Leukemic stem cell frequency in addition to minimal residual disease is an important biomarker to predict outcome in acute myeloid leukemia: prospective data from the HOVON/SAKK 102 study

Wendelien Zeijlemaker¹, Rosa Meijer², Angèle Kelder¹,
Jannemieke C Carbaat-Ham¹, Yvonne JM Oussoren-Brockhoff³,
Alexander N Snel¹, Dennis Veldhuizen¹, Willemijn J Scholten¹,
Johan Maertens³, Dimitri A Breems⁴, Thomas Pabst⁵, Markus G Manz⁶,
Vincent HJ van der Velden⁷, Jennichjen Slomp⁸, Frank Preijers⁹,
Peter JM Valk¹⁰, Jacqueline Cloos¹¹,¹¹, Arjan A van de Loosdrecht¹,
Bob Löwenberg¹⁰, Gert J Ossenkoppele¹, Gerrit J Schuurhuis¹

¹ Department of Hematology, VU University Medical Center, Cancer Center Amsterdam, Amsterdam, The Netherlands.
² Clinical trial Center- HOVON data center, Erasmus University Medical Center, Rotterdam, The Netherlands.
³ Department of Hematology, University Hospitals Leuven, Campus Gasthuisberg. Leuven, Belgium.
⁴ Department of Hematology, Ziekenhuis Netwerk Antwerpen, Antwerp, Belgium.
⁵ Department of Hematology, Inselspital, Bern University Hospital, Bern, Switzerland.
⁶ Department of Hematology, University Hospital Zürich, Zürich, Switzerland.
⁷ Department of immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.
⁸ Department of Clinical Chemistry, Medisch Spectrum Twente/Medlon, Enschede, The Netherlands.
⁹ Department of Laboratory Medicine – Laboratory for Hematology, Radboud University Nijmegen Medical Center, RUNMC, Nijmegen, The Netherlands.
¹⁰ Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands.
¹¹ Department of Pediatric Oncology/Hematology, VU University Medical Center, Amsterdam, The Netherlands.

Manuscript submitted
Abstract

Current baseline and post-remission risk algorithms imperfectly predict patient outcome in acute myeloid leukemia (AML). Here we introduce and validate an approach of leukemic stem cell (LSC) assessment for relapse prediction and survival.

LSC (CD34+CD38-) frequencies were measured using flow cytometry in an add-on study of the HOVON102/SAKK AML trial. Predefined cut-off levels were prospectively evaluated to assess LSC levels at diagnosis (n=594), and, additionally, to identify LSC\textsuperscript{low}/LSC\textsuperscript{high} (n=302) and MRD\textsuperscript{low}/MRD\textsuperscript{high} patients (n=305) in morphological complete remission. In 242 complete remission patients combined MRD and LSC results were available.

At diagnosis the LSC frequency independently predicts overall survival (OS). After achieving complete remission both LSC\textsuperscript{high} and MRD\textsuperscript{high} patients had worse outcome than LSC\textsuperscript{low} and MRD\textsuperscript{low} patients (hazard ratios [HR] 1.87 and 2.50 for cumulative incidence of relapse [CIR] and 1.62 and 1.49 for OS). Combining LSC and MRD showed substantially reduced survival in MRD\textsuperscript{high}/LSC\textsuperscript{high} patients (HR: OS 3.62, CIR 5.89) compared to MRD\textsuperscript{low}/LSC\textsuperscript{high}, MRD\textsuperscript{high}/LSC\textsuperscript{low}, and especially MRD\textsuperscript{low}/LSC\textsuperscript{low} patients.

This is the first study demonstrating that LSC strongly improves prognostic impact of MRD detection, identifying a patient subgroup with an almost 100% treatment failure probability, warranting LSC measurement incorporation in future AML risk schemes.
Introduction

Proper risk assessment to adapt consolidation treatment strategies for acute myeloid leukemia (AML), either chemotherapy, autologous- or allogeneic stem cell transplantation, is of utmost importance. However, with the current risk-adapted strategies, still a considerable number of patients in the good and intermediate risk groups, relapse. Therefore, further improvement of risk group definition is indispensable. Therapy resistance results in survival of leukemic cells, which are detected during therapy within the bone marrow (BM) derived white blood cells (WBCs). This so-called measurable/minimal residual disease (MRD) can be defined by either flow cytometry or molecular assays, and has been shown in many studies to have clear prognostic impact\(^1\), also in a prospective setting.\(^2\)-\(^6\) Currently, in several on-going prospective clinical studies, therapy is adapted based on MRD evaluations. Despite the strong prognostic value, still part of MRD-negative patients relapse.\(^2\)-\(^6\) There is growing evidence that small subpopulations of malignant cells are more therapy resistant than the bulk of leukemia cells and are at the basis of leukemic outgrowth to relapse.\(^7\),\(^8\) Although these leukemia stem cell like populations (further referred to as LSC) may have different immunophenotypes, a CD34+CD38-/dim LSC population is of particular interest because of its independent prognostic impact.\(^9\)-\(^11\)

In a large international multicentre AML clinical trial (HOVON/SAKK 102), we prospectively validate the prognostic relevance of the CD34+CD38- LSC frequency, both at time of diagnosis and after induction therapy. As we have previously done for flow cytometric MRD\(^2\),\(^12\), we prospectively validated predefined LSC threshold levels\(^9\) for predictability. This novel approach is evaluated here for the first time in a prospective clinical trial setting in a large group of 242 newly diagnosed AML patients in morphological complete remission (CR) and shows strong prognostic value as a single parameter, as well as a marked added value compared to simultaneously measured MRD as regards the ability to predict treatment outcome.

Materials and Methods

Patients

In total 890 patients with AML, age between 18-66, were included in the HOVON/SAKK 102 trial between February 2010 and September 2013. Standard treatment consisted of two cycles of induction chemotherapy followed by either a third cycle, autologous- or allogeneic stem cell transplantation, dependent on pre-treatment risk assessment. Half of the patients received standard treatment and half standard treatment with the addition of clofarabine. For detailed information regarding treatment protocols we refer to the supplementary file and the recently published paper of the H102 clinical trial.\(^13\) A consort flow diagram of all 890 patients is shown in Figure 1. According to the study protocol, well known risk parameters, like t(8;21), inv16/t(16;16), monosomal karyotype, CEBPA, FLT3/NPM1 status, EVI1 over-expression, WBC count, and response to chemotherapy treatment, were used to distinguish patients with a good, intermediate, poor or very poor prognosis (Table S1). Patient characteristics, segregated for patients with and without an MRD/LSC sample available, are outlined in Table 1.
MRD assessment

At diagnosis peripheral blood (PB) was used when BM was not available to define leukemia associated immunophenotypes (LAIPs). At follow-up, only BM was used. MRD assessment was performed as described in detail before. More information regarding the used antibodies and instruments is outlined in the Supporting data. As shown in Figure 1, 305 samples were available for MRD analyses after the 2nd chemotherapy cycle. The established and already prospectively validated cut-off of 0.1% was used to define MRD-negative and MRD-positive patients.2,12

CD34/LSC status

CD34 status was defined at diagnosis as described by Zeijlemaker et al.14 CD34-negative samples are characterized by the absence of neoplastic CD34+ cells, implying that all leukemic cells are in the CD34- compartment. Consequently, no CD34+CD38- LSCs are present in CD34-negative samples. For simplicity reasons these CD34-negative CD34+CD38-LSC=0 AML patients will be further referred to as CD34-negative. In CD34-positive cases CD34+CD38- LSCs are present, and a retrospectively defined cut-off of 0.03% to discriminate LSClow and LSChigh cases was used.9 The antibody panels used at diagnosis are shown in Table S2. Depending on the aberrant immunophenotype(s) as found at AML diagnosis and available BM material, one or more of the diagnosis tubes was used for follow-up LSC determination. For follow-up LSC determination a cut-off of 0.00000% (% of WBC; 0% is no events measured) was used. Detailed information regarding this used cut-off is outlined in the Supporting file, Figures S1, S2 and Table S3.

In total 890 patients were included, of which in total 594 patients were available for combined CD34/LSC analyses at time of diagnosis. After the second cycle there were 305 and 302 samples available for MRD and LSC analyses, respectively. In total there were 242 bone marrow samples in which both MRD and LSC data were available. The different reasons for drop-off are shown in the boxes in the diagram. ‘Unfit material’ refers to a poor quality BM aspirate (e.g. too few blood cells available) hindering reliable measurements. ‘No right markers’ refers to samples in which not all (backbone) markers were measured enabling proper CD34+CD38- LSC detection (e.g. lack of CD34 or CD38). CR, complete remission; LSC, leukemic stem cell; LAIP, leukemia associated immunophenotype; MRD, minimal/measurable residual disease.
FIGURE 1 | Consort flow chart for sample availability of patients included in the HOVON/SAKK 102 study.
<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Patient characteristics of CR patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients with MRD &amp;LSC results</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>242</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>119</td>
</tr>
<tr>
<td>Female</td>
<td>123</td>
</tr>
<tr>
<td><strong>Age, median</strong></td>
<td>52</td>
</tr>
<tr>
<td>Range</td>
<td>19-66</td>
</tr>
<tr>
<td><strong>WBC at diagnosis x10⁹/L</strong></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>8.4</td>
</tr>
<tr>
<td>Range</td>
<td>0.6-229</td>
</tr>
<tr>
<td><strong>AML type</strong></td>
<td></td>
</tr>
<tr>
<td>De Novo</td>
<td>202</td>
</tr>
<tr>
<td>Secondary</td>
<td>24</td>
</tr>
<tr>
<td>MDS</td>
<td>16</td>
</tr>
<tr>
<td><strong>CD34 status</strong></td>
<td></td>
</tr>
<tr>
<td>CD34 negative</td>
<td>35</td>
</tr>
<tr>
<td>CD34 positive LSC ≤0.03%</td>
<td></td>
</tr>
<tr>
<td>CD34 positive LSC &gt;0.03%</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>9</td>
</tr>
<tr>
<td><strong>MRD / LSC status after 2nd course</strong></td>
<td></td>
</tr>
<tr>
<td>MRD neg LSC neg</td>
<td>136</td>
</tr>
<tr>
<td>MRD neg LSC pos</td>
<td>58</td>
</tr>
<tr>
<td>MRD pos LSC neg</td>
<td>28</td>
</tr>
<tr>
<td>MRD pos LSC pos</td>
<td>20</td>
</tr>
<tr>
<td><strong>Karyotype classification</strong></td>
<td></td>
</tr>
<tr>
<td>Core binding factor</td>
<td>36</td>
</tr>
<tr>
<td>Normal karyotype</td>
<td>119</td>
</tr>
<tr>
<td>Monosomal karyotype</td>
<td>20</td>
</tr>
<tr>
<td>Cytogenetic rest abnormalities</td>
<td>59</td>
</tr>
<tr>
<td>Missing</td>
<td>8</td>
</tr>
<tr>
<td><strong>Mutated CEBPα</strong></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>171</td>
</tr>
<tr>
<td>Pos</td>
<td>7</td>
</tr>
<tr>
<td>Missing</td>
<td>64</td>
</tr>
<tr>
<td><strong>FLT3/NPM1 status</strong></td>
<td></td>
</tr>
<tr>
<td>FLT3mt / NPM1mt</td>
<td>122</td>
</tr>
<tr>
<td>FLT3mt / NPM1mut</td>
<td>45</td>
</tr>
<tr>
<td>FLT3mut / NPM1mt</td>
<td>26</td>
</tr>
<tr>
<td>FLT3mut / NPM1mut</td>
<td>26</td>
</tr>
<tr>
<td>Missing</td>
<td>23</td>
</tr>
<tr>
<td><strong>EVI1</strong></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>171</td>
</tr>
<tr>
<td>Pos</td>
<td>19</td>
</tr>
<tr>
<td>Missing</td>
<td>52</td>
</tr>
<tr>
<td><strong>IDH</strong></td>
<td></td>
</tr>
<tr>
<td>IDH1mut / IDH2mut</td>
<td>154</td>
</tr>
<tr>
<td>IDH1mt / IDH2mt</td>
<td>18</td>
</tr>
<tr>
<td>IDH1mut / IDH2mut</td>
<td>27</td>
</tr>
<tr>
<td>IDH1mt / IDH2mt</td>
<td>1</td>
</tr>
<tr>
<td>Missing</td>
<td>42</td>
</tr>
<tr>
<td><strong>DNMT3a</strong></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>139</td>
</tr>
<tr>
<td>Pos</td>
<td>61</td>
</tr>
<tr>
<td>Missing</td>
<td>42</td>
</tr>
<tr>
<td>Patients with MRD &amp; LSC results</td>
<td>%</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---</td>
</tr>
<tr>
<td><strong>First consolidation therapy</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>31</td>
</tr>
<tr>
<td>3rd cycle</td>
<td>90</td>
</tr>
<tr>
<td>Autologous SCT</td>
<td>32</td>
</tr>
<tr>
<td>Allogeneic SCT</td>
<td>89</td>
</tr>
<tr>
<td><strong>CR reached</strong></td>
<td></td>
</tr>
<tr>
<td>After cycle 1</td>
<td>207</td>
</tr>
<tr>
<td>After cycle 2</td>
<td>35</td>
</tr>
<tr>
<td><strong>HOVON risk group</strong></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>80</td>
</tr>
<tr>
<td>Intermediate</td>
<td>60</td>
</tr>
<tr>
<td>Poor</td>
<td>67</td>
</tr>
<tr>
<td>Very poor</td>
<td>35</td>
</tr>
<tr>
<td><strong>Treatment group</strong></td>
<td></td>
</tr>
<tr>
<td>Standard treatment</td>
<td>129</td>
</tr>
<tr>
<td>Clofarabine 10mg</td>
<td>106</td>
</tr>
<tr>
<td>Clofarabine 15mg</td>
<td>7</td>
</tr>
</tbody>
</table>

Patient characteristics of the 242 patients where both MRD and LSC results were present after 2 cycles of chemotherapy. Comparison with the 444 eligible patients who also received a second cycle of chemotherapy treatment and achieved CR (right column). Results show that the patients with and without MRD/LSC results are largely the same. Significant changes between patients with and without MRD/LSC sample are probably due to small patients numbers. Core-binding factor AML defined as translocation[8;21]/inv[16] or t[16;16]; CR, complete remission; NA, not applicable; SCT, stem cell transplantation. WBC, white blood cell. Significant changes with Chi-square test: p=0.029 (karyotype), p=0.010 (first consolidation therapy), and p=0.044 (CR reached). Borderline significance with Chi-square test: p=0.073 (sex), p=0.092 (AML type), and with Kruskal-Wallis test: p=0.091 (WBC). All other differences were not significant.

### Statistical analysis

Survival analyses were performed for data obtained at time of diagnosis and data acquired after the second course of induction chemotherapy. For diagnosis analyses, overall survival (OS) was used, defined as time from date of diagnosis until date of death. For follow-up MRD/LSC analyses, primary end points were OS and cumulative incidence of relapse (CIR). Event free survival (EFS; measured from sample date after achieving CR) was a secondary end point. For EFS both relapse and death were defined as an event. CIR was calculated using the Fine and Gray model from sample date until date of relapse. Death was hereby included as a competing event. Patients without an event were censored at date of last follow-up. For OS and EFS analyses, Kaplan-Meier curves were generated and outcome between groups was compared using log-rank tests. Cox-regression multivariate analyses were performed to calculate hazard ratios (HR) and associated 95% confidence interval (CI) for OS and EFS. Multivariate analyses for CIR were calculated using the Fine and Gray model. All variables significant in univariate analyses and all variables with known clinical importance in AML were used in these analyses, included age, AML type (de novo/secondary/high risk MDS), risk group, FLT3/NPM1 status, CD34 status at diagnosis, WBC count at diagnosis, type of consolidation therapy and H102 treatment arm. More information regarding multivariate analyses is available in the supporting information. Median follow-up time of all eligible surviving patients was 41 months.
Results

CD34/LSC status at diagnosis

Baseline characteristics
Baseline characteristics of the in total 594 patients with a sample available at time of diagnosis are shown in Table S4. Of the 594 samples at diagnosis there were 77 (13%) CD34-negative cases, 338 (57%) CD34-positive cases with low CD34+CD38- LSC levels (≤0.03%), further referred to as LSC\text{low}, and 179 (30%) CD34-positive cases with high CD34+CD38- LSC levels (>0.03%), further referred to as LSC\text{high}. Median LSC percentage of all 517 CD34-positive cases (CD34-negative samples with CD34+CD38-LSC=0 excluded) was 0.0079% (range 0.0000%-19.8761%), i.e. 79 in a million cells.

Validation of prognostic relevance
To establish the value of LSCs in prognostics, first the role of LSC at diagnosis was established. In univariate analysis, a significant decrease in OS with increasing LSC frequencies was apparent (Figure 2B). Multivariate analysis revealed LSC frequency as an independent prognostic factor for OS: HR in LSC\text{low} was 1.94 (95%CI 1.11-3.39) and in LSC\text{high} 2.38 (95%CI 1.34-4.25) relative to the CD34-negative group (Table S5). The relation between CD34+CD38- status and achievement of hematological CR was also investigated: in the CD34-negative cohort (LSC=0), only 4% (3/77) never achieved CR as compared to 12% (41/338) of the LSC\text{low} group and 18% (33/179) of LSC\text{high} group (p<0.001, Figure 2A). In multivariate analyses, odds ratios for achieving CR were 0.65 for LSC\text{low} and 0.45 for LSC\text{high}. These results were not significant (Table S5), probably due to associations with molecular/cytogenetic parameters as outlined below.

At diagnosis associations exist with molecular aberrancies as shown in Table S6. As expected\cite{16,17}, associations were found between CD34-negative AML (CD34+CD38- LSC=0) and the good prognostic FLT3\text{wt}/NPM1\text{mut} profile (Table S6A, left part). Moreover, the percentage of FLT3\text{mut}/NPM1\text{wt} patients was higher in LSC\text{low} than in LSC\text{high} patients (Table 6A, right part). In addition, the percentage of the poor prognostic FLT3\text{mut}/NPM1\text{wt} increased with increasing LSC frequency (LSC\text{high}> LSC\text{low}>CD34 negative). Table S6B shows the association of LSC frequency with the poor prognostic presence of EVI1 over-expression: LSC\text{high} > LSC\text{low} >CD34 negative.

Moreover double mutations of CEBPA are known to be associated with strong CD34 expression.\cite{18} We indeed found this good prognostic marker exclusively in CD34-positive patients, however especially in the LSC\text{low} patients groups with relatively good prognosis (Table S6C).

LSC assessment in CR predicts relapse and survival
The LSC frequency was measured in 302 patients in CR after the second cycle of induction therapy. In total 204 (68%) LSC-negative and 98 (32%) LSC-positive patient samples were included. LSC-negative patients (n=204) had better 3-year CIR and 3-yr OS compared to LSC-positive patients (n=98, Figure S3). Moreover, in multivariate analyses LSC status in CR patients was an independent predictor for both CIR (HR 1.87, 95%CI 1.16-3.01) and OS (HR 1.62, 95%CI 1.05-2.51).
MRD assessment in CR predicts relapse and survival

In 305 CR patients MRD was measured after the 2nd chemotherapy cycle. A difference in outcome was apparent between MRD-negative (n=239) and MRD-positive (n=66) patients (Figure S4). In multivariate analysis, MRD was an independent predictor for CIR (HR 2.49, 95%CI 1.55-4.02), with a trend for OS (HR 1.70, 95%CI 0.94-2.35).

![Graph showing MRD assessment in CR predicts relapse and survival](image)

**FIGURE 2. Effects CD34/LSC status at diagnosis on complete remission rates and overall survival**

Figure 2A illustrates the difference between remission status in the total group of 594 patients of which 77 CD34-negative cases, 338 CD34-positive cases with low LSC levels and 179 CD34-positive cases with high LSC levels. Patient numbers are indicated below the figure. Clearly, with increasing CD34+CD38- LSC frequencies (0, LSClow, LSChigh) the percentage of patients not entering CR, or entering CR at a later stage, increases. Figure 2B shows the impact of CD34/LSC status on overall survival, whereby CD34-negative
patients have an improved overall survival as compared to CD34-positive LSC\textsuperscript{low} and CD34-positive LSC\textsuperscript{high} patients, while, in turn, CD34-positive LSC\textsuperscript{low} have a better survival than CD34-positive LSC\textsuperscript{high} patients. 3-Yrs OS and median OS for the three groups are indicated in the text at the bottom of the figure (upper part).

**Evaluating combined MRD/LSC assessments in CR for relapse and survival**

In 242 patients both the LSC frequency and MRD frequency after chemotherapy treatment were known. When combining LSC and MRD (Figure 3), four groups were distinguished: 1. MRD\textsuperscript{neg}/LSC\textsuperscript{neg} (n=136; 56%); 2. MRD\textsuperscript{pos}/LSC\textsuperscript{neg} (n=28; 12%); 3. MRD\textsuperscript{neg}/LSC\textsuperscript{pos} (n=58; 24%); 4. MRD\textsuperscript{pos}/LSC\textsuperscript{pos} (n=20; 8%). CIR and OS analyses demonstrated that patients negative for both MRD and LSC have the best prognosis: 3-year CIR was 35% for MRD\textsuperscript{neg}/LSC\textsuperscript{neg}, 43% for MRD\textsuperscript{pos}/LSC\textsuperscript{neg}, 53% for MRD\textsuperscript{neg}/LSC\textsuperscript{pos} and 100% for MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients. Similar results were found for OS with 3-year OS of 66% for MRD\textsuperscript{neg}/LSC\textsuperscript{neg}, 68% for MRD\textsuperscript{pos}/LSC\textsuperscript{neg}, 53% for MRD\textsuperscript{neg}/LSC\textsuperscript{pos} and 0% for MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients. These results show that double positivity is associated with an excessively poor outcome. Multivariate analyses showed that the combined MRD\textsuperscript{pos}/LSC\textsuperscript{pos} status has profound predictive significance for both OS (HR 3.62), CIR (HR 5.89) and EFS (HR 4.24, Table S7).

**MRD/LSC results in distinct risk categories**

Table S8 shows that MRD\textsuperscript{neg}/LSC\textsuperscript{neg} patients are present in all four different risk groups, although the frequency of MRD\textsuperscript{neg}/LSC\textsuperscript{neg} patients is lower in the very poor risk group (47%) as compared to the good risk group (68%). Moreover, the frequency of MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients was highest in the very poor risk group (15%), although MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients were present as well in the good risk group (4%). These data illustrate that MRD/LSC status at follow-up is an important property across all risk groups.

Clinical outcome was subsequently assessed for the different MRD/LSC subgroups within the different H102-defined risk groups. Figure 4 shows MRD/LSC results for CIR (Figure 4.I.A-D) and for OS (Figure 4.II.A-D). Although patient numbers for some subgroups are small, these results show that MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients have a (very) poor prognosis, even when present in the good or intermediate risk group. In contrast, the double negative population still has a relatively good prognosis even in the poor risk group. Overall, these results illustrate that the LSC frequency has important clinical relevance additional to MRD and currently used risk factors in predicting outcome in AML.
FIGURE 3 | Prognostic value of MRD/LSC status as defined at follow-up
Figure 3A shows cumulative incidence of relapse (CIR) for the four different MRD/LSC patient groups. This figures shows the important difference in both CIR (3A) and OS (3B) for the different MRD/LSC patient groups. Prognosis becomes better in the sequence MRD<sup>neg</sup>/LSC<sup>neg</sup> and MRD<sup>pos</sup>/LSC<sup>neg</sup> > MRD<sup>neg</sup>/LSC<sup>pos</sup> > MRD<sup>pos</sup>/LSC<sup>pos</sup>. At the bottom CIR, 3-years OS and median OS, are summarized for the different groups.
FIGURE 4 | Prognostic value of MRD/LSC status in different risk categories

CIR and OS curves for MRD/LSC status in 80 good risk patients (I.A, II.A), 60 intermediate risk patients (I.B, II.B), 67 poor risk patients (I.C, II.C) and 35 very poor risk patients (I.D, II.D). I.A shows that CIR after 3-years in the good risk patient group was 25% (SE 6) for MRD<sup>neg</sup> + LSC<sup>neg</sup> (n=54), 50% (SE 16) for MRD<sup>pos</sup> + LSC<sup>neg</sup> (n=10), 32% (SE 13) for MRD<sup>neg</sup> + LSC<sup>pos</sup> (n=13), and 100% for MRD<sup>pos</sup> + LSC<sup>pos</sup> patients (n=3). In the intermediate risk group (I.B) this was 38% (SE 10), 40% (SE 15), 50% (SE 14), and 83% (SE 15), respectively. In the poor risk group (I.C) this was 35% (SE 8), 40% (SE 22), 46% (SE 11), and 83% (SE 15), respectively. In the very poor risk group (I.D) patient numbers were too small for reliable results (I.D). Same sequences were found for OS (II.A): in the good risk group (II.A) 3-year OS was 75% (SE 7), 80% (SE 13), 62% (SE 26), and 33% (SE 27). In the intermediate risk group (II.B) this was 81% (SE 7), 60% (SE 16), 50% (SE 14), and 17% (SE 15). In the poor risk patients (II.C) this was 64% (SE 8), 60% (SE 22), 62% (SE 11), and 17% (SE 15). In the very poor risk group (II.D) patient numbers were too small for reliable results.


**Discussion**

MRD as defined by flow cytometry or molecular assays has been proven to be a reproducible marker for relapse prediction.\cite{2,6,19,20} However, still a considerable proportion of MRD-negative patients will relapse and it has been hypothesized that differences in LSC frequencies may be partly responsible.\cite{9,10,21} Support for this notion comes from the prognostic value of LSC frequencies in preliminary retrospective studies.\cite{9,10,21} In the present study for the first time we show, in a prospective large cohort of AML patients, that LSC frequency has prognostic value additive to well-known AML risk factors, including MRD. Already at time of diagnosis CD34+CD38- LSC frequencies were correlated with OS in multivariate analyses. The correlation with good (FLT3\textsuperscript{int}/NPM1\textsuperscript{wt}) and poor (FLT3\textsuperscript{int}/NPM1\textsuperscript{wt}, EVI1 over-expression) molecular aberrancies suggests that these and probably other genetic aberrations may translate into differences in frequencies of CD34+CD38- LSC. These results offer an explanation for the absence of significance of correlation between LSC frequency and CR achievement in multivariate analysis, but on the other hand indicate that LSC may represent a common factor for these and other prognostic factors not only at diagnosis, but also at follow up. Moreover, these CD34+CD38- LSCs, which are known to be therapy resistant and immune-evasive\cite{22,23}, and which, during leukemic growth, may outcompete other stem cell immune-phenotypes.\cite{9}

We show that in all different AML risk categories, MRD- and LSC double positivity, as defined after achievement of morphologic complete remission, predicts a very poor outcome in AML patients. Even in the good risk category, we were able to show that MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients have a high risk of relapse and poor OS (Figure 4.IA and 4.IIA). Our data indicate that MRD\textsuperscript{pos}/LSC\textsuperscript{pos} patients, independent on risk category, should be considered as poor/very poor risk patients, preferably to be treated as such.

To further improve this flow cytometric assay, especially to better discriminate risk of relapse between MRD\textsuperscript{neg}/LSC\textsuperscript{neg}, MRD\textsuperscript{pos}/LSC\textsuperscript{neg} and MRD\textsuperscript{neg}/LSC\textsuperscript{pos} patients, improvements concerning sensitivity/specificity of the MRD and LSC assay, can be expected when using: 1) a smart combination of flow cytometric MRD and molecular MRD (manuscript in preparation), 2) the combination of MRD/LSC assessment after different courses of therapy\cite{2}, and 3) the detection of upcoming MRD and LSC populations. It is known that at AML diagnosis different AML clones/mutations/cell populations are present and that this may lead to a change in constitution of the disease in follow up and at relapse.\cite{24-28} In the present study mainly aberrancies defined at diagnosis were used for both MRD and LSC assessment; including such upcoming populations could reduce frequencies of false negative MRD/LSC results. Lastly, the sensitivity of the LSC assay can be improved by increasing the numbers of WBC analysed. With 5-10 million WBC acquired, for example in “a one-tube-for-both-MRD-and-LSC” approach currently under development, more patients with high LSC frequencies prone to relapse can be distinguished. Such a “one-tube” approach has already been developed for LSC determination.\cite{29}

Overall, we show that the LSC frequency at baseline and after chemotherapy is an independent prognostic factor and, combined with MRD, enables to identify (very) poor risk patients in all different currently used risk categories. Our data warrants including both MRD and LSC in future AML risk classifications.
Acknowledgements

We thank all participating study centers.

References

Supporting information

Patients and treatment

In total 890 acute myeloid leukemia (AML) patients were included in the HOVON 102 multicenter clinical trials ([HOVON/SAKK AML] Dutch-Belgian Hemato-Oncology Cooperative Group/Swiss Group for Clinical Cancer Research Acute Myeloid Leukemia). AML was diagnosed according to the World Health Organization criteria, whereby also high-risk myelodysplastic syndromes (MDS) patients (IPSS ≥ 1.5) were included. Secondary leukemias were also included and were defined as AMLs with prior treatment with chemotherapy and/or radiotherapy, or cases with a pre-existing hematological disease (e.g. MDS or myelofibrosis) who developed AML. All patients gave written informed consent in accordance with the Declaration of Helsinki (central study approval number 2009-293; VUmc local approval number 2010- 56 [LUV]). On the HOVON website (www.hovan.nl) detailed information regarding inclusion/exclusion criteria is available. In general all patients received two cycles of standard induction chemotherapy, whereby the first induction cycle consisted of idarubicin (12 mg/m², days 1-3) and cytarabine (200mg/m², days 1-7), and the second induction cycle contained amsacrine (120 mg/m², days 4-6) and cytarabine (1000mg/m², days 1-6). Half of the patients included in the HOVON102 study were randomized to the experimental treatment arm, in which clofarabine treatment (assigned dose, days 1-5) was added to above described standard chemotherapy treatment regiments. After induction chemotherapy treatment most patients received consolidation therapy, consisting of either a third chemotherapy cycle, autologous or allogeneic stem cell transplantation. In case a third chemotherapy was the consolidation therapy of choice, this consisted of mitoxantrone (10mg/m², days 1-5) and etoposide (100mg/m², days 1-5).

MRD assessment

Measurements were performed using a FACS CANTO flow cytometer (BD Biosciences; San Jose, CA). Infinicyt™ software (Cytognos, Spain) was used for the immunophenotypic analyses. For diagnosis determination of leukemia associated immunophenotype (LAIP), 20 million cells were used, while for MRD determination in follow-up 2x10⁶ WBCs were labelled with a minimum of 500,000 WBCs acquired for immunophenotypic analyses. At diagnosis, LAIPs were determined as described in detail before.²

MRD assessment was performed locally by four other centers, qualified to perform this³: University Medical Center Rotterdam (n=20), University Hospital Leuven (n=6), Medisch Spectrum Twente (n=6) and Radboud University Nijmegen Medical Center (n=1). All other MRD measurements were performed in VU University Medical Center (n=272). Final data analysis of all MRD samples was performed in VU University Medical Center. Clinicians were not informed concerning the MRD status, while people responsible for data analysis were not aware of the clinical situation of the patients.
CD34/LSC status

At time of diagnosis bone marrow (BM; n=495) or peripheral blood (PB; n=97) samples were used to perform multiparameter immunophenotyping by flow cytometry (in 2 cases origin of the material was missing). At follow up only BM samples were used. After lysing of the erythrocytes, cells were labelled with the antibodies (Table S2). In November 2011 the 6-color antibody panel was exchanged for the 8-color panel (patients were included between February 2010 and September 2013). Cells were incubated for 15 minutes with the appropriate antibodies in the dark. Subsequently, unlabelled antibodies were removed via washing with phosphate buffered saline/0.1% human serum albumin. Flow cytometric measurements were performed with a FACSCanto-II from BD (New Jersey, USA). All CD34 and LSC measurements at diagnosis and LSC measurements during follow-up were performed in VU University Medical Center.

Methods to define CD34 status at time of diagnosis are described by Zeijlemaker et al. and LSC status at diagnosis and follow-up by Terwijn et al. CD34 data of 173 patients, as defined at time of diagnosis, have been previously published. In our previous study, a cut-off for LSC positivity and negativity of 0.0001% (% of WBC) was defined. In that study this cut-off was chosen since, out of a wide range of cut-off values (Table S7 in Terwijn et al.), it gave the best discrimination (highest p-value) in outcome between LSC-negative and positive patients. However, in that paper, only CD34-positive patients were studied. In the present study we included all patients, both CD34-positive- and CD34-negative, the latter CD34-negative group making up 13% of the total patient population. Due to the inclusion of these CD34-negative patients, with its inherent CD34+CD38− LSC frequency of 0%, the previously used cut-off value was likely to decrease. We thus also studied cut-off levels lower than the 0.0001% cut-off, as used in the previous study. The cut-off of 0% turned out to offer the most optimal one (together with 0.4 in a million, i.e. 0.00004%) where still a considerable number of patients were classified as LSC positive (i.e. values above 0%) (Table S3). As a control, we applied this 0% cut-off for the data of our previous study and found similar differences in survival between the two patient groups defined by either the 0% and the previously used 0.0001% cut-off (Figure S1).

In part of the follow-up BM samples used for LSC identification, we were unable to acquire > 1 million cells. Figure S2A shows results for event free survival (EFS) with all patients included, irrespective numbers of WBC acquired (range 174.524-2.758.892 WBC; median 1.402.905). When including only cases with ≥ 1 million WBC, similar results were obtained (Figure S2B). For cases with ≥ 2 million WBCs acquired, results were even better despite the fact that the group was considerably smaller (Figure S2C). Future developments in flow cytometry will allow such since the application of one tube for both MRD and LSC is being developed, which will largely decrease the total number of cells needed for all assays (now 7 tubes with 2 million cells each and in future then 1 tube with e.g. 10 million cells available for all MRD and LSC sub-analyses).
Baseline characteristics

Table S4 shows the baseline characteristics for the 594 patients included in the diagnosis study of LSC frequencies (left part of the table). The right part of the table shows the 264 patients included in the HOVON 102 for which no sample was available. The table shows no major differences in baseline characteristics between AML patients with and without a sample available, except for WBC (7.4 \(10^9\)/L vs 5.8 \(10^9\)/L). AML patients with a sample available more often did not receive a consolidation treatment (34% vs 25%) and less often received an allogeneic transplantation as compared to the patients with no sample available (31% vs 44%).

Statistical analysis

Statistical analyses were performed using both STATA version 14.0 and SPSS version 22.0 software. All variables with clinical significant importance or variables with known clinical importance based on earlier studies were included in the Cox-regression multivariate analyses. The baseline multivariate model (Table S5) risk groups according to European LeukemiaNet (ELN) were included, instead of risk groups according to the HOVON protocols, since the ELN classification is based on diagnosis parameters only, while remission status (ie remission attained or not attained after one induction cycle) is a parameter that is included in the HOVON risk categories. In follow-up multivariate models, risk groups were defined according to the HOVON risk categories since here only patients in CR are included. Consolidation therapy was included as a time-dependent factor in both the diagnosis (Table S5) and the follow-up model (Table S7). No important differences were found between the multivariate models for cumulative incident of relapse using Fine & Gray and using cause specific hazard ratios. A p-value of < 0.05 was considered significant.

References

## Supplementary tables

### TABLE S1 | Risk group profiles according to HOVON/SAKK 102 study

<table>
<thead>
<tr>
<th>Risk</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>t(8;21) or AML1-ETO, WBC≤20</td>
</tr>
<tr>
<td></td>
<td>inv16/t(16;16) or CBFB-MYH11</td>
</tr>
<tr>
<td></td>
<td>MK-, CEBPA+</td>
</tr>
<tr>
<td></td>
<td>MK-, FLT3ITD-/NPM1+, CRE</td>
</tr>
<tr>
<td>Intermediate</td>
<td>t(8;21) or AML1-ETO, WBC&gt;20</td>
</tr>
<tr>
<td></td>
<td>CN-X-Y, WBC≤100, CRE</td>
</tr>
<tr>
<td>Poor</td>
<td>CN-X-Y, WBC≤100, not CRE</td>
</tr>
<tr>
<td></td>
<td>CA, non CBF, MK-, no abn3q26, EV11-</td>
</tr>
<tr>
<td>Very poor</td>
<td>non CBF, MK+</td>
</tr>
<tr>
<td></td>
<td>non CBF, abn3q26</td>
</tr>
<tr>
<td></td>
<td>non CBF, EV11+</td>
</tr>
</tbody>
</table>

CA, cytogenetically abnormal; CBF, core-binding factor leukemia; CN-X-Y, cytogenetically normal or only loss of X or Y chromosome; CRE, early complete remission; MK, monosomal karyotype; WBC, white blood cell.

### TABLE S2 | Antibody panels

#### 6-Color antibody panel

<table>
<thead>
<tr>
<th>Tube</th>
<th>FITC</th>
<th>PE</th>
<th>PerCP</th>
<th>PC7</th>
<th>APC</th>
<th>APC-H7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PBS</td>
<td>PBS</td>
<td>CD45</td>
<td>PBS</td>
<td>CD34</td>
<td>PBS</td>
</tr>
<tr>
<td>2</td>
<td>CD2</td>
<td>CD7</td>
<td>CD45</td>
<td>CD13</td>
<td>CD34</td>
<td>HLADR</td>
</tr>
<tr>
<td>3</td>
<td>CD36</td>
<td>CD133</td>
<td>CD45</td>
<td>CD22</td>
<td>CD34</td>
<td>CD19</td>
</tr>
<tr>
<td>4</td>
<td>CD15</td>
<td>CD33</td>
<td>CD45</td>
<td>CD11b</td>
<td>CD34</td>
<td>CD14</td>
</tr>
<tr>
<td>5</td>
<td>CD13</td>
<td>CD56</td>
<td>CD45</td>
<td>CD117</td>
<td>CD34</td>
<td>HLADR</td>
</tr>
<tr>
<td>6</td>
<td>CD7</td>
<td>TIM-3</td>
<td>CD45</td>
<td>CD38</td>
<td>CD34</td>
<td>HLADR</td>
</tr>
<tr>
<td>7</td>
<td>CD11b</td>
<td>CLL-1</td>
<td>CD45</td>
<td>CD38</td>
<td>CD34</td>
<td>CD19</td>
</tr>
</tbody>
</table>

#### 8-Color antibody panel

<table>
<thead>
<tr>
<th>Tube</th>
<th>FITC</th>
<th>PE</th>
<th>PerCP-CY5.5</th>
<th>PC7</th>
<th>APC</th>
<th>APC-H7</th>
<th>HV450</th>
<th>KO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HLA-DR</td>
<td>CD33</td>
<td>CD13</td>
<td>PBS</td>
<td>CD14</td>
<td>PBS</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>2</td>
<td>CD44</td>
<td>CLL-1/CLEC12a</td>
<td>CD13</td>
<td>CD56</td>
<td>CD38</td>
<td>HLA-DR</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>3</td>
<td>CD7</td>
<td>TIM-3</td>
<td>CD13</td>
<td>CD117</td>
<td>CD38</td>
<td>CD19</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>4</td>
<td>CD2</td>
<td>CD133</td>
<td>CD13</td>
<td>CD117</td>
<td>CD38</td>
<td>CD19</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>5</td>
<td>CD36</td>
<td>CD123</td>
<td>CD13</td>
<td>CD33</td>
<td>CD38</td>
<td>CD14</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>6</td>
<td>CD11b</td>
<td>CD96</td>
<td>CD13</td>
<td>CD117</td>
<td>CD38</td>
<td>CD14</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>7</td>
<td>CD15</td>
<td>CD22</td>
<td>CD13</td>
<td>CD117</td>
<td>CD38</td>
<td>HLA-DR</td>
<td>CD34</td>
<td>CD45</td>
</tr>
</tbody>
</table>

In the 6 color antibody panel often a second labeling, including a label switch of one or more antibodies, was performed to increase sensitivity specificity of a LAIP. Leukemic stem cell frequencies were determined in the 6-color panel using tubes 6 and 7, including CD34, CD45 and CD38 as a backbone. In the 8-color panel stem cell frequencies were determined using tubes 2-7.
TABLE S3 | Range of LSC cut-off values after second course of induction therapy used for overall survival and accompanying p-values calculated via log-rank analyses

<table>
<thead>
<tr>
<th>Cut-off %</th>
<th>0.00000</th>
<th>0.000004</th>
<th>0.000006</th>
<th>0.000008</th>
<th>0.000010</th>
<th>0.000015</th>
<th>0.000020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n) above cutoff</td>
<td>78</td>
<td>78</td>
<td>64</td>
<td>53</td>
<td>48</td>
<td>34</td>
<td>29</td>
</tr>
<tr>
<td>Patients (n) below cutoff</td>
<td>164</td>
<td>164</td>
<td>178</td>
<td>189</td>
<td>194</td>
<td>208</td>
<td>213</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000011</td>
<td>0.000011</td>
<td>0.000023</td>
<td>0.000001</td>
<td>0.000014</td>
<td>0.000083</td>
<td>0.000032</td>
</tr>
</tbody>
</table>

TABLE S4 | Patient characteristics at baseline

<table>
<thead>
<tr>
<th></th>
<th>Patients with CD34 &amp;LSC results</th>
<th>%</th>
<th>Other Patients</th>
<th>%</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>594</td>
<td>264</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>329</td>
<td>55</td>
<td>150</td>
<td>57</td>
<td>0.697*</td>
</tr>
<tr>
<td>Female</td>
<td>265</td>
<td>45</td>
<td>114</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Age, median</td>
<td>54</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>18-66</td>
<td>18-66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBC at diagnosis x10⁹/L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>7.4</td>
<td>5.8</td>
<td></td>
<td></td>
<td>0.032*</td>
</tr>
<tr>
<td>Range</td>
<td>0.4-341</td>
<td>0.2-255</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.341*</td>
</tr>
<tr>
<td>M0</td>
<td>51</td>
<td>9</td>
<td>18</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>120</td>
<td>20</td>
<td>43</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>155</td>
<td>26</td>
<td>67</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>59</td>
<td>10</td>
<td>17</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>78</td>
<td>13</td>
<td>43</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>M6</td>
<td>17</td>
<td>3</td>
<td>13</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>M7</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RAEB(-t)</td>
<td>91</td>
<td>15</td>
<td>50</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Not classified</td>
<td>19</td>
<td>3</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>AML type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.133*</td>
</tr>
<tr>
<td>De Novo</td>
<td>476</td>
<td>80</td>
<td>196</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>62</td>
<td>62</td>
<td>33</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>MDS</td>
<td>56</td>
<td>56</td>
<td>33</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>CD34 status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34 negative</td>
<td>77</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34 positive LSC &lt; 0.03%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34 positive LSC &gt; 0.03%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karyotype classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.087*</td>
</tr>
<tr>
<td>Core binding factor</td>
<td>62</td>
<td>10</td>
<td>17</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Normal karyotype</td>
<td>297</td>
<td>50</td>
<td>128</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Monosomal karyotype</td>
<td>80</td>
<td>13</td>
<td>31</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Cytogenetic rest abnormalities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>17</td>
<td>3</td>
<td>10</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Mutated CEBPα</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.056*</td>
</tr>
<tr>
<td>Neg</td>
<td>397</td>
<td>67</td>
<td>192</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Pos</td>
<td>14</td>
<td>2</td>
<td>14</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>183</td>
<td>31</td>
<td>58</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Patients with CD34 &amp;LSC results</td>
<td>%</td>
<td>Other Patients</td>
<td>%</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---</td>
<td>----------------</td>
<td>---</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td><strong>FLT3/NPM1 status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLT3	extsuperscript{mt} / NPM1	extsuperscript{mt}</td>
<td>315</td>
<td>53</td>
<td>150</td>
<td>57</td>
<td>0.817*</td>
</tr>
</tbody>
</table>
| FLT3	extsuperscript{mt} / NPM1
  mut | 90 | 90 | 40 | 15 |       |
| FLT3	extsuperscript{md} / NPM1	extsuperscript{mt} | 54 | 9 | 20 | 8 |       |
| FLT3	extsuperscript{md} / NPM1
  mut | 59 | 10 | 25 | 9 |       |
| Missing                       | 76 | 13 | 29 | 11 |       |
| **EVI1**                      |   |                |   |         |
| Neg                           | 420 | 71 | 191 | 72 | 0.298* |
| Pos                           | 53 | 9 | 31 | 12 |       |
| Missing                       | 121 | 20 | 42 | 16 |       |
| **IDH**                       |   |                |   |         |
| IDH1 neg / IDH2 neg           | 386 | 65 | 126 | 48 | 0.170* |
| IDH1 pos / IDH2 neg           | 55 | 9 | 9 | 3 |       |
| IDH1 neg / IDH2 pos           | 60 | 10 | 23 | 9 |       |
| IDH1 pos / IDH2 pos           | 1 | 0 | 0 | 0 |       |
| Missing                       | 92 | 15 | 106 | 40 |       |
| **DNMT3a**                    |   |                |   |         |
| Neg                           | 350 | 59 | 123 | 47 | 0.048* |
| Pos                           | 152 | 26 | 35 | 13 |       |
| Missing                       | 92 | 15 | 106 | 40 |       |
| **First consolidation therapy** |   |                |   |         |
| None                          | 199 | 34 | 67 | 25 | 0.002* |
| 3rd cycle                     | 159 | 27 | 66 | 25 |       |
| Autologous SCT                | 53 | 9 | 16 | 6 |       |
| Allogeneic SCT                | 183 | 31 | 115 | 44 |       |
| **CR reached**                |   |                |   |         |
| Never CR                      | 77 | 13 | 23 | 9 | 0.211* |
| After cycle 1                 | 408 | 69 | 197 | 75 |       |
| After cycle 2                 | 89 | 15 | 38 | 14 |       |
| Later                         | 20 | 3 | 6 | 2 |       |
| **HOVON risk group**          |   |                |   |         |
| Good                          | 148 | 25 | 65 | 25 | 0.606* |
| Intermediate                  | 132 | 22 | 65 | 25 |       |
| Poor                          | 198 | 33 | 77 | 29 |       |
| Very poor                     | 116 | 20 | 57 | 22 |       |
| **Treatment group**           |   |                |   |         |
| Standard treatment            | 295 | 50 | 138 | 52 | 0.765* |
| Clofarabine 10mg              | 276 | 46 | 117 | 44 |       |
| Clofarabine 15mg              | 23 | 4 | 9 | 3 |       |

Patient characteristics of the 594 patients where both CD34 and LSC results were present at time of diagnosis. Comparison with the 264 eligible patients where no CD34 and LSC results were available (right column) show that baseline characteristics are largely the same. Core-binding factor AML defined as translocation[8;21]/inv[16] or t[16;16]; CR, complete remission; FAB, French-American-British classification; NA, not applicable. RAEB, refractory anaemia with excess blasts; RAEB-t, RAEB in transformation; SCT, stem cell transplantation. WBC, white blood cell. *Chi-square test; *Fisher’s exact test; *Kruskal-Wallis test.
### TABLE S5 | Multivariate analyses for therapy response, overall survival, and event free survival at time of diagnosis

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Multivariate analysis</th>
<th>Multivariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CR OR</td>
<td>95% CI</td>
<td>p</td>
</tr>
<tr>
<td>Total</td>
<td>518</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34 status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34-negative</td>
<td>64</td>
<td>0.65</td>
<td>0.13-3.19</td>
<td>0.591</td>
</tr>
<tr>
<td>CD34-positive LSC&lt;sub&gt;Low&lt;/sub&gt;</td>
<td>294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34-positive LSC&lt;sub&gt;High&lt;/sub&gt;</td>
<td>160</td>
<td>0.45</td>
<td>0.09-2.25</td>
<td>0.329</td>
</tr>
<tr>
<td>Age (continuous)</td>
<td>518</td>
<td>0.94</td>
<td>0.91-0.97</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AML type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Novo</td>
<td>424</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>54</td>
<td>0.67</td>
<td>0.30-1.45</td>
<td>0.307</td>
</tr>
<tr>
<td>MDS</td>
<td>40</td>
<td>2.00</td>
<td>0.64-6.23</td>
<td>0.230</td>
</tr>
<tr>
<td>Risk group ELN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate-I</td>
<td>175</td>
<td>0.48</td>
<td>0.17-1.37</td>
<td>0.171</td>
</tr>
<tr>
<td>Intermediate-II</td>
<td>94</td>
<td>1.39</td>
<td>0.40-4.80</td>
<td>0.605</td>
</tr>
<tr>
<td>Adverse</td>
<td>109</td>
<td>0.33</td>
<td>0.12-0.94</td>
<td>0.037</td>
</tr>
<tr>
<td>FLT3/NPM1 status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLT3&lt;sub&gt;mut&lt;/sub&gt; / NPM1&lt;sub&gt;mut&lt;/sub&gt;</td>
<td>315</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLT3&lt;sub&gt;mut&lt;/sub&gt; / NPM1&lt;sub&gt;mut&lt;/sub&gt;</td>
<td>90</td>
<td>4.78</td>
<td>0.85-26.71</td>
<td>0.075</td>
</tr>
<tr>
<td>FLT3&lt;sub&gt;mut&lt;/sub&gt; / NPM1&lt;sub&gt;mut&lt;/sub&gt;</td>
<td>54</td>
<td>1.41</td>
<td>0.53-3.75</td>
<td>0.494</td>
</tr>
<tr>
<td>FLT3&lt;sub&gt;mut&lt;/sub&gt; / NPM1&lt;sub&gt;mut&lt;/sub&gt;</td>
<td>59</td>
<td>2.94</td>
<td>0.93-9.25</td>
<td>0.066</td>
</tr>
<tr>
<td>WBC at diagnosis (x 10&lt;sup&gt;9&lt;/sup&gt;/L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>351</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-100</td>
<td>130</td>
<td>0.63</td>
<td>0.30-1.31</td>
<td>0.212</td>
</tr>
<tr>
<td>&gt;100</td>
<td>37</td>
<td>0.58</td>
<td>0.17-2.02</td>
<td>0.391</td>
</tr>
<tr>
<td>First consolidation therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>166</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 3</td>
<td>144</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Autologous SCT</td>
<td>48</td>
<td>0.49</td>
<td>0.28-0.85</td>
<td>0.011</td>
</tr>
<tr>
<td>Allogeneic SCT</td>
<td>160</td>
<td>0.47</td>
<td>0.34-0.65</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Treatment arm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>control group</td>
<td>254</td>
<td>1.53</td>
<td>0.87-2.72</td>
<td>0.142</td>
</tr>
<tr>
<td>clofarabine</td>
<td>264</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; CR, complete remission; HR, hazard ratio; NA, not applicable; OR, odds ratio; p, p-value; SCT, stem cell transplantation; WBC, white blood cell count. Consolidation therapy is not included in the analyses for treatment response since consolidation treatment is only given after achieving CR. In OS analyses consolidation therapy is included as a time-dependent variable. *Event defined as relapse or death
### TABLE S6 | Different molecular sub-group frequencies and survival in CD34/LSC defined subgroups

<table>
<thead>
<tr>
<th></th>
<th>CD34-negative&lt;sup&gt;LSC&lt;sup&gt;©&lt;/sup&gt; (n=64)</th>
<th>CD34-positive (n=454)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wt/wt</td>
<td>wt/mut</td>
</tr>
<tr>
<td>FLT3/NPM1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency (%)</td>
<td>20.3</td>
<td>60.9</td>
</tr>
<tr>
<td>Patients(n)</td>
<td>13</td>
<td>39</td>
</tr>
<tr>
<td>A 3-yrs OS (%)</td>
<td>53</td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CD34-negative&lt;sup&gt;LSC&lt;sup&gt;©&lt;/sup&gt; (n=59)</th>
<th>CD34-positive (n=414)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EVI1&lt;sup&gt;overexpression&lt;/sup&gt;    neg</td>
<td>pos</td>
</tr>
<tr>
<td>Frequency (%)</td>
<td>98.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Patients(n)</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>B 3-yrs OS (%)</td>
<td>75</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CD34-negative&lt;sup&gt;LSC&lt;sup&gt;©&lt;/sup&gt; (n=50)</th>
<th>CD34-positive (n=361)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CEBPA&lt;sup&gt;double mut&lt;/sup&gt;               neg</td>
<td>pos</td>
</tr>
<tr>
<td>Frequency (%)</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Patients(n)</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>C 3-yrs OS (%)</td>
<td>70</td>
<td>-</td>
</tr>
</tbody>
</table>

The meaning of the bold text is explained in the manuscript. Abbreviations: OS, overall survival; mut, mutant; neg, negative; NP, not possible; pos, positive; wt, wild-type.
### TABLE S7 | Multivariate analyses for relapse incidence, overall survival, and event free survival for CR patients

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Cumulative relapse incidence</th>
<th>Death incidence</th>
<th>Event incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>HR</td>
<td>95% CI</td>
<td>p</td>
</tr>
<tr>
<td>Total</td>
<td>212</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34 status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD34-negative</td>
<td>30</td>
<td>0.67</td>
<td>0.34-1.32</td>
<td>0.242</td>
</tr>
<tr>
<td>CD34-positive LSC low</td>
<td>120</td>
<td>1.10</td>
<td>0.55-2.17</td>
<td>0.789</td>
</tr>
<tr>
<td>CD34-positive LSC high</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>212</td>
<td>1.01</td>
<td>1.00-1.03</td>
<td>0.138</td>
</tr>
<tr>
<td>AML type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Novo</td>
<td>179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>21</td>
<td>1.10</td>
<td>0.55-2.22</td>
<td>0.783</td>
</tr>
<tr>
<td>MDS</td>
<td>12</td>
<td>0.91</td>
<td>0.41-2.01</td>
<td>0.811</td>
</tr>
<tr>
<td>CR reached</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early CR</td>
<td>180</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late CR</td>
<td>32</td>
<td>1.44</td>
<td>0.77-2.67</td>
<td>0.251</td>
</tr>
<tr>
<td>MRD/LSC status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRD&lt;sup&gt;low&lt;/sup&gt; LSC&lt;sup&gt;low&lt;/sup&gt;</td>
<td>121</td>
<td>2.00</td>
<td>1.04-3.87</td>
<td>0.039</td>
</tr>
<tr>
<td>MRD&lt;sup&gt;low&lt;/sup&gt; LSC&lt;sup&gt;low&lt;/sup&gt;</td>
<td>23</td>
<td>1.61</td>
<td>0.87-2.97</td>
<td>0.127</td>
</tr>
<tr>
<td>MRD&lt;sup&gt;low&lt;/sup&gt; LSC&lt;sup&gt;low&lt;/sup&gt;</td>
<td>51</td>
<td>5.89</td>
<td>3.32-10.47</td>
<td>0.001</td>
</tr>
<tr>
<td>HOVON risk group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>43</td>
<td>2.99</td>
<td>1.39-6.44</td>
<td>0.005</td>
</tr>
<tr>
<td>Poor</td>
<td>60</td>
<td>2.95</td>
<td>1.29-6.74</td>
<td>0.010</td>
</tr>
<tr>
<td>Very poor</td>
<td>34</td>
<td>8.37</td>
<td>3.92-17.83</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FLT3/NPM1 status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLT3&lt;sup&gt;mut&lt;/sup&gt; / NPM1&lt;sup&gt;mut&lt;/sup&gt;</td>
<td>116</td>
<td>0.51</td>
<td>0.23-1.13</td>
<td>0.099</td>
</tr>
<tr>
<td>FLT3&lt;sup&gt;mut&lt;/sup&gt; / NPM1&lt;sup&gt;mut&lt;/sup&gt;</td>
<td>45</td>
<td>1.12</td>
<td>0.53-2.37</td>
<td>0.762</td>
</tr>
<tr>
<td>FLT3&lt;sup&gt;mut&lt;/sup&gt; / NPM1&lt;sup&gt;mut&lt;/sup&gt;</td>
<td>26</td>
<td>0.86</td>
<td>0.37-2.00</td>
<td>0.721</td>
</tr>
<tr>
<td>WBC at diagnosis (x 10&lt;sup&gt;9&lt;/sup&gt;/L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>141</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-100</td>
<td>52</td>
<td>3.07</td>
<td>1.82-5.20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&gt;100</td>
<td>19</td>
<td>2.66</td>
<td>1.32-5.35</td>
<td>0.006</td>
</tr>
<tr>
<td>First consolidation therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 3</td>
<td>81</td>
<td>0.96</td>
<td>0.47-1.97</td>
<td>0.909</td>
</tr>
<tr>
<td>Autologous SCT</td>
<td>26</td>
<td>0.53</td>
<td>0.25-1.13</td>
<td>0.101</td>
</tr>
<tr>
<td>Allogeneic SCT</td>
<td>79</td>
<td>0.22</td>
<td>0.11-0.44</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Treatment arm control group</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clofarabine</td>
<td>100</td>
<td>0.70</td>
<td>0.45-1.10</td>
<td>0.120</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; CR, complete remission; HR, hazard ratio, na, not applicable; SCT, stem cell transplantation; WBC, white blood cell count. OS and CIR are measured from date of the BM sample after the second cycle. An event was defined by either the occurrence of relapse or death. *Defined as CR after the 1st cycle of chemotherapy. Event defined as relapse or death.
TABLE S8 | MRD/LSC status in the four cytogenetic risk groups

<table>
<thead>
<tr>
<th>Risk group</th>
<th>Good</th>
<th>% Intermediate</th>
<th>% Poor</th>
<th>% Very poor</th>
<th>% Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRD&lt;sup&gt;neg&lt;/sup&gt;</td>
<td>54</td>
<td>68</td>
<td>30</td>
<td>54</td>
<td>33</td>
<td>52</td>
</tr>
<tr>
<td>MRD&lt;sup&gt;pos&lt;/sup&gt;</td>
<td>10</td>
<td>13</td>
<td>10</td>
<td>18</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>MRD&lt;sup&gt;neg&lt;/sup&gt;LSC&lt;sup&gt;neg&lt;/sup&gt;</td>
<td>13</td>
<td>16</td>
<td>11</td>
<td>20</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>MRD&lt;sup&gt;pos&lt;/sup&gt;LSC&lt;sup&gt;pos&lt;/sup&gt;</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

In these analysis same patients are included as shown in Table 7A. This table shows that, without taking into account the CD34/LSC status at diagnosis, there is only a moderate decrease of the frequency of the MRD<sup>neg</sup>LSC<sup>neg</sup> population from good risk patients to very poor risk patients (68%, 53%, 52%, 47%). Similarly, there is no clear change in the MRD<sup>neg</sup>LSC<sup>neg</sup> population from good risk to very poor risk (11%, 18%, 8%, 6%), but there is an increase for the MRD<sup>neg</sup>LSC<sup>pos</sup> population (17%, 20%, 32%, 32%), while for MRD<sup>pos</sup>LSC<sup>pos</sup> population there is also an increase (4%, 9%, 8%, 15%)

Supplementary figures

FIGURE S1 | LSC cut-off in HOVON 42a study

This figure shows LSC data as defined in our earlier performed retrospective HOVON study (HOVON/SAKK 42a)<sup>5</sup> Figure S1A shows the HOVON42a data of the 77 AML patients, where LSC negative were discriminated from LSC positive patients using a 0.0000% cut-off. Plot S1B shows the results for the 0.0001% (1 LSC per 1.000.000 WBCs) cut-off as used in the previously published paper.<sup>5</sup> This figure shows that the 0.0000% cut-off as used in our present paper to distinguish LSC positive from LSC negative patients can also be properly applied to the earlier Hovon42a dataset.
FIGURE S2 | LSC results for event free survival depending on the number of WBC events measured

Figure S2A shows EFS for the group of 242 AML patients when all patient samples are included, independent on the number of WBCs acquired using flow cytometry (A). Results are similar when only patient samples in which ≥ 1 million WBCs were acquired, were included (n=168). When including only cases with ≥ 2 million WBCs acquired results were even better, although the patient group was small (n=37). Note that the decrease of numbers did not largely affect the distribution of patients over the two patient groups: ratio LSCneg/LSCpos was 2.1, 1.9 and 1.8, respectively, for A, B and C, showing lack of bias in the use of samples with different WBC count.

FIGURE S3 | Prognostic value of the leukemic stem cell frequency as defined after induction therapy

This figure shows cumulative incidence of relapse (A) and overall survival (B) for 302 patients in which the LSC frequency was determined after induction chemotherapy. In multivariate analysis the LSC-high status had a HR for CIR of 1.87 and 1.62 for OS.
**FIGURE S4 | Prognostic value of minimal residual disease after induction therapy**

Relapse incidence (A) and overall survival (B) results are shown for the 305 AML patients in which MRD status was defined after the second cycle of chemotherapy treatment. In multivariate analysis MRD status had a HR for CIR of 2.49 and for OS 1.70 (trend).