Prospective Validation of a New Method of Monitoring Minimal Residual Disease in Childhood Acute Myelogenous Leukemia

Daniel Steinbach1, Peter Bader2, Andre Willasch2, Stephan Bartholomae1, Klaus-Michael Debatin1, Martin Zimmermann3, Ursula Creutzig3, Dirk Reinhardt3, and Bernd Gruhn4

Abstract

Purpose: This study evaluated the prognostic impact of a novel, simple, and standardized assay for monitoring minimal residual disease (MRD) in pediatric acute myelogenous leukemia (AML).

Experimental Design: The expression of seven leukemia-associated genes (WT1, PRAME, CCL23, GAGED2, MSLN, SPAG6, and ST18) was measured by TaqMan Low Density Arrays in 112 patients and 52 healthy controls. Patients were treated according to the multicenter study AML-BFM 2004. Samples were collected prospectively at standard time points. The laboratory that measured MRD was blinded to patient outcome.

Results: Relapse-free survival (RFS) was 95% (N = 19; SE = 5%) if expression of all genes was down to normal on day 15, 63% (N = 41; SE = 8%) if expression was normalized on day 28, and 38% (N = 21; SE = 11%) in patients who still showed elevated expression on day 28. The prognostic impact of MRD remained significant (P = 0.002) when patients were stratified for the AML-BFM 2004 risk group. Multivariate analysis identified the MRD risk group and day 28 cytology as the only independent prognostic factors. Patients with a cytologic nonremission on day 28, which was confirmed by MRD, had a dismal prognosis. Only 1 out of 8 patients survived without relapse.

Conclusions: This novel method of monitoring MRD has a strong prognostic impact that is independent from established risk factors in childhood AML.

Introduction

Major improvements in the treatment of childhood acute myelogenous leukemia (AML) have been achieved over the last 40 years. During this period, the survival rates changed from less than 20% to more than 60%. This improvement was mainly achieved by an enormous intensification of chemotherapy (1, 2).

From early studies, we know that some patients can be cured with much less therapy and hence less toxicity. Other patients' outcome is fatal despite the most intense therapy. The design of risk-adapted therapies represents one of the greatest challenges in the treatment of pediatric AML.

In acute lymphoblastic leukemia (ALL), the monitoring of minimal residual disease (MRD) has become the most important determinant for risk-adapted therapy (3, 4). In AML, the best method of monitoring MRD is still a matter of debate. In some patients, leukemic cells show specific chromosomal rearrangements, which are detectable by PCR or FISH (5–10). Flow cytometry can be used to detect a combination of antigens present on leukemic cells. Monitoring residual leukemic cells with this method can provide strong prognostic information, but it is difficult to standardize and results strongly depend on the laboratory that is performing the analysis (11–16).

Measuring the expression of genes, which are overexpressed in leukemic cells compared with healthy bone marrow cells, is another way of monitoring MRD in AML. The Wilms tumor gene (WT1) has been studied intensively as MRD marker either alone or in combination with flow cytometry (17–22). The preferentially expressed antigen in melanoma (PRAME) was also suggested as a possible MRD marker (23, 24).

In a retrospective analysis, we have identified a set of seven genes (25), including WT1 and PRAME, which allowed monitoring MRD in a larger proportion of patients. These seven genes are vastly overexpressed in AML and they are expressed at very low levels in healthy bone marrow as well as in leukemia-free bone marrow that is recovering from chemotherapy (25).

In the present study, we determined the prognostic relevance of monitoring MRD by those seven genes in a prospective multicenter setting. The expression of the genes was measured using TaqMan Low Density Arrays, which have the advantage of being simple, standardized, and commercially available.

Materials and Methods

Patients and healthy controls

Pretreatment samples from 112 children with primary AML were analyzed. The initial diagnosis of AML was determined by standard methods (1). All patients were treated according to the multicenter study AML-BFM 2004 (1). The main patient characteristics are summarized in Table 1. We included all patients...
in the MRD analysis who were treated in the main study and in whom a sufficient amount of cryopreserved leukemic cells could be sent to the MRD laboratory. The follow-up samples were also obtained from bone marrow punctures that were part of routine diagnostics. Standard time points for diagnostic bone marrow punctures were day 1, day 15, and before the second cycle of chemotherapy. The scheduled day for the start of the second cycle was day 28.

Samples of healthy bone marrow were obtained from 52 healthy individuals who donated for bone marrow transplantation. Written consent was given for the use of all samples. This study was approved by institutional review board and was conducted in accordance with the Declaration of Helsinki.

AML-BFM 2004 risk group definition
Chemotherapy in study AML-BFM 2004 consisted of four courses for standard-risk (SR) patients and five courses for high-risk patients (1). SR was defined as FAB (French–American–British) M1/M2 with Auer rods; FAB M4eo or favorable cytogenetics [t(8;21)/AML1-ETO or inv(16) or t(16;16) and/or CBFB/MYH11]; bone marrow blasts <5% on day 15; or FAB M3 (all patients). High risk was defined as all others. SR patients were reclassified to the high-risk group if FLT3-ITD positive.

Sample collection and processing
Anticoagulated bone marrow samples of all patients were sent to the laboratory of the AML-BFM study center in Hannover by overnight mail. That was part of the routine diagnostic workup, including morphology, flow cytometry, cytogenetics, cytochemistry, etc. (1). If enough material was left over, samples were cryopreserved in liquid nitrogen. Aliquots of the frozen cells were then sent to the Jena University Hospital on dry ice for MRD analysis.

After thawing the cells, total RNA was isolated using the RNeasy Mini Kit including DNase digestion (Qiagen). The amount of RNA was measured by photometry. RNA (0.5 μg) was transcribed into cDNA using Omniscript (Qiagen).

TaqMan real-time PCR
TaqMan Low Density Arrays were performed according to the manufacturer’s instructions using the Applied Biosystems 7900HT Fast Real Time PCR System (Applied Biosystems). The MRD markers were chemokine C-C motif ligand 23 (CCL23), WT1, PRAME, suppression of tumorigenicity 18 (ST18), sperm associated antigen 6 (SPAG6), mesothelin (MSLN), and G antigen family D member 2 (GAGED2).

The expression of the MRD genes was standardized for the expression of three housekeeping genes, beta-actin (ACTB),

Table 1. Initial patient data

<table>
<thead>
<tr>
<th>Number of patients</th>
<th>Median age in years (range)</th>
<th>Median WBC count in 10^9/L (range)</th>
<th>Risk group SR/HRa</th>
<th>Bone marrow blasts day 15 &gt;5%</th>
<th>FAB type</th>
<th>Auer rod</th>
<th>inv(16) (CBFβ/MYH11)</th>
<th>t(15;17) (PML-RARα)</th>
<th>t(8;21) (AML1-ETO)</th>
<th>MLL rearrangement</th>
<th>Trisomy 8</th>
<th>Normal karyotype</th>
<th>Complex karyotypesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>112</td>
<td>9.0 (0–17.8)</td>
<td>52 (0.9–590)</td>
<td>46/66</td>
<td>21</td>
<td>M0</td>
<td>38</td>
<td>14</td>
<td>7</td>
<td>15</td>
<td>19</td>
<td>6</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>40</td>
<td>8.9 (0–16.7)</td>
<td>80 (2.3–585)</td>
<td>15/25</td>
<td>5</td>
<td>M1</td>
<td>19</td>
<td>14</td>
<td>7</td>
<td>4</td>
<td>19</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>68</td>
<td>9.0 (0–17.5)</td>
<td>52 (0.9–585)</td>
<td>31/57</td>
<td>10</td>
<td>M2</td>
<td>28</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>11</td>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M3</td>
<td>27</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>13</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M4</td>
<td>23</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M6</td>
<td>38</td>
<td>15</td>
<td>2</td>
<td>23</td>
<td>23</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>4</td>
<td>11</td>
<td>1</td>
<td>13</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>13</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
<td>9</td>
<td>13</td>
<td>1</td>
<td>13</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*SR and HR (high risk) according to AML-BFM 2004 (1).

bThree or more aberrations, at least one structural aberration, without favorable genetics, without MLL rearrangement.
beta-2-microglobulin (B2M), and Abelson gene 1 (ABL1). All three genes were shown to be expressed with little variability in healthy hematologic cells and AML (26). The geometrical mean of the three standardizations was used for the calculation of the expression of the MRD markers. All housekeeping genes and MRD genes were measured in triplicate.

The assay IDs were Hs99999903_m1 for ACTB, Hs00187842_m1 for B2M, Hs00245445_m1 for ABL1, Hs00270756_m1 for CCL23, Hs00542625_m1 for SPAG6, Hs00196132_m1 for PRAME, Hs00608494_m1 for ST18, Hs00240913_m1 for WT1, Hs00245879_m1 for MSLN, and Hs00220764_m1 for GAGED2.

Statistical methods
Kaplan–Meier statistics and log-rank tests were calculated to estimate the prognostic significance of MRD results. Multivariate Cox regression analysis was performed to analyze prognostic relevance of multiple risk factors. Relapse-free survival (RFS) was defined as the time from diagnosis to first relapse of AML or death of any course. RFS was 0 in patients who failed to achieve remission. Patients were censored in case of a secondary malignancy. Patient characteristics in children with initial samples, children with samples from day 15, and children with samples from day 28 were compared using the Kruskal–Wallis test or metric variables and Fisher exact test for dichotomous variables. All P values are given for two-sided tests. All calculations were performed using the IBM SPSS Statistics 21 program.

Results
Healthy controls and definition of units of expression
The expression of the seven MRD markers was first analyzed in 52 samples of healthy bone marrow. We defined units of expression to achieve comparability between the different MRD markers. For each marker, 1 unit was defined as the 90th percentile of the expression of this marker in healthy controls. The 95th percentile for all seven markers was below 3 units. Therefore, an expression above 3 units was defined as abnormal expression. Table 2 shows how units of expression are calculated for each marker given the threshold cycle of the marker itself and the threshold cycles of the three housekeeping genes.

Pretreatment samples
Pretreatment samples from 112 children were analyzed. Table 2 shows how often each MRD marker reached levels above 100, 30, and 10 units, respectively. The number of patients in whom the marker was expressed at higher levels than all other markers is also given.

The best marker was higher than 100 units in 66% of the patients and higher than 30 units in 83% of the patients. Using the set of seven genes allows for a suitable MRD marker in many more patients than any of the markers alone (Table 2).

Prognostic impact of MRD at different time points
Follow-up samples were available from 81 patients. Day 15 was available in 40 patients and day 28 in 68 patients. At both time points, patients with a normalized expression of all markers had a significantly superior prognosis compared with patients who showed an elevated expression of at least one marker (Fig. 1).

On the basis of the results of day 15 and day 30, three risk groups were prospectively defined: MRD low risk: normalized expression of all markers on day 15 MRD intermediate risk: all patients not qualified for either high risk or low risk MRD high risk: elevated expression of at least one marker on day 28

Both, RFS and overall survival (OS) differed significantly between those three groups (Fig. 2).

Multivariate analysis
To analyze whether the prognostic impact of MRD stratification provides additional information to the established risk group definition, Kaplan–Meier statistics and log-rank tests were repeated, using the AML-BFM 2004 risk group as stratification variable (Fig. 3). MRD risk groups had a strong and independent prognostic impact (P = 0.001). Multivariate Cox regression analysis was performed including the following variables: MRD risk group, AML-BFM 2004 risk group, favorable cytogenetics [t(8;21), inv(16) or t(15,17)], MLL rearrangement, FLT3 internal tandem duplication, cytology day 15 (<5% leukemic cells in bone marrow) and cytology day 28 (<5% leukemic cells in bone marrow). Only the MRD risk group (P < 0.001) and day 28 cytology (P < 0.001) remained as independent prognostic factors.

The combination of the MRD risk group and cytology day 28 identified a small group of patients with dismal prognosis. Eight patients showed a cytologic nonremission on day 28, which was confirmed by MRD. Only 1 of these patients survived without relapse. Patients had a much better prognosis than relapsed patients (P < 0.001).

Table 2. Expression of MRD markers in initial bone marrow samples from 112 children with AML

<table>
<thead>
<tr>
<th>Marker</th>
<th>Calculation of expression in units</th>
<th>&gt;100 units</th>
<th>&gt;30 units</th>
<th>&gt;10 units</th>
<th>Best markera</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST18</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>10</td>
<td>17</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>CCL23</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>0</td>
<td>7</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>GAGED2</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>23</td>
<td>25</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>WT1</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>18</td>
<td>42</td>
<td>57</td>
<td>23</td>
</tr>
<tr>
<td>PRAME</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>29</td>
<td>38</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td>MSLN</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>15</td>
<td>32</td>
<td>43</td>
<td>9</td>
</tr>
<tr>
<td>SPAG6</td>
<td>2[(CtACTB/3+CtMRD)]</td>
<td>4</td>
<td>12</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Best markerb</td>
<td></td>
<td>66</td>
<td>85</td>
<td>96</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: CT, threshold cycle.

aIndicates the number of patients in whom the respective marker was expressed at a higher level than all other markers.
bIndicates the number of patients in whom the best marker reached the respective threshold.
(P = 0.005) if they showed a cytologic nonremission, which was not confirmed by MRD, or if they still had elevated MRD markers on day 28 but were in cytologic remission (Fig. 4).

Definition of MRD risk groups by trend of expression

The definition of MRD risk groups by cutoffs for normal versus elevated expression was prospectively determined. After opening...
the patient data to the laboratory that performed the MRD analysis, we also analyzed other ways to define the risk groups. For instance, all calculations were repeated using the following definition:

- **MRD low risk**: more than 10-fold reduction of all initially elevated markers on day 15
- **MRD intermediate risk**: all patients not qualified for either high risk or low risk
- **MRD high risk**: less than 10-fold reduction of at least one initially elevated marker on day 28

The groups that resulted from this definition were largely overlapping with the original definition. The Kaplan–Meier graphs and the P values were very similar. There was no improvement compared to the original definition.

**WT1 alone versus the set of genes**

WT1 alone is an established MRD Marker (17–22). The present study included 42 follow-up samples that were classified as MRD positive, i.e., at least one of the seven genes was elevated (21 samples from day 15 and 21 samples from day 28; Fig. 1). The gene that most frequently indicated MRD positivity was WT1 but it was only positive in 20 (48%) of the 42 samples, i.e., the set of genes allowed detecting more than twice as many MRD positive samples than WT1 alone. WT1 was followed by PRAME (19 samples), GAGED2 (15 samples), MSLN (11 samples), SPAG6 (11 samples), ST18 (8 samples), and CCL23 (5 samples). A separate analysis of those patients in whom MRD positivity was detected by other genes but not by WT1 still showed a significant prognostic impact for day 15 (P = 0.006) and for day 28 (P = 0.036). The only gene that could be left out was CCL23 because all of the follow-up samples where MRD positivity was indicated by CCL23 were also positive for at least one other gene.

**Excluding patients with a translocation as MRD marker**

The chromosomal aberrations t(15;17), t(8;21), inv(16), t(4;11) and t(9;11) are possible MRD markers in AML that might be superior to our set of genes in patients in whom they are applicable. We therefore performed a separate analysis of all patients in whom none of these translocations were present. Also in this subgroup, we observed strong and statistically significant (P = 0.032) differences between MRD low risk (n = 11;
RFS = 89%; SE = 11%), MRD intermediate risk (n = 25; RFS = 45%; SE = 12%), and MRD high risk (n = 13; RFS = 27%; SE = 13%).

**Discussion**

This study provides a prospective evaluation of the prognostic impact of monitoring treatment response in AML by measuring the expression of seven leukemia-associated genes. We have previously shown that these seven genes are vastly overexpressed in many patients with AML and that they are expressed at very low levels in healthy bone marrow as well as in leukemia-free bone marrow that is recovering from chemotherapy (25).

The expression of all seven marker genes and the three housekeeping genes was measured on the same TaqMan Low Density Array. This makes it a very quick and easy method of monitoring MRD. TaqMan Low Density Arrays are fully standardized. Using the assay IDs given above, any group can order the same array and perform the same analysis. This is an important advantage of this method compared with flow cytometry where results from different laboratories are difficult to compare.

Monitoring MRD using individual DNA marker for each patient (translocations, rearrangements, single gene mutations) should have the advantage of being more specific. However, to cover as many patients, this would be a much more complex approach with many different assays and it would not necessarily result in a stronger prognostic impact. Especially at the early time points, a reduction in leukemia-specific gene expression might have a different meaning than the eradication of cells carrying leukemia-specific DNA.

Because we could only include patients in whom sufficient material was sent to the MRD laboratory, a possible selection bias is a limitation of this study. However, the patient characteristics given in Table 1 are very similar to the total population of study AML-BFM 2004 (1), suggesting that this group is a good representation. It is also shown in Table 1, that patient characteristics were similar in patients in whom day 15 or day 28 samples were available.

Patients who reach a normal expression of all seven genes by day 15 have an excellent prognosis. This holds true, even if they have other established high risk characteristics (Fig. 3). Future clinical trials should consider a controlled reduction of treatment intensity in this group.

Patients with the combination of cytologic nonremission on day 28 and MRD high risk had an extremely poor prognosis. This result needs to be interpreted with particular caution because of the small number of patients and because this risk group was not prospectively defined but resulted from the multivariate analysis. A possible explanation is that only patients with nonremission and elevated MRD markers on day 28 have a really high leukemic burden before the second course of chemotherapy. Among the patients with nonremission on day 28 but low MRD, there seems to be a substantial number in whom a heavily regenerating, left shifted hematopoiesis cannot be distinguished from malignant proliferation. This hypothesis is supported by the finding that there were 7 patients with low MRD on day 15 but nonremission on day 28. Only one of these patients experienced a relapse.

Standard risk patients with high MRD confirmed cytologic nonremission on day 28 should be shifted to the high-risk group. If our result for these patients is fully corroborated, they will require even more intensive therapy or should be offered phase I/II studies.

This study shows that this novel method of monitoring MRD has a strong prognostic impact, which is independent from established risk factors in childhood AML. It was recently shown that one of these genes (WT1) can improve MRD risk stratification when used in combination with flow cytometry (19), which is so far the best established single method of monitoring MRD in the majority of patients with AML. Future studies should evaluate the potential of combining this whole set of genes with other methods of monitoring MRD.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Authors’ Contributions**

Conception and design: D. Steinbach, P. Bader, B. Gruhn
Development of methodology: D. Steinbach, S. Bartholomae, B. Gruhn
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): D. Steinbach, P. Bader, A. Willasch, S. Bartholomae, K.-M. Debatin, D. Reinhardt, B. Gruhn
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D. Steinbach, S. Bartholomae, U. Creutzig, B. Gruhn, M. Zimmermann
Writing, review, and/or revision of the manuscript: D. Steinbach, P. Bader, A. Willasch, S. Bartholomae, K.-M. Debatin, U. Creutzig, B. Gruhn
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D. Steinbach, D. Reinhardt, B. Gruhn
Study supervision: D. Steinbach, K.-M. Debatin, D. Reinhardt, B. Gruhn

**Grant Support**

The authors thank the Deutsche Krebshilfe for financial support to this study (grant 107942).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.


**References**


Minimal Residual Disease in Childhood AML
Prospective Validation of a New Method of Monitoring Minimal Residual Disease in Childhood Acute Myelogenous Leukemia

Daniel Steinbach, Peter Bader, Andre Willasch, et al.


Updated version Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-1999

Cited articles This article cites 26 articles, 15 of which you can access for free at: http://clincancerres.aacrjournals.org/content/21/6/1353.full#ref-list-1

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.