virtually all BCP-ALL patients. By increasing the number of acquired cells to $10^7$ (e.g. by running both tubes with $5x10^6$ cells), the sensitivity and robustness likely can even be further increased.

**Acknowledgements**

The research was performed within the EuroFlow Consortium, which started with an EU-FP6 grant (LSHB-CT-2006-018708) and obtained sustainability by protecting and licensing intellectual property, thereby obtaining royalties, which are exclusively being used for supporting the EuroFlow research program (chairmen: JJMvD and AO). We gratefully acknowledge Tomas Kalina and the technicians of the Laboratory for Medical Immunology, Erasmus MC, for technical assistance. We thank Christa Homburg and colleagues for performing molecular MRD analyses of part of the Dutch patients. We thank Marieke Comans-Bitter for organizational support. EF was supported by the Grant Agency of the Czech Republic (GACR, project of Centre of Excellence No. P302/12/G101). LS, TS and PT were supported by ERA-NET PrioMedChild, grant 40-41800-98-027. EM was supported by Ministry of Health of the Czech Republic, grant nr. 15-28525A. MK was supported by the University Hospital Motol, Prague, Czech Republic (00064203). ECS, QL and AO acknowledge the Bilateral Cooperation Program between Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES (Brasília/Brazil) and Dirección General de Políticas Universitárias – Ministério de Educación, Cultura y Deportes DPGU (Madrid/Spain) (311/15). ESC acknowledge FAPERJ, Rio de Janeiro, Brazil (E26/110.105/2014; E26/102.191/2013) and Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPQ of Brazil (400194/2014-7). GG, CB and PB were supported by Fondazione Tettamanti. We gratefully acknowledge the contribution of the EuroClonality/EuroMRD NGS network (chair: A.W. Langerak) for support with the NGS data. The research for this manuscript was in part performed within the framework of the Erasmus Postgraduate School Molecular Medicine.

**Author's Contributions**

VHJvdV, EM, JJMvD, and AO designed the research;
PT, EM, QL, GG, AO, JJMvD and VHJvdV developed the methodology and the research strategies;
PT and VHJvdV interpreted the data;
PT, EM, LS, AvdS, CB, EO, GGa, MB, ASdC, MN, ESC, PB, TS, LH, GGr, AO and VHJvdV analysed flow cytometry data;
PT, EM, LS, AvdS, MB, MN, OH, ESC, JtM, CB, EO, PB, LH, performed flow cytometry experiments;
MK, EF, JT and MB designed and performed the NGS experiments and analysed NGS data;
VHJvdV wrote the paper; all authors reviewed and approved the manuscript.

**Disclosure of Conflicts of Interest**

JJMvD: contract research for Roche, Amgen, and BD Biosciences. VHJvdV: contract research for Roche, Amgen, Pfizer, Janssen, and BD Biosciences; consultancy fees: Celgene. GGE is employee of Cytognos SL, Salamanca, Spain. The other authors declare no conflict of interest.

## References


1. Flohr T, Schrauder A, Cazzaniga G, Panzer-Grumayer R, van der Velden V, Fischer S*, 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.
2. van der Velden VH, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer ER*, et al.* Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* 2007; 21 (4):604-611.
3. van der Velden VH, van Dongen JJ. MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR. *Methods Mol Biol* 2009; 538:115-150.
4. Basso G, Veltroni M, Valsecchi MG, Dworzak MN, Ratei R, Silvestri D*, 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.
5. Borowitz MJ, Wood BL, Devidas M, Loh ML, Raetz EA, Salzer WL*, et al.* Prognostic significance of minimal residual disease in high risk B-ALL: a report from Children's Oncology Group study AALL0232. *Blood* 2015; 126 (8):964-971.
6. Cheng SH, Lau KM, Li CK, Chan NP, Ip RK, Cheng CK*, et al.* Minimal residual disease-based risk stratification in Chinese childhood acute lymphoblastic leukemia by flow cytometry and plasma DNA quantitative polymerase chain reaction. *PLoS One* 2013; 8 (7):e69467.
7. Stow P, Key L, Chen X, Pan Q, Neale GA, Coustan-Smith E*, et al.* Clinical significance of low levels of minimal residual disease at the end of remission induction therapy in childhood acute lymphoblastic leukemia. *Blood* 2010; 115 (23):4657-4663.
8. Weng XQ, Shen Y, Sheng Y, Chen B, Wang JH, Li JM*, et al.* Prognostic significance of monitoring leukemia-associated immunophenotypes by eight-color flow cytometry in adult B-acute lymphoblastic leukemia. *Blood Cancer J* 2013; 3:e133.
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 (6):1952-1958.
10. Denys B, van der Sluijs-Gelling AJ, Homburg C, van der Schoot CE, de Haas V, Philippe J*, et al.* Improved flow cytometric detection of minimal residual disease in childhood acute lymphoblastic leukemia. *Leukemia* 2013; 27 (3):635-641.
11. Ryan J, Quinn F, Meunier A, Boublikova L, Crampe M, Tewari P*, et al.* Minimal residual disease detection in childhood acute lymphoblastic leukaemia patients at multiple time-points reveals high levels of concordance between molecular and immunophenotypic approaches. *Br J Haematol* 2009; 144 (1):107-115.
12. Gaipa G, Cazzaniga G, Valsecchi MG, Panzer-Grumayer R, Buldini B, Silvestri D*, et al.* Time point-dependent concordance of flow cytometry and RQ-PCR inminimal residual disease detection in childhood acute lymphoblasticleukemia. *Haematologica* 2012; 97:1582-1593.

13. Thorn I, Forestier E, Botling J, Thuresson B, Wasslavik C, Bjorklund E*, et al.* Minimal residual disease assessment in childhood acute lymphoblastic leukaemia: a Swedish multi-centre study comparing real-time polymerase chain reaction and multicolour flow cytometry. *Br J Haematol* 2011; 152 (6):743-753.
14. Karawajew L, Dworzak M, Ratei R, Rhein P, Gaipa G, Buldini B*, et al.* Minimal residual disease analysis by eight-color flow cytometry in relapsed childhood acute lymphoblastic leukemia. *Haematologica* 2015; 100 (7):935-944.
15. Shaver AC, Greig BW, Mosse CA, Seegmiller AC. B-ALL minimal residual disease flow cytometry: an application of a novel method for optimization of a single-tube model. *Am J Clin Pathol* 2015; 143 (5):716-724.
16. Pedreira CE, Costa ES, Almeida J, Fernandez C, Quijano S, Flores J*, et al.* A probabilistic approach for the evaluation of minimal residual disease by multiparameter flow cytometry in leukemic B-cell chronic lymphoproliferative disorders. *Cytometry A* 2008; 73A (12):1141-1150.
17. Pedreira CE, Costa ES, Barrena S, Lecrevisse Q, Almeida J, van Dongen JJ*, et al.* Generation of flow cytometry data files with a potentially infinite number of dimensions. *Cytometry A* 2008; 73 (9):834-846.
18. van Dongen JJ, Lhermitte L, Bottcher S, Almeida J, van der Velden VH, Flores-Montero J*, et al.* EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* 2012; 26 (9):1908-1975.
19. Kalina T, Flores-Montero J, van der Velden VH, Martin-Ayuso M, Bottcher S, Ritgen M*, et al.* EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia* 2012; 26 (9):1986-2010.
20. Aalbers AM, van den Heuvel-Eibrink MM, Baumann I, Dworzak M, Hasle H, Locatelli F*, et al.* Bone marrow immunophenotyping by flow cytometry in refractory cytopenia of childhood. *Haematologica* 2015; 100 (3):315-323.
21. Arroz M, Came N, Lin P, Chen W, Yuan C, Lagoo A*, et al.* Consensus guidelines on plasma cell myeloma minimal residual disease analysis and reporting. *Cytometry B Clin Cytom* 2016; 90 (1):31-39.
22. van der Velden VHJ, Willemse MJ, van der Schoot CE, Hahlen K, van Wering ER, van Dongen JJM. Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time quantitative PCR. *Leukemia* 2002; 16 (5):928-936.
23. van der Velden VHJ, Wijkhuijs JM, Jacobs DC, van Wering ER, van Dongen JJM. T cell receptor gamma gene rearrangements as targets for detection of minimal residual disease in acute lymphoblastic leukemia by real-time quantitative PCR analysis. *Leukemia* 2002; 16 (7):1372-1380.
24. van der Velden VH, de Bie M, van Wering ER, van Dongen JJ. Immunoglobulin light chain gene rearrangements in precursor-B-acute lymphoblastic leukemia: characteristics and applicability for the detection of minimal residual disease. *Haematologica* 2006; 91 (5):679-682.

25. Szczepanski T, van der Velden VHJ, Hoogeveen PG, De Bie M, Jacobs DCH, Van Wering ER*, et al.* V{delta}2-J{alpha} gene rearrangements are frequent in precursor-B-acute lymphoblastic leukemia but rare in normal lymphoid cells. *Blood* 2004; 103:3798-3804.
26. Bruggemann M, van der Velden VHJ, Raff T, Droese J, Ritgen M, Pott C*, et al.* Rearranged T-cell receptor beta genes represent powerful targets for quantification of minimal residual disease (MRD) in childhood and adult T-cell acute lymphoblastic leukemia (T-ALL). *Leukemia* 2004; 18:709-719.
27. Kotrova M, Muzikova K, Mejstrikova E, Novakova M, Bakardjieva-Mihaylova V, Fiser K*, et al.* The predictive strength of next-generation sequencing MRD detection for relapse compared with current methods in childhood ALL. *Blood* 2015; 126 (8):1045-1047.
28. Lucio P, Gaipa G, van Lochem EG, van Wering ER, Porwit-MacDonald A, Faria T*, 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.
29. Lucio P, Parreira A, van den Beemd MW, van Lochem EG, van Wering ER, Baars E*, 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.
30. Mejstrikova E, Fronkova E, Kalina T, Omelka M, Batinic D, Dubravcic K*, et al.* Detection of residual B precursor lymphoblastic leukemia by uniform gating flow cytometry. *Pediatr Blood Cancer* 2010; 54 (1):62-70.
31. Veltroni M, De Zen L, Sanzari MC, Maglia O, Dworzak MN, Ratei R*, 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.
32. Coustan-Smith E, Song G, Clark C, Key L, Liu P, Mehrpooya M*, et al.* New markers for minimal residual disease detection in acute lymphoblastic leukemia. *Blood* 2011; 117 (23):6267-6276.
33. Solly F, Angelot F, Garand R, Ferrand C, Seilles E, Schillinger F*, et al.* CD304 is preferentially expressed on a subset of B-lineage acute lymphoblastic leukemia and represents a novel marker for minimal residual disease detection by flow cytometry. *Cytometry A* 2012; 81 (1):17-24.
34. Vaskova M, Kovac M, Volna P, Angelisova P, Mejstrikova E, Zuna J*, et al.* High expression of cytoskeletal protein drebrin in TEL/AML1pos B-cell precursor acute lymphoblastic leukemia identified by a novel monoclonal antibody. *Leuk Res* 2011; 35 (8):1111-1113.
35. Vaskova M, Mejstrikova E, Kalina T, Martinkova P, Omelka M, Trka J*, et al.* Transfer of genomics information to flow cytometry: expression of CD27 and CD44 discriminates subtypes of acute lymphoblastic leukemia. *Leukemia* 2005; 19 (5):876-878.
36. Wang W, Gao L, Li Y, Li ZL, Gong M, Huang FZ*, et al.* The application of CD73 in minimal residual disease monitoring using flow cytometry in B-cell acute lymphoblastic leukemia. *Leuk Lymphoma* 2015:1-8.
37. van Dongen JJ, van der Velden VH, Bruggemann M, Orfao A. Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. *Blood* 2015; 125 (26):3996-4009.

38. Behm FG, Smith FO, Raimondi SC, Pui CH, Bernstein ID. Human homologue of the rat chondroitin sulfate proteoglycan, NG2, detected by monoclonal antibody 7.1, identifies childhood acute lymphoblastic leukemias with t(4;11)(q21;q23) or t(11;19)(q23;p13) and MLL gene rearrangements. *Blood* 1996; 87 (3):1134-1139.
39. Jansen MW, Corral L, van der Velden VH, Panzer-Grumayer R, Schrappe M, Schrauder A*, et al.* Immunobiological diversity in infant acute lymphoblastic leukemia is related to the occurrence and type of MLL gene rearrangement. *Leukemia* 2007; 21 (4):633-641.
40. Kalina T, Vaskova M, Mejstrikova E, Madzo J, Trka J, Stary J*, et al.* Myeloid antigens in childhood lymphoblastic leukemia: clinical data point to regulation of CD66c distinct from other myeloid antigens. *BMC Cancer* 2005; 5:38.
41. Kiyokawa N, Iijima K, Tomita O, Miharu M, Hasegawa D, Kobayashi K*, et al.* Significance of CD66c expression in childhood acute lymphoblastic leukemia. *Leuk Res* 2014; 38 (1):42-48.
42. Djokic M, Bjorklund E, Blennow E, Mazur J, Soderhall S, Porwit A. Overexpression of CD123 correlates with the hyperdiploid genotype in acute lymphoblastic leukemia. *Haematologica* 2009; 94 (7):1016-1019.
43. Cherian S, Miller V, McCullouch V, Dougherty K, Fromm JR, Wood BL. A novel flow cytometric assay for detection of residual disease in patients with B-lymphoblastic leukemia/lymphoma post anti-CD19 therapy. *Cytometry B Clin Cytom* 2016.
44. van Lochem EG, van der Velden VHJ, Wind H, te Marvelde JG, Westerdaal NAC, van Dongen JJM. Immunophenotypic Differentiation Patterns of Normal Hematopoiesis in Human Bone Marrow: reference patterns for age-related changes and disease-induced shifts. *Clinical Cytometry* 2004; 60B:1-13.
45. Wood BL. Principles of minimal residual disease detection for hematopoietic neoplasms by flow cytometry. *Cytometry B Clin Cytom* 2016; 90 (1):47-53.
46. Fronkova E, Muzikova K, Mejstrikova E, Kovac M, Formankova R, Sedlacek P*, et al.* B-cell reconstitution after allogeneic SCT impairs minimal residual disease monitoring in children with ALL. *Bone Marrow Transplant* 2008; 42 (3):187-196.
47. van der Velden VH, Wijkhuijs JM, van Dongen JJ. Non-specific amplification of patient-specific Ig/TCR gene rearrangements depends on the time point during therapy: implications for minimal residual disease monitoring. *Leukemia* 2008; 22 (3):641-644.

**Table 1. Patient characteristics.**

| | Initial Phase | | Phase 1 | | Phase 2 | | Phase 3 | | Phase 4 | | MRD phase | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number | 69 | | 61 | | 28 | | 78 | | 83 | | 178 | |
| Age (years) | | | | | | | | | | | | |
| median | 6 | | 5 | | 5 | | 5 | | 3 | | 5 | |
| min | 0 | | 1 | | 1 | | 0 | | 0 | | 0 | |
| max | 76 | | 75 | | 17 | | 17 | | 55 | | 77 | |
| Gender | | | | | | | | | | | | |
| M | 35 | 56% | 40 | 66% | 14 | 50% | 46 | 59% | 43 | 52% | 94 | 53% |
| F | 27 | 44% | 21 | 34% | 14 | 50% | 32 | 41% | 40 | 48% | 84 | 47% |
| WBC (x10e9/L) | | | | | | | | | | | | |
| median | 13.4 | | 8.2 | | 13.7 | | 10.9 | | 15.9 | | 8.5 | |
| min | 0.4 | | 0.8 | | 1.2 | | 0.7 | | 0.5 | | 0.5 | |
| max | 818 | | 87.6 | | 312 | | 595 | | 1250 | | 900 | |
| No data | 10 | | 1 | | 0 | | 2 | | 19 | | 11 | |
| Immunophenotype | | | | | | | | | | | | |
| pro-B-ALL | 8 | 13% | 2 | 3% | 4 | 14% | 5 | 6% | 6 | 7% | 5 | 3% |
| common-ALL | 44 | 73% | 48 | 79% | 20 | 71% | 59 | 76% | 56 | 67% | 137 | 77% |
| pre-B-ALL | 8 | 13% | 11 | 18% | 4 | 14% | 13 | 17% | 21 | 25% | 35 | 20% |
| No data | 7 | | 0 | | 0 | | 1 | | 0 | | 1 | |
| Genetic data[a] | | | | | | | | | | | | |
| *TEL-AML1* | 6 | 11% | 15 | 27% | 6 | 27% | 11 | 16% | 11 | 18% | 45 | 26% |
| *BCR-ABL* | 4 | 6% | 1 | 2% | 0 | 0% | 3 | 4% | 1 | 2% | 4 | 2% |
| *E2A-PBX1* | 0 | 0% | 3 | 6% | 0 | 0% | 0 | 0% | 0 | 0% | 4 | 2% |
| *MLL-AF4* | 4 | 6% | 1 | 2% | 2 | 7% | 3 | 4% | 4 | 6% | 3 | 2% |
| *MLL* other | 2 | 3% | 1 | 2% | 1 | 4% | 1 | 1% | 1 | 2% | 2 | 1% |
| hyperdiplod | 14 | 25% | 27 | 48% | 6 | 26% | 26 | 37% | 22 | 37% | 67 | 43% |
| hypodiploid | 0 | 0% | 1 | 2% | 1 | 4% | 1 | 1% | 1 | 2% | 2 | 1% |

[a] Genetic data were not available for all patients; percentages refer to positive patients per all analyzed patients.

**Table 2**. Development and design of the EuroFlow BCP-ALL MRD panel[a]

| Phase | Violet laser | | Blue laser | | | | Red laser | |
|---|---|---|---|---|---|---|---|---|
| **Initial**[b] | **PB** | **PO** | **FITC** | **PE** | **PerCP Cy5.5** | **PECy7** | **APC** | **APC H7** |
| | CD20 | CD45 | CD58 | CD66c | CD34 | CD19 | CD10 | CD38 |
| | CD9 | CD45 | TdT | CD13 | CD34 | CD19 | CD22 | CD24 |
| | CD21 | CD45 | CD15/CD65 | NG2 | CD34 | CD19 | CD123 | CD81 |
| **1** | **PB** | **PO** | **FITC** | **PE** | **PerCP Cy5.5** | **PECy7** | **APC** | **APC H7** |
| | CD20 | CD45 | CD58 | CD66c | CD34 | CD19 | CD10 | CD38 |
| | CD20 | CD45 | TdT | SmIgK/L | CD34 | CD19 | CD10 | CD24 |
| | CD20 | CD45 | CD9 | CD123 | CD34 | CD19 | CD10 | CD81 |
| **2** | **PB** | **PO** | **FITC** | **PE** | **PerCP Cy5.5** | **PE Cy7** | **APC** | **APC H7** |
| | CD20 | CD45 | CD9 | CD22 | CD34 | CD19 | CD10 | CD38 |
| | CD20 | CD45 | CD123 | CD66c | CD34 | CD19 | CD10 | CD24 |
| **3** | **PB** | **PO** | **FITC** | **PE** | **PerCP Cy5.5** | **PE Cy7** | **APC** | **APC C750** |
| | CD20 | CD45 | CD9 (ML13) | CD66c/CD123 | CD34 | CD19 | CD10 | CD38 |
| | CD20 | CD45 | CD58 | CD22 | CD34 | CD19 | CD10 | CD81 |
| **4** | **PB** | **PO** | **FITC** | **PE** | **PerCP Cy5.5** | **PE Cy7** | **APC** | **APC C750** |
| | CD20 | CD45 | CD81 | CD66c/CD123 | CD34 | CD19 | CD10 | CD38 |
| | CD20 | CD45 | CD81 | CD73/CD304 | CD34 | CD19 | CD10 | CD38 |

[a] Markers that were changed as compared to the previous panel are marked grey.

[b] Only tube 1, 3 and 4 of the EuroFlow BCP-ALL diagnosis panel were evaluated, since most markers in tube 2 (Cyμ, SmIgK, SmIgL, SmIgM) were not expected to contribute to separation of normal and malignant BCP cells.

**Table 3. Concordance between flowcytometric and molecular MRD data.[a]**

| | Day 15 | Day 33 | Day 78 | Other TP | Total | % | % |
|---|---|---|---|---|---|---|---|
| **Concordant** | | | | | | | |
| FCM+/PCR+ | 102 | 47 | 11 | 9 | 169 | 74 | **93** |
| FCM-/PCR- | 1 | 5 | 31 | 4 | 41 | 18 | |
| **Discordant** | | | | | | | |
| FCM-/PCR+ | 0 | 5 | 5 | 0 | 10 | 4 | **7** |
| FCM+/PCR- | 1 | 4 | 2 | 0 | 7 | 3 | |
| **Total** | **104** | **61** | **49** | **13** | **227** | **100** | |

[a] Only samples in which MRD could clearly be detected by FCM-MRD or samples which had at least $4x10^6$ cells acquired were included in the analyses. Based on a LLOQ of 40 events, this allowed a sensitivity of at least $10^{-5}$.

## Figure Legends

**Figure 1. Data analysis strategy used to optimize the antibody panel for distinguishing between BCP-ALL cells and their nearest normal/reactive BCP counterpart.** First multiple normal/reactive BM samples and/or regenerating BM samples were merged (phase 1: n=7; phase 2: n= 11; phase 3: n=14; phase 4: n=10) and CD19-positive B-cells were selected. These were subdivided in four B-cell subsets, based on the backbone markers (CD19/CD10/CD20/CD34/CD45): CD34+ pre-B-I cells (light green), CD34-/CD10+/CD20- to dim pre-B-II/immature cells (dark blue), CD34-/CD10+/CD20+ immature/transitional B-cells (light blue), and CD34-/CD10-/CD20+ mature B-cells (dark green). Dot plots of CD34 versus CD10 (**A.**) and CD10 versus CD20 (**B.**) are shown. The one SD (dashed line) and 2SD lines (solid line) of the two most immature BCP subsets (pre-B-I (light green) and pre-B-II/immature (dark blue)) were displayed in an APS view, which was subsequently fixed (supervised; **C.**). Each individual BCP-ALL case was added to the fixed APS plot and the normal BCP population nearest to each of the BCP-ALL populations was defined (**D.**). The BCP-ALL cells and nearest normal BCP subset were then visualized in a separate (non-fixed and balanced) APS plot, one using the backbone markers only (**E.**) and one using all eight markers (**F.**), by plotting the 1 standard deviation (SD) curve and 2SD curves of the two populations. To prevent an influence of the number of cells on the PCA, we opted to use a balanced PCA, implying a fixed ratio between normal and pathological events. Finally, the separation between the two populations was scored, based on: no overlap between 2SD curves: 3 points; overlap of the 2SD curves: 2 points; overlap of the 2SD and the 1SD curve: 1 point; overlap of both 1SD curves: 0 points. An example of this scoring is shown in **Supplemental Figure 1**. It should be noted that the above described strategy was only used for optimizing the antibody panel for the BCP-ALL MRD tubes, and not for actual MRD analyses.

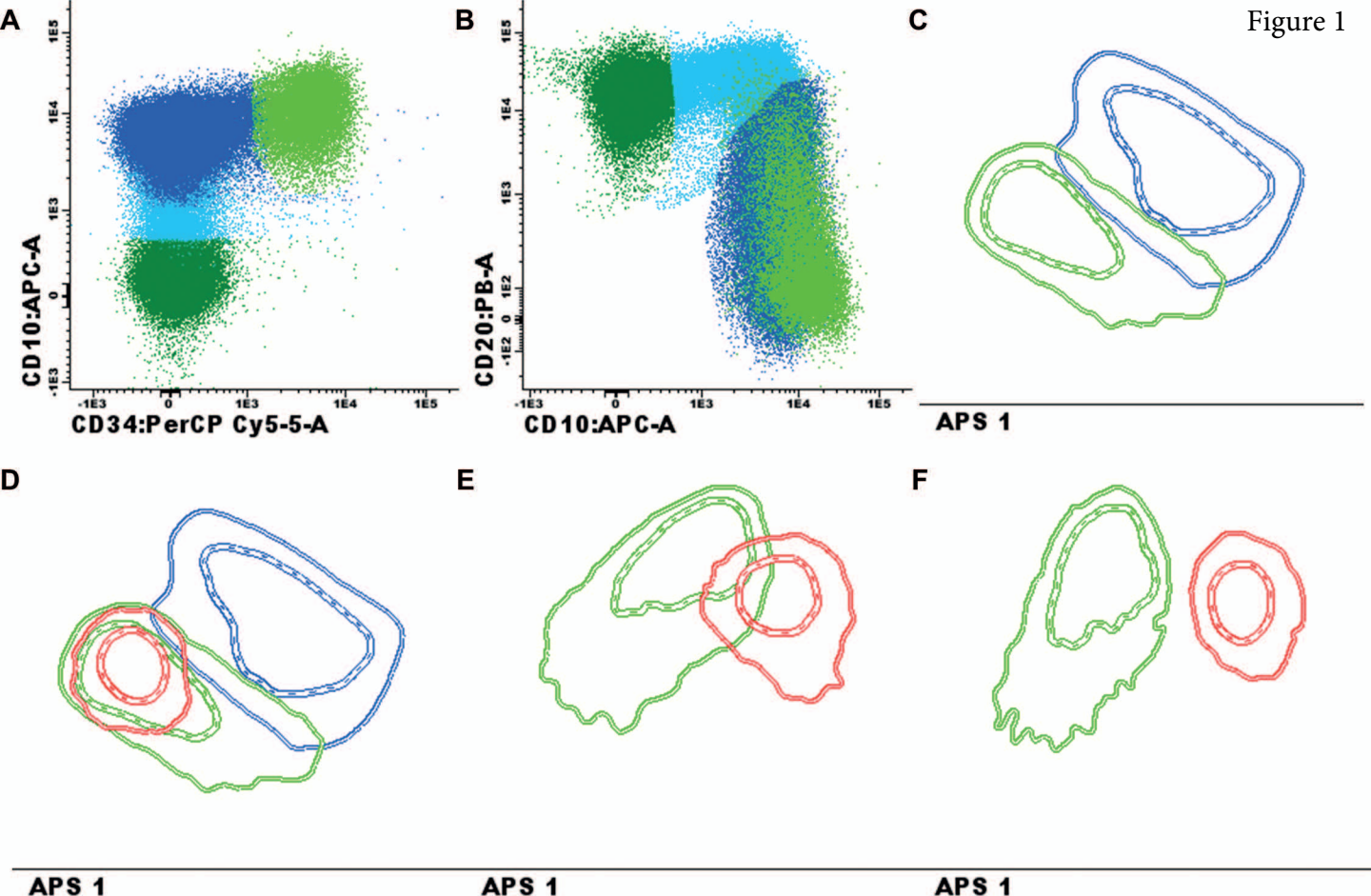
**Figure 2. Power to distinguish BCP-ALL cells from their nearest normal BCP counterpart using the EuroFlow BCP-ALL MRD tubes.** Data reflect the percentage of patients that reached the specified score, obtained as described in Supplemental Figure 1. Briefly, for each patient BCP-ALL cells and their nearest normal BCP subset were visualized in a (non-fixed and balanced) APS plot showing the median, 1SD curves and 2SD curves for both populations. Each patient was subseqently scored as follows: no overlap between 2SD curves: 3 points; overlap of the 2SD curves: 2 points; overlap of the 2SD and the 1SD curve: 1 point; overlap of both 1SD curves: 0 points. Max refers to the maximal score of the individual tubes. Phase 1: seven normal/reactive BM samples and 61 BCP-ALL patients; Phase 2: seven normal BM samples, four regenerating BM samples, and 28 BCP-ALL patients; Phase 3: five normal/reactive BM samples, nine regenerating BM samples, and 78 BCP-ALL patients; Phase 4: ten normal/reactive BM samples and 83 BCP-ALL patients).

**Figure 3. Evaluation of the EuroFlow bulk-lysis protocol.** For reasons of comparison, each of the BM samples (day 15: n=15; day 33: n=15; day 78: n=12) was processed according to the standard EuroFlow protocol (FL) and in parallel according to the EuroFlow bulk-lysis protocol (BL). **A.** Number of leukocytes, debris and doublets, calculated as percentage of acquired events. Using the bulk-lysis method, significantly less debris (p=0.032 by paired t-test) and significantly more leukocytes (p=0.03

by paired t-test) were measured. There were no significant differences between the two methods for the percentage of doublets. **B**. Absolute number of leukocytes acquired. Using BL, on average 12-fold more leukocytes could be acquired ($p<0.0001$). Please note that we included relatively many day 15 samples in order to be able to evaluate the impact of the two methods on the MRD levels as well. However, these day 15 samples generally have a very low WBC; consequently, the number of leukocytes acquired after BL is still relatively low in a subset of samples. **C**. Distribution of leukocyte subpopulations, defined as percentage of leukocytes. By paired t-test (two-sided), small but statistically significant differences were observed for T/NK cells (mean: 24% versus 26%, $p=0.0047$), granulocytes (mean: 33% versus 38%, $p<0.001$) and monocytes (mean: 3.2% versus 4.5%, $p<0,001$), whereas no significant differences were observed for the remaining populations. Of note, in two samples MRD was only detected using the bulk-lysis method (0.013% and 0.018%) but not using the whole BM method. In the 11 samples MRD positive by both methods, MRD levels were not significantly different from each other (paired t-test: $p=0.30$), with mean values of 6.3% and 6.7% by whole BM and bulk-lysed BM method, respectively. Correlation analysis showed a Spearman r of 0.964 (95% CI: 00.857-0.991; $p<0.0001$). PC=plasma cells.

**Figure 4. Performance of FCM-MRD versus PCR-based MRD is dependent on the number of acquired cells. A.** The percentage of discordant cases by FCM-MRD and PCR-MRD is shown for individual time-points (day 15, day 33, and day 78) as well as for all samples together. Data are presented for variable numbers of acquired cells: all samples (independent of cell number) and samples with at least 1, 2, 3, 4, or 5 million cells acquired. **B.** The sensitivity of FCM-MRD relative to PCR-MRD is shown for individual time-points (day 15, day 33, and day 78) as well as for all samples together. Data are presented for variable numbers of acquired cells: all samples (independent of cell number; n=377) and samples with at least 1 (n=330), 2 (n=287), 3 (n=255), 4 (n=227), or 5 million cells (n=191) acquired. Sensitivity is calculated as the number of samples positive by both FCM and PCR divided by the total number of samples positive by PCR (i.e. the reference method).

**Figure 5. Comparison of MRD data obtained by 8-color EuroFlow flow cytometry and routinely obtained molecular MRD data.** Flowcytometric MRD data were compared with molecular MRD data and are shown for samples obtained at day 15 (**A**), day 33 (**B**), or day 78 (**C**). In the lower right part of each panel the number of FCM-/PCR+, FCM+/PCR+, FCM+/PCR-, and FCM-/PCR- is indicated. Only samples in which MRD could clearly be detected by FCM-MRD or samples which had sufficient cells acquired for reaching a sensitivity of $\leq 10^{-5}$ (i.e. $\geq 4x10^{6}$ cells acquired) were included in the analyses. Based on a LOD of 10 events, the sensitivity of FCM-MRD was $2,5x10^{-6}$; the quantitative range (based on a LLOQ of 40 events) is $10^{-5}$. Consequently, FCM-MRD data between $2,5x10^{-6}$ and $10^{-5}$ should be considered as positive, but are below the limit of quantitation.


A
CD10:APC-A
CD34:PerCP Cy5-5-A
B
CD20:PB-A
CD10:APC-A
C
APS 1
D
APS 1
E
APS 1
F
APS 1

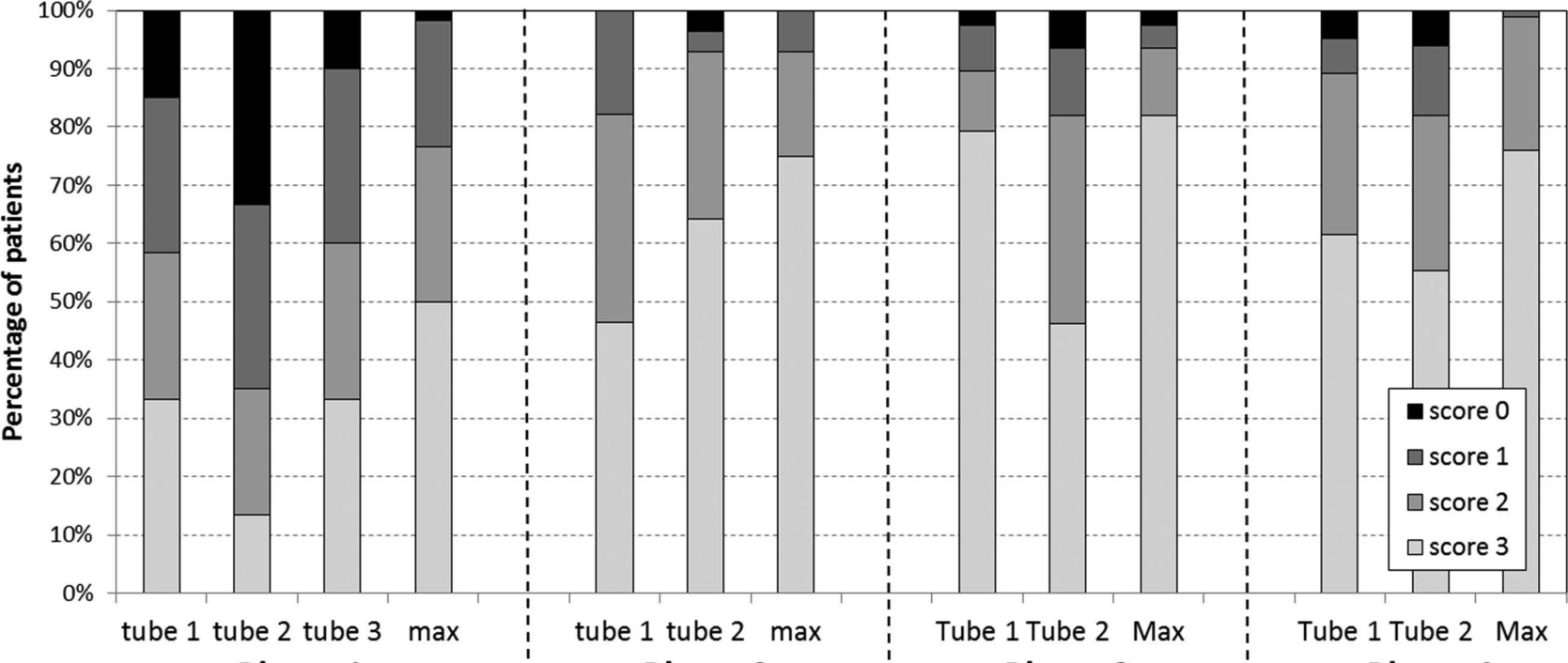
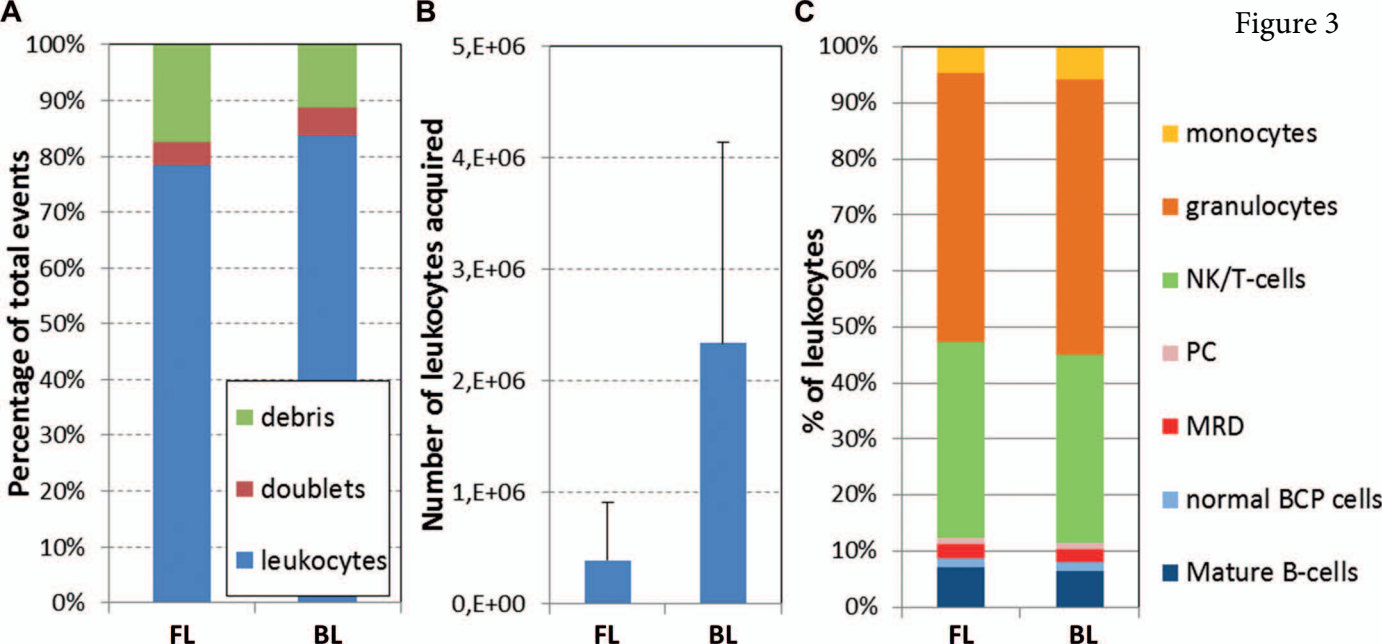

Percentage of patients
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
tube 1
tube 2
tube 3
max
Phase 1
tube 1
tube 2
max
Phase 2
Tube 1
Tube 2
Max
Phase 3
Tube 1
Tube 2
Max
Phase 4
score 0
score 1
score 2
score 3

Figure 2

Figure 3


A
Percentage of total events
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
debris
doublets
leukocytes
FL
BL
B
Number of leukocytes acquired
5,E+06
4,E+06
3,E+06
2,E+06
1,E+06
0,E+00
FL
BL
C
% of leukocytes
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
monocytes
granulocytes
NK/T-cells
PC
MRD
normal BCP cells
Mature B-cells
FL
BL

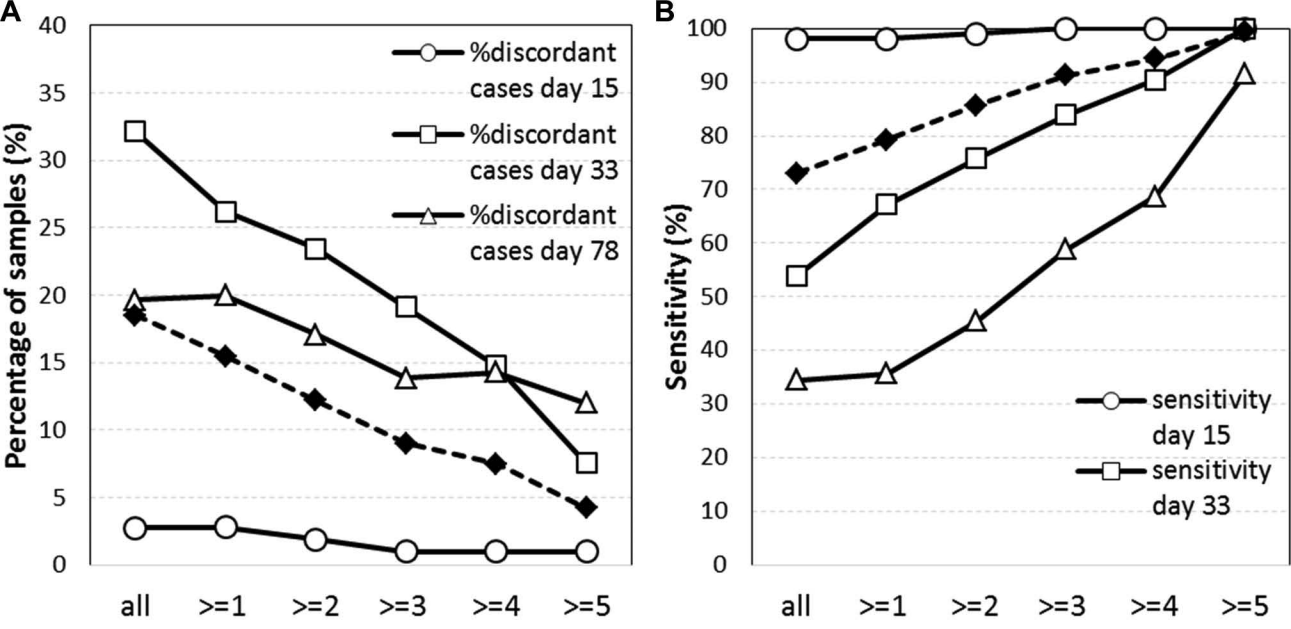
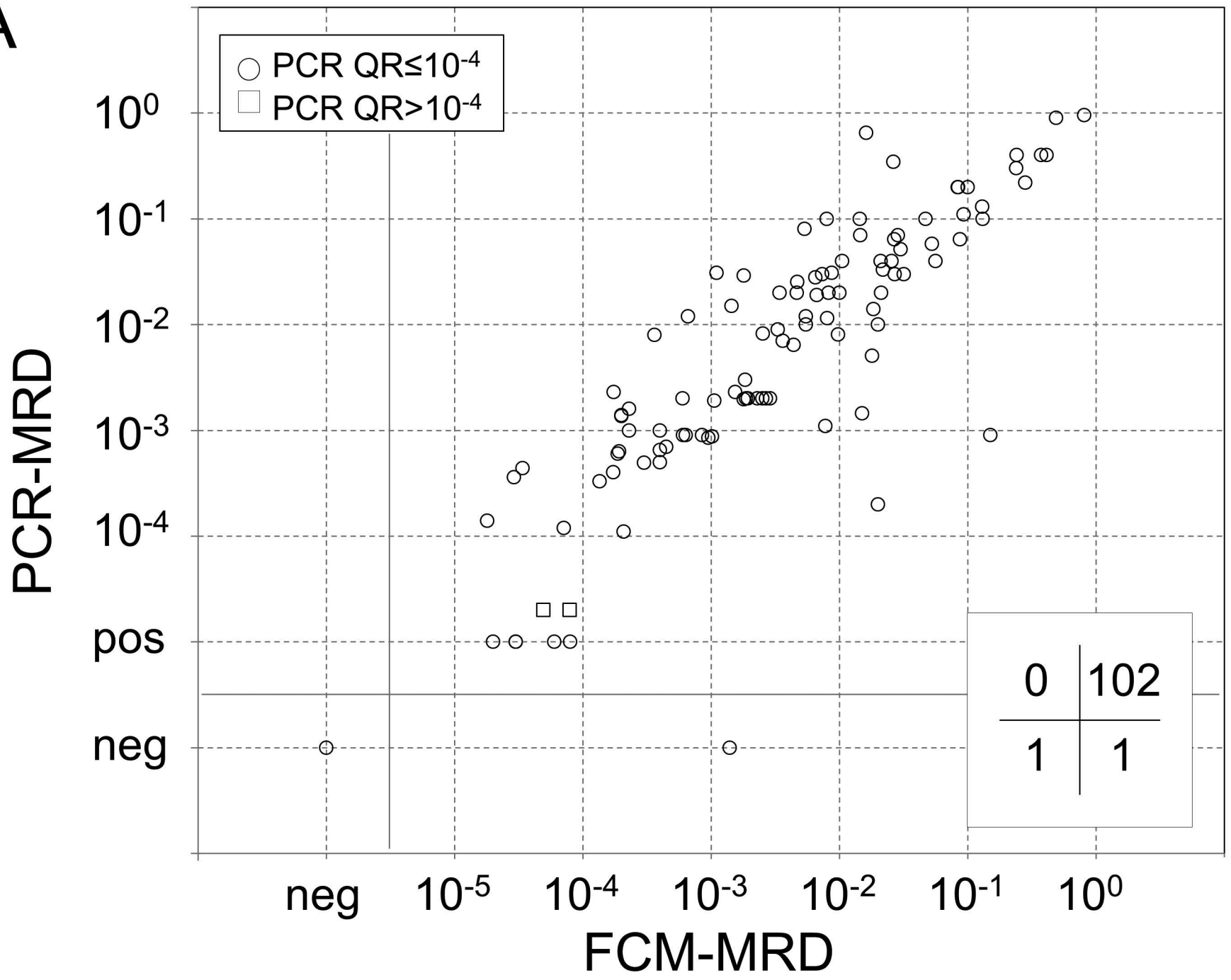
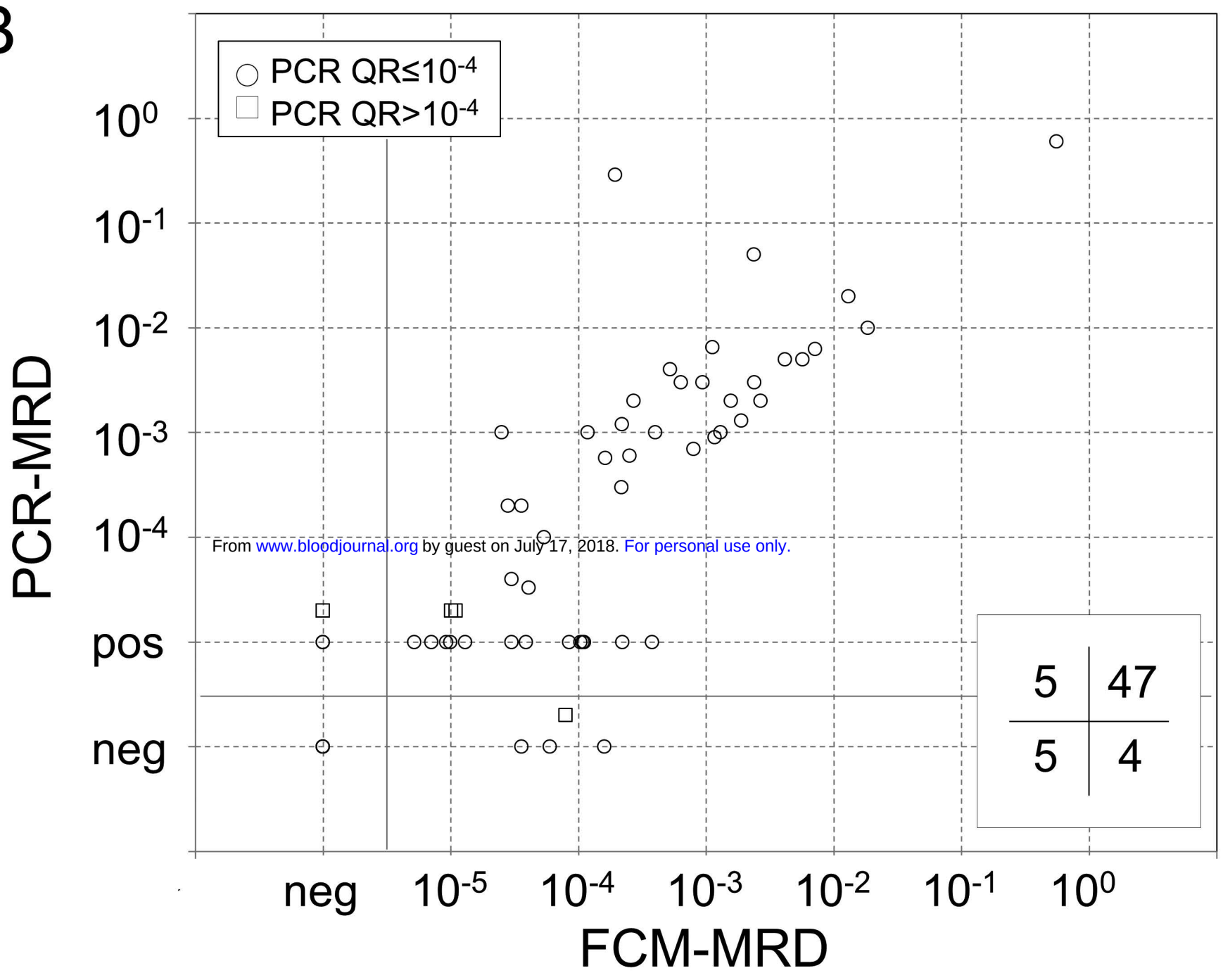

A
Percentage of samples (%)
40
35
30
25
20
15
10
5
0
%discordant cases day 15
%discordant cases day 33
%discordant cases day 78
all
>=1
>=2
>=3
>=4
>=5
Number of cells acquired (million)
B
Sensitivity (%)
100
90
80
70
60
50
40
30
20
10
0
sensitivity day 15
sensitivity day 33
all
>=1
>=2
>=3
>=4
>=5
Number of cells acquired (million)

Figure 4

A




B






C

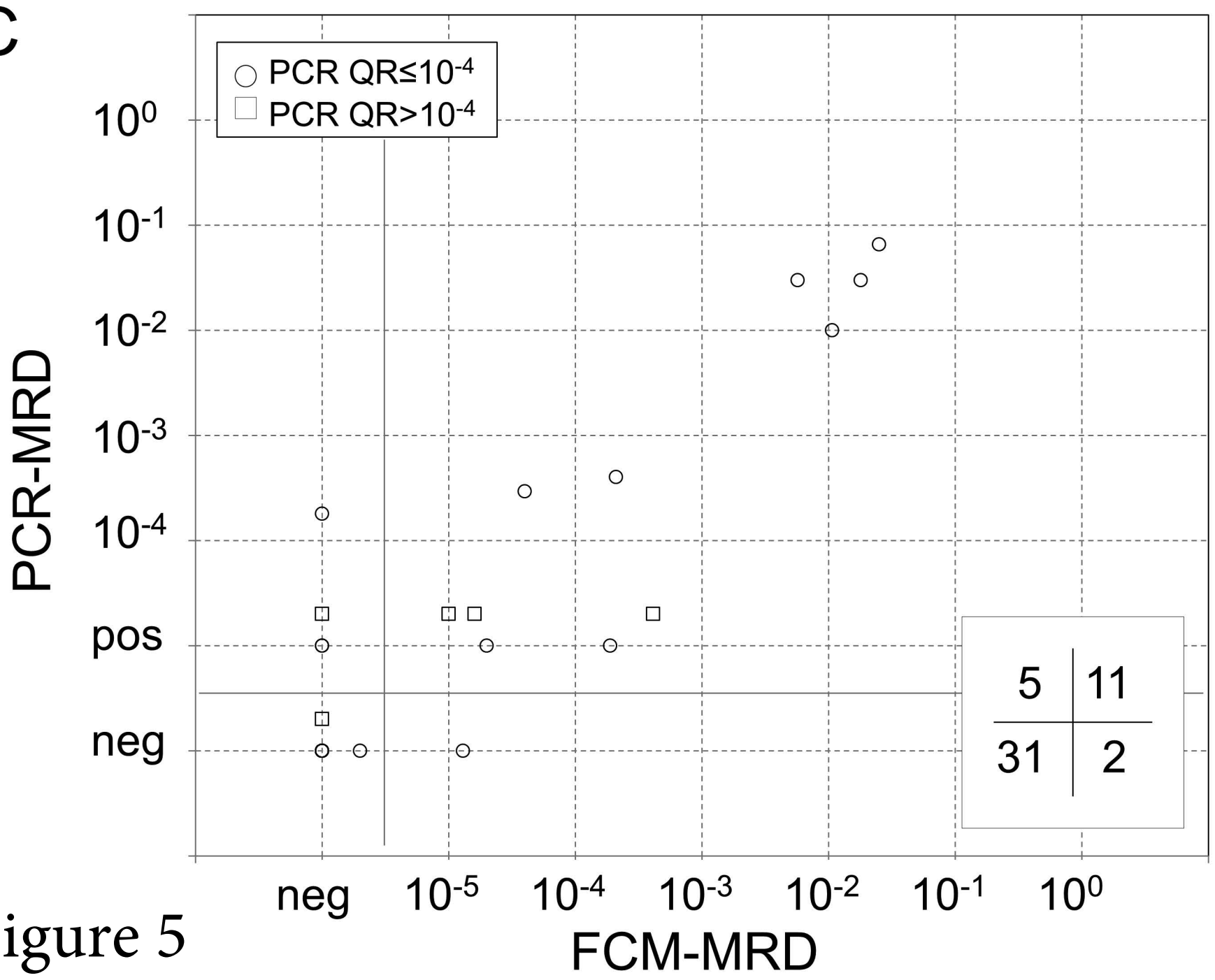



Figure 5

Prepublished online November 30, 2016;
doi:10.1182/blood-2016-07-726307

# Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia

Prisca Theunissen, Ester Mejstrikova, Lukasz Sedek, Alita J. van der Sluijs-Gelling, Giuseppe Gaipa, Marius Bartels, Elaine Sobral da Costa, Michaela Kotrová, Michaela Novakova, Edwin Sonneveld, Chiara Buracchi, Paola Bonaccorso, Elen Oliviera, Jeroen G. te Marvelde, Tomasz Szczepanski, Ludovic Lhermitte, Ondrej Hrusak, Quentin Lecrevisse, Georgiana Emilia Grigore, Eva Fronková, Jan Trka, Monika Brüggemann, Alberto Orfao, Jacques J.M. van Dongen and Vincent H.J. van der Velden



Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.