Strikingly different molecular relapse kinetics in *NPM1c, PML-RARA, RUNX1-RUNX1T1* and *CBFB-MYH11* acute myeloid leukemias

Hans Beier Ommen¹, Susanne Schnittger², Jelena V. Jovanovic³, Ingrid Beier Ommen¹, Henrik Hasle⁴, Mette Østergaard¹, David Grimwade³, and Peter Hokland¹

From

1) Laboratory of Immunohematology, Department of Hematology, Aarhus University Hospital, Århus, Denmark, 2) Munich Leukemia Laboratory, Munich, Germany, 3) Department of Medical & Molecular Genetics, King’s College London School of Medicine, London, United Kingdom, and 4) Department of Pediatrics, Aarhus University Hospital Skejby, Århus, Denmark

Short title: Relapse kinetics in AML

Scientific section: Myeloid Neoplasia

Correspondence: Prof. Peter Hokland (phokland@ki.au.dk)
Abstract

Early relapse detection in AML is possible using standardized real time quantitative polymerase chain reaction (RQ-PCR) protocols. However, optimal sampling intervals have not been defined and are likely to vary according to the underlying molecular lesion. In 74 patients experiencing hematological relapse (HR) and harboring aberrations amenable to RQ-PCR (mutated NPM1 [designated NPM1c], PML-RARA, RUNXI-RUNXIT1 and CBFB-MYH11), we observed strikingly different relapse kinetics. The median doubling time of the CBFB-MYH11 leukemic clone was significantly longer (36 days) than that of clones harboring other markers (RUNXI-RUNXIT1 14 days, PML-RARA, 12 days and NPM1c, 11 days, P<0.001). Furthermore, we employed a mathematical model to determine frequency of relapse detection (RDF) and median time from detection of minimal residual disease (MRD) to HR (tm) as a function of sampling interval length. For example, to obtain an RDF of 90% and a tm of 60 days, blood sampling every 6th month should be performed for CBFB-MYH11 leukemias. By contrast, in NPM1c+/FLT3-ITD-, NPM1c+/FLT3-ITD+, RUNXI-RUNXIT1 and PML-RARA leukemias bone marrow sampling is necessary every 6th, 4th, 4th and 2nd month, respectively. These data carry important implications for the development of optimal RQ-PCR monitoring schedules suitable for evaluation of MRD-directed therapies in future clinical trials.
Introduction

Relapse remains the event that heralds ultimate treatment failure for most acute myeloid leukemia (AML) patients.\(^1\) Thus, while reacquisition of complete remission (CR) is often possible, it always poses a greater therapeutical challenge than the initial cytoreduction, most probably because of selection for therapy-resistant clones.\(^1, 2\)

Consequently, detection of impending relapse remains a major challenge in these patients. Several tools with varying sensitivities are available for this task, most notably multicolor flow cytometry (sensitivity as low as 0.01\%)\(^3\) and real-time quantitative RT-PCR (RQ-PCR) (sensitivity as low as 0.0001\%).\(^4, 5\) RQ-PCR assays are being applied to fusion transcripts, such as \textit{PML-RARA},\(^6-11\) \textit{CBFB-MYH11},\(^8, 10, 12-14\) \textit{RUNX1-RUNXIT1},\(^8, 10, 15-19\) and \textit{DEK-CAN},\(^20-22\) over-expressed genes, such as \textit{WT},\(^23-26\) and \textit{PRAME},\(^27\) or mutated genes such as \textit{NPM1}.\(^28-32\) There is evidence to suggest that assay sensitivity not only varies between MRD markers but can also differ significantly between patients possessing the same molecular target.\(^9, 10, 32\) Furthermore, as follow-up MRD samples are often collected sparsely, even in prospective studies, it has hitherto been difficult to provide firm recommendations for sampling, not only regarding intervals, but equally so with regard to source of sample material (i.e. bone marrow (BM) or peripheral blood (PB)).

To address these issues we recently employed the Wilms tumor gene 1 (\textit{WT1}) as a molecular marker to derive a mathematical model enabling the delineation of a number of quantitative parameters related to relapse kinetics, such as the power to detect molecular relapse (MR) (relapse detection fraction (RDF)) and the median time (\(t_m\)) from molecular positivity to hematological relapse (HR) for different sampling intervals.\(^33\)

Given that the sensitivity of the \textit{WT1} assay is inferior to assays detecting recurrent molecular aberrations such as fusion transcripts and mutations, we here present data collected from three centers handling large numbers of minimal residual disease (MRD) samples. By analyzing relapsing patients with several pre-relapse samples available and applying our mathematical model, we have now been able to delineate relapse kinetics in patients harboring four different target transcripts, namely \textit{RUNX1-RUNXIT1, CBFB-MYH11, PML-RARA}, and mutated \textit{NPM1} (designated \textit{NPM1c}).\(^34\) Application of the model which takes into account differences in assay sensitivity and relapse kinetics will...
allow development of monitoring schedules tailored according to the molecular lesion, suitable for evaluating the clinical utility of MRD-directed therapies in multicenter clinical trials.

Materials and methods

Patient samples

Patient samples were analyzed at the Laboratory of Immunohematology (IHL), Aarhus University Hospital from October 1995 to November 2007, the Munich Leukemia Laboratory (MLL) from July 2005 to March 2008 and the Department of Medical and Molecular Genetics (MMG)/Molecular Oncology Diagnostics Unit, Guy’s Hospital, London from May 2002 to March 2009. All patients were diagnosed and followed for MRD detection using standard published methods.8-10, 32, 35 MRD analyses were conducted with informed patient consent in accordance with the Declaration of Helsinki and subject to ethics committee approval from all participating institutions. The patients with PML-RARA acute promyelocytic leukemia were treated with all-trans retinoic acid (ATRA) and anthracycline-based chemotherapy as detailed in Grimwade, et al.9 Non-APL patients were treated with standard combination therapy regimens, as described.10, 32, 33, 36

MRD determination and reporting

MRD values were calculated using the ΔΔC(t) or the absolute quantification method using plasmid standards (Ipsogen, Marseille) as described by Beillard, et al37 according to the laboratory and molecular target (for NPM1c quantification, see Schnittger, et al32). Samples with low quality RNA (defined as a threshold cycle number of the control gene β-2-microglobulin exceeding 25 and/or a threshold cycle number of the control gene Abelson exceeding 30) were excluded.

Assays were run in duplicate (NPM1c) or triplicate (other markers), and in accordance with Europe Against Cancer criteria amplification in at least two of three replicates with C(t)-values ≤ 40 (threshold 0.1 [0.05 for PML-RARA]) was required to define a result as PCR positive for the MRD marker in question.37 For the depiction of relative values on
an absolute scale, data were transformed from $\Delta C(t)$ values using the PCR efficiencies found in routine laboratory assay testing.

**Definition of MR, HR and molecular complete remission (mCR)**

Definitions of MR, HR and mCR followed the recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia with the modifications described in the results section. Thus, patients were considered having a MR at first recurrence of MRD marker PCR positivity, except for those followed using the $NPM1c$ or $RUNX1-RUNX1T1$ molecular markers where an MRD level was above $5 \times 10^{-5}$ or $1 \times 10^{-4}$, respectively was required to define MR. Similarly, patients with the two latter markers were considered in mCR if PCR expression levels dropped to below these thresholds.

**Inclusion and exclusion criteria**

To be eligible for relapse kinetics analysis, patients needed to enter CR, have PCR samples taken after the discontinuation of chemotherapy but before HR and not receive preemptive treatment upon MR. Thus, patients PCR positive after treatment, who were not given further high-dose chemotherapy could be included in this part of the study (n=5).

To be eligible for relapse modeling based on PCR conversion, patients needed to enter mCR, have samples taken the last year before HR and not receive preemptive treatment upon MR. Thus, patients experiencing a HR with no previous positive samples could be included (n=28).

One patient with a CNS relapse was excluded from both analyses.

**Statistical analyses**

To compare relative levels of expression of the molecular target at diagnosis and at relapse as well as increments in normalized leukemic transcripts prior to HR, Wilcoxon rank-sum test was used. To compare the performance of PB vs. BM testing, the number
of paired MRD measurements where only the BM was positive was compared to the number of MRD measurements were only the PB was positive using the binomial distribution. These analyses were restricted to paired samples in which the BM and PB yielded RNA of comparable quality, as indicated by the respective levels of control gene expression.

**Modeling of relapse patterns**

Relapse patterns were modeled as described in Ommen, et al.\(^\text{33}\) Briefly, pre-relapse samples were divided into monthly intervals based on how long before time of HR they were taken. Patients were considered RQ-PCR negative or positive in all intervals between two negative or positive samples, respectively. The fraction of positive samples in each interval was then plotted against time to relapse. For a given positive sampling interval, I, the chance of detecting the relapse, RDF, was then given by:

\[
RDF = \frac{\int_0^{I} F(t) \, dt}{I},
\]

where \(F(t)\) describes the distribution of the time span from the leukemic burden exceeds the sensitivity of a given MRD marker to HR. This function can be approximated based on the distribution of individual patients’ conversion to PCR positivity prior to HR, as described in Ommen, et al.\(^\text{33}\) The median time from MR to HR, \(t_m\), can be found by solving the integral equation:

\[
\int_{t_0}^{t_m} (F(t) - F(t + I)) \, dt = \int_{t_0}^{t_m} (F(t) - F(t + I)) \, dt,
\]

where \(t_0\) is the intercept with the x-axis.\(^\text{33}\) In this manner, RDFs and \(t_m\)'s for sampling intervals 1, 2, 3, 4, 6, 9 and 12 months were found for each molecular marker in both PB and BM. \(F(t)\) is approximated using high-grade polynomials for each of the MRD markers. The used polynomials can be found in the data supplement.
Results

MRD marker expression at diagnosis and relapse

A major determinant of the capacity for RQ-PCR to detect residual AML in a background of normal hematopoietic cells is assay sensitivity, which depends on the relative level of expression of the MRD target in leukemic blasts (reviewed in Freeman, et al\(^5\)). To compare assay sensitivities, we determined the MRD marker expression in relation to control gene expression in bone marrow at diagnosis from 365 patients (151 PML-RARA, 31 CBFB-MYH11, 42 RUNXI-RUNXITI and 141 NPM1c) and at HR for all relapsing patients.

As can be seen from supplementary Figure 1 and supplementary Table 1, median expression of NPM1c was higher than that of the fusion transcript MRD markers that exhibited comparable levels of expression. While diagnosis levels were generally higher than relapse levels, median values were within the same order of magnitude, allowing the use of diagnosis expression level as a measurement for MRD assay sensitivity.

Low-level positive expression in patients in continuous complete remission

Simple qualitative PCR positivity has been reported in a number of cases in patients who did not later relapse.\(^{39-41}\) To investigate this situation in this large cohort of patients, we analyzed the number of positive samplings in patients who had earlier tested negative (for PML-RARA patients, two consecutive negative samples were required, as described in Grimwade, et al\(^9\)). For all four aberrations, a small percentage of positive samples could in fact be observed, the highest fraction in NPM1c patients (26% of CCR samplings), possibly due to the higher level of MRD marker transcripts in these cells or in some cases due to cross-reactivity of the mutant assay with the wild-type allele.\(^{10}\) However, when a cut-off of 5 x 10\(^{-5}\), relative to the diagnostic level, was used to define molecular relapse (MR), only one positive reaction was not followed by a HR (out of a total of 104 CCR determinations). Applying a similar cut-off level for RUNXI-RUNXITI patients at 1 x 10\(^{-4}\) resulted in 1 positive reaction out of 87 CCR determinations not followed by a HR. By contrast, positive samplings in CCR using the CBFB-MYH11 molecular marker occurred in only three samples accounting for 4.8% of the CCR samples. For this aberration we therefore chose not to use an MR cut-off level, and resorted to the
requirement of two consecutive positive samplings to define an MR given that \textit{CBFB-MYH11} relapses are generally indolent allowing for such an approach.\textsuperscript{42} For \textit{PML-RARA} follow-up, positive samples were rare, and we chose not to use a cut-off threshold in the case of this aberration. For the two assays (\textit{NPM1c} and \textit{RUNXI-RUNXIT1}) where thresholds were introduced to exclude very low level positivity not followed by a HR, the maximum sensitivity corresponds to the level of the threshold. As the sensitivity of the assays using the other two MRD markers is approximately $1 \times 10^{-4}$ (with a range of $1 \times 10^{-5}$ to $1 \times 10^{-3}$ reflecting differences in the relative level of expression of the MRD target in the leukemic blasts between patients, as well as variation in the quality of follow-up samples) we found that the sensitivities of MRD detection, and thereby the ability to detect MR, using the four different markers are comparable (Figure 1).

\textit{Relapse kinetics}

The advantage of employing the RQ-PCR methodology in MRD follow-up is its potential for revealing relapse kinetics in the individual patient, provided that enough RQ-PCR positive samples have been obtained prior to HR.

From Figure 2 it will be seen that the \textit{CBFB-MYH11} patients display a slower rate of rise of leukemic transcripts prior to HR than all the other aberrations (median BM doubling time 36 days, range 7.4-175 days, vs. all other aberrations, \textit{P}<0.001). Compared to this, the incremental rise in \textit{NPM1c} transcripts was more rapid (median BM doubling time 11 days, range 2.2-33 days, \textit{P}=0.0016): however there was marked heterogeneity with some patients experiencing indolent relapses while others exhibited the shortest doubling times in the data set. A major reason for this duality was found to be the presence or absence of the \textit{FLT3-ITD} aberration (\textit{NPM1c+/FLT3-ITD-}, median BM doubling time 15 days, range 2.2-33 days, \textit{NPM1c+/FLT3-ITD+}, median BM doubling time 7.4 days, range 3.0-14 days, \textit{P}=0.031). \textit{RUNXI-RUNXIT1} relapsing clone reappearance was faster than \textit{CBFB-MYH11} growth (median BM doubling time, 14 days, range 12-51 days, \textit{P}=0.105), as was \textit{PML-RARA} clone growth (median BM doubling time, 12 days, range 4.2-462 days, \textit{P}=0.093), with the exception of one patient who displayed the slowest relapse growth rate in the entire cohort (BM and PB doubling times of 462 and 54 days, respectively). Both these leukemias progressed with approximately
the same speed as the \textit{NPM1+/FLT3-ITD-} ones ($P=0.97$ and 0.93, respectively), but slower than the \textit{NPM1+/FLT3-ITD+} leukemias ($P=0.02$ and 0.05, respectively).

\textit{Relapse modeling based on conversion to PCR positivity}

In some patients subjected to longitudinal RQ-PCR testing no positive samples prior to HR will be obtained, either because of very rapid clone growth, poor sampling quality giving rise to “false negative” results or simply unfortunate scheduling of MRD assessment relative to the timing of recurrent PCR positivity. We devised a mathematical model, which takes this lack of data into account and allows inclusion of information from such patients.\textsuperscript{33} As can be seen from Figure 3 which depicts time before relapse as a function of the fraction of positive samples in each interval for each molecular marker, we were able to confirm and extend the relapse kinetics findings shown in Figure 2. Thus, \textit{CBFB-MYH11} emerged as the MRD marker with the longest lag from MR to HR (50% of the patients tested RQ-PCR positive in BM eight months before relapse, Figure 3C). By comparison, \textit{RUNX1-RUNX1T1} relapses showed more rapid relapse kinetics with 50% being positive in BM as close as 3 months prior to relapse (Figure 3D. Moreover, the dichotomy observed for \textit{NPM1c} transcript-based relapse detection was preserved in the model with a 3-month difference in when 50% of \textit{FLT3-ITD-} and \textit{FLT3-ITD+} patients became PCR positive (6.5 months \textit{vs.} 3.5 months prior to relapse, Figure 3A). For \textit{PML-RARA} positive relapses, the heterogeneity was even more pronounced, with one relapse being detectable 14 months before HR, and two being undetectable 71 days and 41 days before HR, respectively (50% detectable 3.5 months before HR, Figure 3B).

A recurring question when RQ-PCR is used for MRD detection has been to what extent PB can substitute for BM sampling. When comparing PB \textit{vs.} BM-based MR detection, it is now apparent that for the majority of \textit{PML-RARA} cases, BM sampling is superior. Thus, in 7 paired samplings MR was detected in BM only in 5 cases and in PB only in none ($P=0.031$). From Figure 3C and 3D it will on the other hand be suggested that PB and BM sampling are equally useful for \textit{CBFB-MYH11} and \textit{RUNX1-RUNX1T1}, even though the number of paired samplings was too low to draw any firm conclusions.
Taking advantage of the mathematical modeling we were finally able to calculate the relapse detection fractions (RDFs) and median time to hematological relapse (t_m) as a function of sampling interval using the formulae described in Materials and Methods (Figure 4). For NPM1c-based follow-up, modeling was done for the subgroups containing and lacking FLT3-ITD separately. As conclusions regarding NPM1c-based PB sampling would be based on very few patients (median number of MRD courses per interval 2.5), no modeling was done for NPM1c-based follow-up in this tissue. For comparative purposes and considering that some patients are negative for all the markers described here, we included WT1 data which complement the ones presented here resulting in more than 80% of AML patients having a valid molecular marker for MRD detection.25

Once more, CBFB-MYH11 leukemia displayed the most tardy relapses with high RDFs and t_m’s for sampling intervals as long as 6 months (PB; RDF 90% and t_m 180 days, BM; RDF 85% and t_m 150 days). Sampling interval of this length also yielded satisfactory results in the group of patients with NPM1c leukemia without FLT3-ITD (6 month intervals, BM: RDF 90%, t_m 120 days). As an intermediate group, for NPM1c (with FLT3-ITD) and RUNX1-RUNXI1TI leukemias application of a three to four monthly sampling frequency, still yielded satisfactory relapse detection (4 month intervals, NPM1c+FLT3-ITD+, BM, RDF 85% and t_m 65 days; RUNX1-RUNXI1TI PB; RDF 75% and t_m 55 days, BM; RDF 95% and t_m 85 days). Somewhat surprisingly, relapse detection by WT1 fell into this category, too, at least when BM was used (4 month intervals, WT1; BM, RDF 95% and t_m 75 days). Due to the occurrence of two patients who were negative in BM close to HR, PML-RARA relapses were the most difficult to detect, and two-monthly BM sampling will be necessary to obtain the relapse detection efficiencies of the other markers (two-monthly intervals, PML-RARA; RDF 95%, t_m 70 days).
Discussion
Close molecular monitoring using RQ-PCR in AML patients in CR holds the promise of detecting subclinical levels of residual disease in time to institute treatment intervention to prevent overt relapse. Despite an impressive body of data showing that RQ-PCR is excellently suited for early detection of AML relapses, few investigators have taken clinical action on these findings, though it is by now evident that the risk of a false positive result of an RQ-PCR is minimal, at least when the thresholds to exclude irrelevant low level amplification in CCR are employed, and especially so when molecular conversion is confirmed in a subsequent sample. Thus far, benefit from early salvage after conversion has been shown in PML-RARA acute promyelocytic leukemia (APL) and in a preliminary report on the use of donor-lymphocyte infusion (DLI) upon recurrent WT1 positivity after allogeneic transplantation, but has not been otherwise evaluated in non-APL AML patients.

Major reasons for this lack of translation of a powerful and technically standardized molecular method into clinical decision-making beyond APL include uncertainty as to 1) the most informative schedules for MRD monitoring, 2) the most appropriate management of confirmed molecular relapse, and 3) whether early treatment intervention is likely to confer any clinical benefit compared to retreatment in overt relapse. In order to develop optimal MRD monitoring schedules to allow reliable assessment of the clinical utility of MRD monitoring in non-APL patients within large scale clinical trials, we have analyzed data from a very large cohort of patients subject to hematological relapse (n=114).

We used two different ways of describing the behavior of the leukemic clone prior to HR. First, we used the quantitative data obtained from patients in whom MR was identified to describe relapsing clone growth before HR. Assuming a constant doubling time, we were able to compare the different leukemia subsets and show that CBFB-MYH11 leukemia displayed slower leukemic clone growth than AML with mutant NPM1 or the RUNX1-RUNXIT1 fusion and, with a single exception, PML-RARA+ APL. Interestingly, but perhaps not surprisingly, NPM1c+/FLT3-ITD+ AML displayed relapses occurring significantly faster than all other molecular subtypes studied.
In the other model presented here we examined pre-HR conversion to PCR positivity. This approach has the advantage of allowing the inclusion of information on patients for whom MR was not necessarily detected prior to HR. Also, the assumption of a constant doubling time of the malignant clone is not necessary, although the cost is the lower degree of integration of quantitative values in this model. Thus, the two models complement each other, and it is reassuring that the results using these distinct approaches were near-identical.

By directly comparing paired samples of PB and BM we were able to show that PML-RARA BM MRD testing is superior to that of PB. Moreover, in the absence of a suitable number of paired samples for the other aberrations, the PCR positivity conversion model suggests that PB and BM are comparable for MRD detection in core binding factor leukemia. This is an encouraging conclusion, as PB sampling is much less troublesome for the patients, even if slightly higher sampling frequencies are necessary. However, further analyses are needed, especially regarding NPM1c-based MRD detection in PB, as our data do not support any definitive conclusions in this matter.

One great challenge when optimizing MRD follow-up is to determine the optimal sampling interval. The model presented here allows for the evaluation of suitable sampling intervals for the different analyzed markers.

Establishing guidelines for follow-up sampling should include considerations about their predictive value of MRD assessment, the options for intervention, as well as cost-benefit estimations. Thus, PB sampling every 6 months in patients with CBFB-MYH11 leukemia from PB will result in an RDF of 90% and a tm of 180 days. Even with this sampling cadenza only 10% of relapses will be missed before hematological relapse. Moreover, if the first sample is obtained three months after discontinuation of therapy and each patient is followed for only three years (given the literature on AML relapse and that 15/16 CBFB-MYH11 relapses in the study cohort occurred during the first three years after diagnosis), based on the present study only 14 MRD samples would have to be taken to detect one MR. Given such considerations, which can be applied equally well to other RQ-PCR assays, it will be apparent that molecular monitoring might prove to be cost-effective with its promise of early, possibly less intensive, intervention.
In conclusion, we here present data enabling us to model relapse kinetics in four common AML subtypes. We show great difference between these subtypes, and show that the studied markers are generally superior to non-leukemia-specific marker WT1. Furthermore, we show that this superiority can be used in the CBF leukemias to employ PB as the sampling tissue. These data should be useful in cost-benefit calculations regarding MRD monitoring implementation, in clinical decision-making in the individual patient, and in statistical power calculations when designing clinical trials.

Acknowledgements

This study was supported by grants to PH from the Danish Medical Research Council, the Danish Cancer Society, the John and Birthe Meyer foundation, and the Karen Elise Jensen foundation. DG and JVJ gratefully acknowledge support from Leukaemia Research Fund (Great Britain) and the European LeukemiaNet. This study was conducted within the framework of the MRD Workpackage (WP12) of the European LeukemiaNet. We would like to thank Drs Morten Frydenberg and Steffen Hokland for statistical help. We thank Dr Robert Hills for helpful discussions and DG gratefully acknowledges Prof Alan Burnett and the UK National Cancer Research Institute AML Working group for access to data from APL patients treated in the Medical Research Council AML15 trial. PH thanks Prof Finn Bo Petersen, Intermountain Health Care Center, Salt Lake City, USA for helpful discussions.

Authorship

HBO and PH designed the study, analyzed the data and wrote the first draft of the manuscript. SS was responsible for the analyses performed at MLL. DG and JVJ were responsible for the analyses performed at MMG. IBO analyzed the data, HH was responsible for the clinical care of the pediatric patients followed at IHL, MØ was responsible for the analyses performed at IHL. All authors made significant contributions to the final manuscript.

Conflicts of interests

SS in part owns Munich Leukemia Laboratory (MLL)
References


Table 1: Study subjects

<table>
<thead>
<tr>
<th>MRD marker</th>
<th>NPM1c+</th>
<th>PML-RARA</th>
<th>CBFB-MYH11</th>
<th>RUNX1-RUNX1T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>180</td>
<td>406*</td>
<td>47</td>
<td>56</td>
</tr>
<tr>
<td>Patients with HR</td>
<td>54</td>
<td>22*</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Patients in CCR</td>
<td>126</td>
<td>363*</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>CCR samples</td>
<td>193</td>
<td>1865</td>
<td>151</td>
<td>170</td>
</tr>
</tbody>
</table>

**Relapse kinetics**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients included</td>
<td>22</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Patients excluded</td>
<td>32</td>
<td>13</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Reasons for exclusion:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No MR/HR sample</td>
<td>32</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Administration of pre-emptive therapy in MR</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CNS relapse</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Relapse modelling**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients included</td>
<td>31</td>
<td>13</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>Samples</td>
<td>70</td>
<td>61</td>
<td>59</td>
<td>46</td>
</tr>
<tr>
<td>Patients excluded</td>
<td>23</td>
<td>8</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Reasons for exclusion:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No mCR</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>No CR samples**</td>
<td>20</td>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Administration of pre-emptive therapy in MR</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CNS relapse</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

CCR – continuous complete remission, patients who have not experienced relapse at last point of follow-up, CR – complete remission, HR – hematological relapse, MR – molecular relapse, mCR – molecular complete remission defined as detailed in the results section, * Numbers does not add as some patients were successfully treated in 1st MR. ** in the 12 months prior to HR.
Figure Legends

**Figure 1. Low-level positive RQ-PCR samplings in patients in CCR.** X-axis MRD target, Y-axis: MRD marker expression level relative to diagnostic level, arbitrarily set to 1. Numbers below the figure depict the number of low-level positive samples, and negative samples in patients with available diagnostic levels. Suggested cut-off levels for RUNX1-RUNX1T1 and NPM1c are shown in red bars and given in numbers. Sensitivity of each test can be understood as the lower of the difference between diagnostic level and the sample quality dependent sensitivity (typically between $1 \times 10^{-5}$ and $1 \times 10^{-3}$) and cut-off level ($NPM1c$ and $RUNX1-RUNX1T1$).

**Figure 2. Relapse kinetics according to molecular lesion.** Increasing MRD marker levels are shown in patients between time of molecular relapse and diagnosis of hematological relapse (HR). Expression normalized to HR level (set to 1). X-axis: time before HR. Y-axis: Normalized MRD levels with PB and BM values depicted separately.

**Figure 3. Conversion to PCR positivity.** X-axis: time, HR at t=0. Y-axis: fraction positive samples in each pre-relapse interval. Median number of evaluable MRD profiles in each interval is shown for each molecular marker. (Range in parenthesis) Solid lines: BM, broken lines: PB. $NPM1c$: Blue: FLT3-ITD-, red: FLT3-ITD+

**Figure 4. Comparisons between different molecular markers regarding RDF and $t_m$.** X-axis: sampling time in months. Y-axis: relapse detection fraction (RDF) (top), median time from MR to HR ($t_m$) (bottom). Horizontal black lines: top panels: RDF equaling 90%. Bottom panels: $t_m$ equaling 60 days. Blue lines in top panel represent sampling interval necessary to ensure 90% relapse detection. Blue lines in bottom panel show corresponding $t_m$ values. Solid lines, BM, Broken lines, PB, $NPM1c$-based MRD follow-up: black line FLT3-ITD-, red line FLT3-ITD+. PML-RARA based MRD follow-up: PB sampling Intervals above 6 months not shown as model handles situation of long sampling intervals in very heterogeneous populations poorly.
Continuous complete remission measurements:

<table>
<thead>
<tr>
<th></th>
<th>Positive samples</th>
<th>Negative samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPM1c+</td>
<td>28</td>
<td>76</td>
</tr>
<tr>
<td>PML-RARA</td>
<td>2</td>
<td>1865</td>
</tr>
<tr>
<td>CBFB-MYH11</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>RUNX1-RUNXI T1</td>
<td>7</td>
<td>87</td>
</tr>
</tbody>
</table>

Figure 1
Figure 2
Figure 3

A: NPM1c+
BM:
FLT3-ITD−: 7 (6-13)
FLT3-ITD+: 7 (4-10)
PB: 2.5 (1-4)

B: PML-RARA
BM: 10.5 (6-12)
PB: 8 (4-8)

C: CBFB-MYH11
BM: 8 (5-9)
PB: 5 (3-6)

D: RUNX1-RUNX1T1
BM: 6 (4-11)
PB: 5 (3-8)
Figure 4
Strikingly different molecular relapse kinetics in NPM1c, PML-RARA, RUNX1-RUNX1T1 and CBFB-MYH11 acute myeloid leukemias

Hans Beier Ommen, Susanne Schnittger, Jelena V. Jovanovic, Ingrid Beier Ommen, Henrik Hasle, Mette Ostergaard, David Grimwade and Peter Hokland