A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia

Wendelien Zeijlemaker, Angèle Kelder,
Yvonne JM Oussoren-Brockhoff, Willemijn J Scholten,
Alexander N Snel, Dennis Veldhuizen, Jacqueline Cloos,
Gert J Ossenkoppele, Gerrit J Schuurhuis

Published in: Leukemia 2016
Abstract

Relapses after initial successful treatment in acute myeloid leukemia are thought to originate from the outgrowth of leukemic stem cells. Their flowcytometrically assessed frequency is of importance for relapse prediction and is therefore assumed to be implemented in future risk group profiling. Since current detection methods are complex, time- and bone marrow consuming (multiple-tubes approach), it would be advantageous to have a broadly applicable approach that enables to quantify leukemia stem cells both at diagnosis and follow-up.

We compared 15 markers in 131 patients concerning their prevalence, usefulness and stability in CD34+CD38 leukemic stem cell detection in healthy controls, acute myeloid leukemia diagnosis and follow-up samples. Ultimately, we designed a single 8-color detection tube including common markers CD45, CD34 and CD38, and specific markers CD45RA, CD123, CD33, CD44, and a marker cocktail (CLL-1/TIM-3/CD7/CD11b/CD22/CD56) in one fluorescence channel. Validation analyses in 31 patients showed that the single tube approach was as good as the multiple-tube approach. Our approach requires the least possible amounts of bone marrow, and is suitable for multi-institutional studies. Moreover, it enables detection of leukemic stem cells both at time of diagnosis and follow-up, thereby including initially low-frequency populations emerging under therapy pressure.
Introduction

After achieving complete remission (CR) many patients with acute myeloid leukemia (AML) experience a relapse, resulting in dismal outcome.1 It is generally accepted that chemotherapy resistant leukemic cells often referred to as leukemic stem cells (LSCs) or leukemia initiating cells are responsible for relapse. The frequency of such minimal or measurable residual disease (MRD) cells offers an important post-remission risk factor in AML.2-7 MRD, determined using immunophenotypical and/or molecular procedures, is implemented in risk classifications of many major AML trial groups. However, relapses do occur in a considerable number (20%-70%) of patients with low/negative MRD levels, even in the intermediate risk group.2-7 To account for these “false-negative” cases, one may focus on these surviving LSCs that are at the basis of outgrowth of MRD cells to overt relapse. Although different cellular compartments (CD34+CD38+, CD34+CD38 , CD34 ) may contain LSCs,8-12 CD34+CD38 stem cells seem to be most therapy resistant and least immunogenic.13-16 This is in accordance with the finding that CD34+CD38 or CD34+CD38dim LSC frequency is highly predictive for relapse in AML.9 Additionally, CD34+CD38 LSC frequency adds important prognostic information to MRD assessment, reducing the number of false-negative MRD cases.9 Overall, the LSC assessment is instrumental in defining remission quality and predicting relapse risk. However, identification of LSCs can be challenging due to the very low frequency of this stem cell population (frequency 0.2 to 625 cells per 106 mononuclear cells).14,18 It is known that compared to CD34+CD38 hematopoietic stem cells (HSCs), CD34+CD38 LSCs often aberrantly express cell surface markers,19 whereby patterns may be very heterogeneous. Markers expressed on LSCs include myeloid antigens CD13, CD33 and CD123,20,21 CLL-1 (also known as CLEC12A)22 and the lineage markers CD2, CD7, CD11b, CD14, CD15, CD19, CD22 and CD56.12,23 The expression of CD96, also known as Tactile, and T-cell Ig mucin 3 (TIM-3), is also higher on LSCs as compared to HSCs.24,25 Although CD44 is already highly expressed on HSCs, LSCs often show over-expression for CD44.26 Using aberrant markers and well-established functional assays like aldehyde dehydrogenase (ALDH)27-29 and the side population30, the normal and neoplastic nature of such immunophenotypically defined CD34+CD38 LSCs and HSCs could be confirmed using molecular and cytogenetic assays in multiple samples.5,27,30,31 Moreover, current LSC studies have revealed that marker positive CD34+CD38 cells show leukemia engraftment in different mouse models.12,20-22,24-26,32 Marker expression differs between and within patients9,20,22-24 and therefore different immunophenotypically defined LSC compartments may be associated with specific sub-populations showing different sensitivity for therapy.33 Due to tumor heterogeneity accurate flow cytometric LSC detection at time of diagnosis requires extensive antibody panels. In addition, besides molecular constitution,34,35 also immunophenotypes can change between diagnosis and relapse.36 As a consequence, a broader panel of markers is also required during follow-up, enabling the detection of populations emerging during disease. Recent studies have shown that HSCs can also harbor leukemia-specific mutations.37 Although these so-called pre-leukemic HSCs are non-leukemic, they are significantly different from real HSCs since they are supposed to be of importance in the process of leukemogenesis.38 In case the relapse aberrancies are very distinct from diagnosis, it can be suggested that the relapse
evolved from such a “pre-leukemic” clone. To facilitate complete and accurate CD34+CD38+ LSC detection in AML patients, often with small amounts of bone marrow (BM) cells available, we composed a single flow cytometric tube. Expression and usefulness of 15 different cell surface markers (apart from the backbone markers, CD45, CD34 and CD38) were compared in a large cohort of AML patients. This enabled the design of a single 8-color LSC detection tube consisting of a cocktail of 6 markers in one fluorescence channel combined with single channels for the backbone markers (CD45, CD34, CD38) and stem cell markers (CD44, CD33 and CD123). Overall, we defined a single tube containing in total 13 antibodies, which enables CD34+CD38+ LSC detection in a broadly applicable, less expensive and more efficient manner than the current detection strategies.

Patients, Materials and Methods

Patients
In total 236 patients were screened for initial analyses at time of AML diagnosis. For data concerning marker stability 132 follow-up samples were acquired. All samples were gathered between 2010 and 2015. All patients had a cytopathologically confirmed diagnosis of AML according to the WHO classification (excluding acute promyelocytic leukemia) or a diagnosis of refractory anemia with excess of blasts and IPSS score ≥1.5. Patients were treated according to HOVON/SAKK clinical trials (www.hovon.nl) and provided their written informed consent before entrance into the study (for approval numbers see supplementary file). BM samples from pathological controls (details in supplementary file) and healthy donors were used to investigate cell surface expression on HSCs.

Immunophenotyping
Erythrocyte-lysed (Pharm lyse, Becton and Dickinson [BD]) fresh BM samples were used to perform 8-color multiparameter flow cytometry using a FACS Canto II from BD (San Jose, CA, USA). After lysis, cells were washed with phosphate buffered saline/0.1% human serum albumin and labeled with the appropriate antibodies. Cells and antibodies were incubated for 15 minutes in the dark at room temperature and subsequently, cells were washed to remove unlabeled antibodies. In general a minimum of 500 000 white blood cells (WBCs) were acquired. Detailed information concerning the used antibodies is provided in the data supplement. Table 1 shows the conventional 7-tube 8-color antibody panel at diagnosis.
TABLE 1 | 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>HV500c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PBS</td>
<td>PBS</td>
<td>CD13</td>
<td>PBS</td>
<td>PBS</td>
<td>PBS</td>
<td>CD34</td>
<td>CD45</td>
</tr>
<tr>
<td>2</td>
<td>CD44</td>
<td>CLL-1</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>

Gating strategy and marker selection

Samples were analyzed using Infinicyt™ software, v. 1.7 (Cytognos SL, Spain). The gating strategy is shown in Figure S1A-D. Within the blast population presumed HSCs were defined as CD34+/CD38/FSC<low/SSC<low/CD44+/Lin−/CD123<low/CD33<low/CLL-1/TIM-3− and if available CD45RA+ (Lin− means CD2/CD7/CD11b/CD14/CD15/CD19/CD22/CD56). Presumed LSCs were defined as CD34+/CD38− and CD44+ and/or CD123− and/or CD33+ and/or CLL-1+, and/or TIM-3− and/or Lin− and/or if available CD45RA+. CD45RA, a CD45 isoform generated by alternative splicing, was included at a later stage, since it also was a broadly useful LSC marker. For reasons of simplicity, presumed HSCs and LSCs are further referred to as HSCs and LSCs. In previous work, where we used the same flow cytometric method to discriminate HSCs and LSCs, we provided proof that flow cytometrically defined CD34+CD38− HSCs, after injection of this cell fraction in mice, indeed gave multilineage engraftment, while in mice injected with CD34+CD38− LSCs, leukemic engraftment was seen. To determine which markers were needed for the LSC detection tool, a scoring system was used to, retrospectively, define and compare the usefulness of above mentioned cell surface markers. This scoring system takes into account background autofluorescence of individual markers which were sometimes measured in different fluorochrome channels and consequently had differences in brightness. Therefore, this scoring system was created to ensure a less subjective comparison as compared to solely investigating Median Fluorescence Intensities (MFIs) of the different markers. This scoring system ranked from 0-3 points, whereby, in short, one point was attributed to a marker for a) clear distinction of two populations within the CD34+CD38− fraction, b) high negative predictive value of the particular marker (no LSC pollution in marker negative fraction), and c) high sensitivity (maximal LSC coverage) (Figure 1).

Based on the number of points given, one (or sometimes multiple) best marker(s) were defined in each AML sample. The details are outlined in the legend of Figure S1. If a marker scored 3 points, it was by definition a ‘best marker’. In the minority of patients (8/131), however, the ‘best marker’ scored less than 3 points. The scores were used to compare overall marker prevalence and usefulness regarding LSC detection in AML in 131 cases in which enough LSCs were present (cluster of ≥5 cells) to evaluate marker performance of the in total 236 screened cases. For examples and more detailed information concerning this scoring system, see Figure S1.
FIGURE 1 | Scoring system criteria to evaluate marker performance
Marker positive and marker negative cells are shown in red and green, respectively. Section I shows the first scoring system criterion: a) clear distinction between two populations within the CD34^+ CD38^- fraction. A and B show two different patients whereby both markers (CLL-1 and CD7) show two different populations within CD34^+ CD38^- . Section II shows the second criterion: b) high negative predictive value of the marker. CLL-1 expression in patient #1711 is shown (A) in which a clear population of HSCs is present (based on both marker negativity and FSC^low/SSC^low scatter properties, as shown in B). CLL-1 has only little expression on CD34^+ CD38^- cells and all marker negative cells are presumed HSCs, implying that there are no/few LSCs present within the CLL-1 negative CD34^+ CD38^- fraction. Finally, section III shows the last criterion: c) high sensitivity of the marker. Both CD33 (A) and TIM-3 (B) are highly positive on CD34^+ CD38^- cells and almost all LSCs were marker positive. In case the marker of interest fulfilled one of the criteria, 1 point was given; thereby the total number of points in the scoring system ranged from 0-3 points. Gating strategy of CD34^+ CD38^- cells and examples of how scores were defined are shown in Figure S1.
Statistics
Statistical analyses were performed with IBM SPSS Statistic 20 and Graphpad prism 5. The Mann-Whitney U test was used to compare non-parametric unpaired variables. Correlation coefficients were calculated using the Spearman’s Rho test (two-sided) and scatter plots were produced to compare LSC data obtained with the two different panels. P-values were considered significant below 0.05.

Results
Prevalence of aberrant cell surface markers on CD34+CD38− cells at diagnosis
The aim of this study was to design a single 8-color tube that would allow specific identification of CD34+CD38− LSCs. For 15 different cell surface markers, we compared general prevalence, regardless of the percentage and intensity of expression on CD34+CD38− cells. A marker was defined as present on CD34+CD38− when it scored at least 1 point according to the scoring system described in the methods section. Figure 2 shows that CD123 and CD33 were most often aberrantly expressed with a prevalence of 82.3% (107/130) for both these markers. Furthermore, the following cell surface markers showed aberrant expression on CD34+CD38− cells in more than half of the cases: CLL-1 (70%), TIM-3 (62%), CD11b (55%) and CD22 (51%). Additionally, CD7 (43%), CD96 (33%), CD56 (32%), and CD15 (30%) were all markers that were present between 30%-50% of the evaluated cases. CD44 is an exceptional marker since it is highly expressed on HSCs and indeed CD44 was found to be expressed on the total CD34+CD38− compartment in 100% of the evaluated AML cases (n=131, example in Figure S1.V.D). Aberrant CD44 expression, characterized by an over-expression of CD44 (usually MFI of CD44++ on CD34+CD38− >25 000) was found in 31% (40/131) of the cases (Figure 2). Finally, CD13 (29%), CD2 (18%), CD19 (8%) and CD14 (2%) were aberrantly expressed in less than 30% of the cases.

![Figure 2](image)

**FIGURE 2 | Performance of different cell surface markers for LSC detection**
This figure shows performance of 15 different cell surface markers for detection of CD34+CD38− LSCs in AML. Scores were defined based on criteria shown in Figure 1 and as described in the legends of Figure S1. Numbers above the bar represent general prevalence of the specific marker as compared to the total number of AML samples studied. Overall, this shows that in general CD123 and CD33 are the best markers for specific detection of CD34+CD38− LSCs, followed by CLL-1 and others.
CD33 and CD123 perform best in detection of leukemic and normal CD34+CD38- cells

Performance of the individual markers for LSC detection was investigated using the scoring system described in the methods section. Figure 2 gives an overview of the scores of the different markers. CD123 and CD33 not only had the highest prevalence, but scored the maximum number of three points in 60% and 59% of all evaluable cases, respectively. This implies that these markers not only had high expression levels, but also most often showed a clear distinction in antigen expression between presumed CD34+CD38- HSCs and LSCs. TIM-3 (24%) and CD11b (16%) had 3 points in >10% of the samples, followed by CD22 (10%), CLL-1 (9%), CD44 (8%), CD96 (8%), CD56 (6%), CD13 (5%), and CD7 (4%). Additionally, CD15 (2%), CD2 (2%), and CD19 (2%) were very useful in only very few cases. Finally, CD14 had no 3 points scored at all. Note that more than one marker may score 3 points in the same AML case.

Essential markers for identifying CD34+CD38- LSCs

Of the 131 evaluable cases there were 48 cases (37%) in which only 1 best marker for LSC detection was present (not shown). In 47/48 cases the best marker scored 3 points, however in 1 case the best marker scored 2 points. These 48 cases thus provided the most essential information as to defining which particular markers are definitely needed in the single LSC detection tube and which markers were redundant. The definition of such antibodies is that, for each AML case studied, at least one of the other antibodies was better. Of these 48 cases, CD123 was best in 44% of the samples (21/48), CD33 in 33% (16/48), CLL-1 in 8% (4/48), TIM-3 in 8% (4/48), CD22 in 4% (2/48) and CD56 in 2% of the cases (1/48). Furthermore, in all remaining 83/131 (63%) cases, it was one of these 6 markers that was an optimal marker. Consequently, these 6 markers were considered essential in the formation of a diagnostic tool with the aim to specifically detect LSCs in a maximal number of patients. Therefore, CD2, CD7, CD11b, CD13, CD14, CD15, CD19, CD44 and CD96 were redundant. However, seen the high incidence of CD7 and CD11b (Figure 2), we left these markers in the study for further experiments. CD13 was not only highly expressed on HSCs, but also showed low incidence (Figure 2). This marker was therefore excluded from further analysis.

CD33 and CD123 are expressed on CD34+CD38- HSCs in follow-up BM

To enable specific LSC detection in a follow-up BM it is of importance that marker expression on stem cells remains stable during disease and/or treatment. At follow-up, in most cases the majority of CD34+CD38- cells are presumed HSCs. Up-regulation on HSC of markers used at diagnosis for LSC identification might lead to largely over-estimated LSC frequencies. Table 2 and Table S1 show that marker expression on HSCs of all markers is low at diagnosis (CD44 was excluded from these experiments since CD44 is highly expressed on HSCs both at diagnosis and follow-up). High MFI ratios are seen for CD123 in follow-up BM (Table 2), suggesting an up-regulation of this marker on HSCs after treatment, especially since CD123 is expressed, albeit at low levels, in NBM. Consistent with previously published results,19 we
found CD33 to be clearly expressed on HSCs in NBM (Table S1). Table 2 shows that CD33 expression is strongly suppressed on HSCs present in AML diagnosis BM as compared to HSCs in NBM (p<0.001) and follow-up BM (p<0.001). Figure S2 shows the MFI ratios of Table 2 including MFI values of both the CD34⁺CD38⁻ HSCs and the lymphocyte sub-fractions. This figure shows that the high MFI ratios as found for CD33 and CD123 in follow-up BM and in the normal control cases cannot be explained by differences in background expression of the control population (lymphocyte sub-populations). Although for many markers MFI ratios are significantly different between the different type of samples, the wide range of ratios as seen for CD33 and CD123 in both AML follow-up and NBM (Table 2) shows that although CD33 and especially CD123 are markers that are important to discriminate LSC and HSC at diagnosis, they should be used with care at follow-up.

### TABLE 2 | Median MFI ratios of CD34⁺CD38⁻ HSC population as compared to lymphocytes

<table>
<thead>
<tr>
<th>AML diagnosis BM</th>
<th>AML follow-up BM</th>
<th>NBM</th>
<th>Diagnosis vs f-up</th>
<th>Diagnosis vs NBM</th>
<th>Follow-up vs NBM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (range)</td>
<td>n</td>
<td>Median (range)</td>
<td>n</td>
<td>Median (range)</td>
</tr>
<tr>
<td>CL-1</td>
<td>0.7 (0.1-2.1)</td>
<td>27</td>
<td>4.1 (0.3-14.0)</td>
<td>30</td>
<td>1.6 (0.3-3.0)</td>
</tr>
<tr>
<td>TIM-3</td>
<td>0.6 (0.2-4.9)</td>
<td>32</td>
<td>2.2 (0.3-10.8)</td>
<td>28</td>
<td>0.8 (0.6-1.6)</td>
</tr>
<tr>
<td>CD7</td>
<td>1.0 (0.4-3.1)</td>
<td>32</td>
<td>2.4 (0.4-5.9)</td>
<td>37</td>
<td>1.7 (0.2-3.6)</td>
</tr>
<tr>
<td>CD11b</td>
<td>1.0 (0.2-2.7)</td>
<td>24</td>
<td>3.1 (0.5-5.9)</td>
<td>22</td>
<td>2.5 (0.7-5.4)</td>
</tr>
<tr>
<td>CD22</td>
<td>2.0 (0.9-6.6)</td>
<td>12</td>
<td>3.5 (1.4-10.6)</td>
<td>9</td>
<td>5.4 (2.0-11.0)</td>
</tr>
<tr>
<td>CD33</td>
<td>2.1 (0.6-2.3)</td>
<td>30</td>
<td>5.5 (0.1-910.6)</td>
<td>35</td>
<td>19.4 (1.0-130.5)</td>
</tr>
<tr>
<td>CD56</td>
<td>1.1 (0.3-4.0)</td>
<td>25</td>
<td>4.3 (0.7-26.4)</td>
<td>24</td>
<td>2.5 (1.2-6.8)</td>
</tr>
<tr>
<td>CD123</td>
<td>3.9 (0.4-17.7)</td>
<td>30</td>
<td>14.5 (1.3-40.1)</td>
<td>25</td>
<td>5.8 (0.5-86.3)</td>
</tr>
</tbody>
</table>

AML, acute myeloid leukemia; BM, bone marrow; MFI, median fluorescence intensity; NBM, normal bone marrow; n, number of samples. P-values were calculated using the Mann-Whitney U test. MFI ratios were calculated as follows: MFI of marker on CD34⁺CD38⁻ divided by MFI of marker on negative control population (lymphocytes).

### Composition and validation of CD34⁺CD38⁻ LSC detection tool

Based on above described results concerning performance, prevalence, redundancy and stability on HSCs of the different markers, one tube could be designed that enables specific LSC detection in almost all AML cases, both at time of diagnosis and follow-up. CLL-1, TIM-3, CD7, CD11b, CD22, CD33, CD56 and CD123, chosen for specific LSC detection, had to be included in the LSC tube, together with the backbone markers CD45, CD34 and CD38, making up 11 antibodies. To enable such for a single 8-color tube, we tested whether we could combine multiple markers in one fluorescence channel. To enable this for both diagnosis and follow-up, a prerequisite for the combination of these markers was to be and remain marker negative on HSCs. Therefore, a cocktail of such stable cell surface markers, including CLL-1, TIM-3, CD7, CD22, CD56 and CD11b, was used in PE-fluorescence channel, which usually provides strong fluorescence for all single markers. Cell surface markers CD33 and CD123 are useful for follow-up since, despite up-regulation of expression on HSCs, expression differences between LSC
and HSC remain due to usually over-expression on LSCs. Nevertheless, due to their instability, these markers should be evaluated in a separate fluorescence channel. Table 3 shows the newly developed single 8-color flow cytometry tube.

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>8-Color LSC detection tube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube</td>
<td>FITC</td>
</tr>
<tr>
<td>1</td>
<td>CD45RA</td>
</tr>
<tr>
<td>2</td>
<td>CLL-1</td>
</tr>
</tbody>
</table>

Accuracy of the marker cocktail in the PE-channel of the one tube panel was checked by comparing CD34−CD38− LSC levels (as percentage of the total amount of WBCs) with the conventional antibody panel from Table 1. In 9 NBM samples it was validated that the marker cocktail of the single LSC tube indeed had no expression on HSCs (Example in Figure 3A; in all 9 samples there was 0% positivity on CD34−CD38− cells; median MFI ratio of marker cocktail on CD34−CD38− HSCs was 1.06, range 0.59-1.54). Additionally, in 6 pathological control samples also no expression on HSCs was found (median 0% positivity, median MFI ratio 1.52, range 1.09-1.92). Furthermore, as negative control, we verified in 9 CD34-negative AML samples with only HSCs present⁴³, that no expression on CD34−CD38− cells was found with the marker cocktail (Example in Figure 3B; 0% positivity, n=9; median MFI ratio of marker cocktail was 1.24, range 0.46-7.43). These control samples show that the LSC detection tube is highly specific.

![FIGURE 3](image_url) | Specificity of marker cocktail in the LSC detection tube
This figure shows the absence of expression of the marker cocktail on CD34−CD38− cells in NBM (A) and a CD34-negative AML (B, in this type of AML only CD34−CD38− HSCs are present). In C, expression of the marker cocktail is shown in a CD34-positive AML in which both HSCs (marker cocktail negative) and LSCs (marker cocktail positive, 0.0586%) are present. Percentages represent the amount of marker positive CD34−CD38− cells as compared to the total amount of WBCs.
When comparing 22 AML samples, with CD34⁺CD38⁻ LSCs present (Example in Figure 3C), a correlation coefficient of 0.968 between LSC measured with the best single marker in the cocktail and the 6 markers combined in PE-channel, was found (Figure 4A, p<0.001). This clearly shows that for quantification, the combination of 6 markers is as good as the best individual marker and in 18 cases slightly better (Figure 4A; median difference in LSC% obtained with the marker cocktail and the best single marker is 0.001%, range 0.000%-0.836%). Only in 4 cases we found that the marker cocktail had slightly lower expression than the highest individual marker (median difference in LSC% obtained with the best single marker and the marker cocktail is 0.459%, range 0.000%-3.538%).

![Graph showing correlation between LSC marker cocktail and best individual marker](image)

**FIGURE 4 | Performance of marker cocktail: comparison with best single marker and with CD45RA expression**

Comparison of LSC results (LSC percentage calculated as percentage of the total amount of WBC, log transformed) in 22 AML diagnosis samples with CD34⁺CD38⁻ LSC present. Results are obtained with the best individual marker of the conventional panel (x-axis, A) and obtained with the marker cocktail (y-axis, A). The dashed grey line represents the y=x line. B shows a comparison of LSC results obtained with the marker cocktail and CD45RA, a potential stem cell marker. Samples are ordered according to increasing percentages obtained with the cocktail.

An example of the efficacy of LSC detection using our single tube is shown in Figure 5. Figure S3 and S4 illustrate the efficacy of the marker cocktail in case of a stable or, on the other hand, instable immunophenotype during disease progression. CD45RA was included in the LSC tube, not only to objectify possible clonal differences within the LSC fraction but also since it is an important marker for LSC detection in the majority of AML patients. Expression of CD45RA was studied in the same 22 samples used for the comparison of single markers with the cocktail (Figure 4A). LSC results obtained with the marker cocktail and CD45RA were in general comparable, however, the marker cocktail being slightly better than CD45RA (Figure 4B; Median difference between marker cocktail and CD45RA: 0.002%, range 0.000%-1.996%). In the 7 samples where CD45RA had higher LSC% the median difference between CD45RA and the marker cocktail was 0.010%, range 0.000%-0.797%.
**FIGURE 5. Efficacy of LSC detection tube in diagnosis AML (#2128)**

After labeling of the cells with antibodies, CD34+ blasts were identified as shown in Figure S1.I.A-D and subsequently CD34+CD38+ cells were gated within this fraction (CD34+CD38+ blasts shown in dark grey). Figure 5.I.A-F shows the expression of the 6 individual markers (CLL-1/CLEC12A, TIM-3, CD7, CD11b, CD22 and CD56) on CD34+CD38+ cells. The percentage of identified LSCs is shown for each of these individual markers. In this particular patient, TIM-3 (0.0282% LSCs) and especially CD56 (0.0257% LSCs) are most useful markers for LSC detection. Figure 5.II shows the result of the marker cocktail in the LSC-tube, whereby CLL-1, TIM-3, CD7, CD11b, CD22 and CD56 are combined in the PE fluorescence-channel. Although individual expression patterns are no longer visible, the result of the marker cocktail in the LSC tube (Figure 5.II) is as good as the best individual marker (here CD56 in Figure 5.I.F). Note that expression percentages of the best individual markers (TIM-3, CD56) and the cocktail are very similar: 0.0282%, 0.0257% and 0.0292%, respectively. Percentage of LSCs identified by the rest of the markers in the LSC detection tube (markers not shown) were less accurate in this sample as compared to the marker cocktail (CD45RA 0.0123%, CD33 0.0043%, CD123 0.0008% and CD44 0.0000%).
Discussion

In this study we developed a broadly applicable one tube approach to identify specific immunophenotypic CD34+/CD38 LSC. Previous reports have demonstrated the prognostic importance of CD34+/CD38 LSC frequency both at time of diagnosis and after treatment.\(^5,41\) Moreover, they further refine current MRD based risk classification, hence improving the prediction of an emerging relapse.\(^9\) We speculate that the CD34+/CD38 LSC frequency will be incorporated in future risk classification in multicenter trials and in monitoring allogeneic transplantation.\(^17\) Consequently, a CD34+/CD38 LSC detection approach that is broadly applicable is very desirable. For reliable LSC detection marker expression should not only be stable on HSCs (Table 2, Table S1, Figure S2), but also on LSCs. However, losses and gains of individual antigens during disease have been frequently observed.\(^36\) For CLL-1 and the lineage markers it has been shown, in small patient groups (n=2-9), that expression on CD34+/CD38 was relatively stable between diagnosis and relapse.\(^22,23\) Using a marker cocktail in one fluorescence channel, as included in the LSC detection tube, any immunophenotypic shift between the markers in that channel does not affect accurate LSC measurement (Figure S3 and Figure S4). This LSC tube thus also anticipates on new LSC populations that can emerge during treatment and/or disease.\(^31\) Importantly, by using this marker cocktail we were able to achieve similar or even better results as compared to the conventional 7 tube antibody panel (Figure 4A). Moreover, our LSC detection tube enables to include additional newly discovered LSC markers of interest, provided that these should not be up-regulated on HSCs. Although not shown in this paper, the specificity of CD34+/CD38 LSC detection can be further improved by using secondary gating strategies based on differences in CD34 and/or CD45 expression and differences in light scatter properties.\(^9\) Importantly, in this paper we define HSCs both on the absence of aberrant marker expression and specific light scatter characteristics (FSC\(^{low}/SSC\(^{low}\)). We have demonstrated that injection of such defined HSCs indeed gave multilineage engraftment in mice.\(^9\) It has however to be emphasized that recent studies\(^37,42,43\) have shown that these HSCs can harbor leukemia-specific mutations such as IDH1 and IDH2\(^42\), TET2\(^43\), and DNMT3a.\(^37,42\) The establishment of the role of these so-called pre-leukemic HSCs in leukemogenesis and relapse development needs further investigation.\(^38\) Furthermore, it must be taken into account that leukemia initiating capacity is not only present in CD34+/CD38, but also in CD34+/CD38- and CD34- immunophenotypic compartments.\(^8,12\) However, the CD34+/CD38 population in a CD34-positive AML has the highest engraftment potential in mice\(^13,14,41\) and is most therapy resistant in vitro and in vivo.\(^13,16\) In addition, in diagnosis AML there was a very strong prognostic value for CD34+/CD38 LSCs, but no impact whatsoever for CD34+/CD38- and CD34- cells.\(^9\) Importantly, final analyses of a large clinical study are currently ongoing, wherein the prognostic value of the CD34+/CD38 LSC frequency, as determined using the here described method, should be confirmed. In the absence of CD34+/CD38 LSCs, it is however plausible that the CD34+/CD38- and/or CD34- cells cause the leukemic engraftment. In CD34-negative patients (~20%) the leukemia initiating cells are present in the CD34-negative compartment.\(^27,31,44\) Since this compartment makes up the vast majority of the blast cells, the
putative stem cells here may be hidden in a sub-compartment which may be identified by functional assays like the side population.\textsuperscript{30,45} Interestingly, the LSC detection tube also enables identification of CD34-negative patients, who experience fewer relapses and have a longer survival as compared to CD34-positive patients.\textsuperscript{31} Overall, our results clearly demonstrate that the use of both a cocktail of markers in one fluorescence channel and of single markers with specific properties in the other fluorescence channels is feasible for accurate LSC detection, which is broadly applicable in a multi-institutional setting. For specific purposes, the panel can be extended with additional antibodies in the PE-channel. The LSC detection tube is lower in costs and requires less BM material as compared to a multiple-tubes approach. In addition, this tube allows to detect not only residual cells that have the immunophenotype established at diagnosis, but also LSCs with emerging immunophenotypes.

Recently, a lyophilized version of the here described LSC tube was manufactured and validation experiments are currently ongoing for an improved standardized LSC detection procedure in the near future.

**Acknowledgements**

We thank all participating study centers for including their patients in the HOVON trials.
References


Supplementary information

Patients

The HOVON (Hemat-Oncology foundation for adults in Netherlands) studies were centrally reviewed and approved by Medical Ethical Review Committee (METC), Erasmus Medical Center, Rotterdam for all participating centers for the total study (clinical plus side studies). Protocol numbers and approval numbers are: H102 (2009-293) and H103 (2010-009). In addition, the VU Amsterdam METC review board approved these centrally approved studies for local feasibility with HOVON/METc numbers: H102 (2010/56), H103 (2010/256). In both studies, patients had to provide their written informed consent before entrance into the study.

Immunophenotyping

Antibodies used for the analyses in this paper were as follows: CLL-1/CLEC12A-PE (clone 50C1, BD), TIM-3-PE (clone 332774, BD), CD2-FITC/PE (clone MT910, DakoCytomation), CD7-FITC/PE (clone M-T701, BD), CD11b-FITC (clone Bear1, BC), CD11b-PE (clone D12, BD) CD13-PerCP-CY5.5 (clone WM15, BD), CD14-APC-H7 (clone MoP9, BD), CD15-FITC (clone MMA, BD), CD19-APC-H7 (clone SJ25C1, BD), CD22-PE (clone S-HCL-1, BD), CD33-PE/PE-Cy7 (clone P67.6, BD), CD34-HV450 (clone 8G12, BD), CD34-BV421 (clone 581, BD), CD38-APC (clone HB7, BD), CD44-FITC (clone J173, BC), CD44-APC-H7 (clone G44-26, BD), CD45-HV500c (clone 2D1, BD), CD45RA-FITC (clone L48, BD), CD56-PE (clone My31, BD) CD56-PC7 (clone N901, BC), CD96-PE (clone 6F9, BD), CD123-PE (clone 9F5, BD) and CD123-PerCP-CY5.5 (clone 7G3, BD).

Sample selection

In general, AML samples were considered eligible for the calculation of scores when a cluster of at least 5 cells was present in the CD34+CD38- compartment. To enable a correct interpretation of a particular marker, the CD34+CD38dim fraction was also studied, since aberrant expression of a marker on CD34+CD38 most often coincides with expression on CD34+CD38dim cells (Figure S1.II.C for example shows that CD38dim cells are all CLL-1+). In total 236 AML samples were screened for initial analyses and 197 (83.5%) cases fulfilled the criteria concerning amount of stem cells. In 66/197 (33.5%) cases only HSCs were present in the CD34+CD38- compartment (CD34-negative AML samples and CD34-positive samples with only normal CD34+CD38- cells present). Since these samples were not helpful in evaluating marker performance (LSC versus HSC) concerning the detection of CD34+CD38- LSCs, these 66 cases were excluded from the scoring system analyses. Consequently, scoring system analyses were performed on the remaining 131 cases in which CD34+CD38- LSCs were present at time of diagnosis. The 66 cases with only normal CD34+CD38- cells present were used to assess possible and unwanted up-regulation of marker expression on normal CD34+CD38- cells. To that end, Median Fluorescence Intensity (MFI) of normal CD34+CD38- cells was divided by MFI of the control population, the negative fraction of the lymphocytes. Lymphocytes were hereby
defined as CD45<sup>-</sup>/FSC<sub>low</sub>/SSC<sub>low</sub> whereby for the calculation of MFI ratios the negative fraction of the lymphocytes was used (e.g. CD7 positive T-cells were excluded for the calculation of the MFI ratio for CD7). Using this formula, MFI ratios were calculated for the different antigens expressed on the HSCs present in healthy controls (normal BM [NBM]), at AML diagnosis and in AML follow-up samples. Importantly, gating strategies were similar for these three different sample types. Validation experiments to compare the accuracy of the new LSC design with that of the conventional antibody panel was checked by comparing CD34<sup>+</sup>CD38<sup>+</sup> positivity levels (as percentage of the total amount of white blood cells [WBC]) and MFI ratios in 9 NBM samples, 6 pathological control samples (1 patients with leukopenia most likely caused by an iron disorder, 3 lymphoma patients and 2 multiple myeloma patients) and 9 AML samples with only CD34<sup>+</sup>CD38<sup>+</sup> HSCs and 22 AML samples with CD34<sup>+</sup>CD38<sup>-</sup> LSCs present.

Supplementary Table

**TABLE S1 | Marker positivity percentages on normal CD34+CD38- cells**

<table>
<thead>
<tr>
<th></th>
<th>AML diagnosis BM</th>
<th>AML follow-up BM</th>
<th>NBM</th>
<th>Diagnosis vs f-up</th>
<th>Diagnosis vs NBM</th>
<th>Follow-up vs NBM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (range)</td>
<td>Median (range)</td>
<td>Median (range)</td>
<td>n</td>
<td>p</td>
<td>p</td>
</tr>
<tr>
<td><strong>CLL-1</strong></td>
<td>0.0 (0.0-0.0)</td>
<td>0.0 (0.0-66.7)</td>
<td>0.0 (0.0-0.0)</td>
<td>27</td>
<td>0.05</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>TIM-3</strong></td>
<td>0.0 (0.0-2.9)</td>
<td>0.0 (0.0-25.0)</td>
<td>0.0 (0.0-0.0)</td>
<td>32</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>CD7</strong></td>
<td>0.0 (0.0-11.1)</td>
<td>0.0 (0.0-0.0)</td>
<td>0.0 (0.0-0.0)</td>
<td>37</td>
<td>0.13</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>CD11b</strong></td>
<td>0.0 (0.0-6.7)</td>
<td>0.0 (0.0-20.0)</td>
<td>0.0 (0.0-0.0)</td>
<td>22</td>
<td>0.45</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>CD22</strong></td>
<td>0.0 (0.0-13.6)</td>
<td>0.0 (0.0-0.0)</td>
<td>0.0 (0.0-0.0)</td>
<td>9</td>
<td>0.12</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>CD33</strong></td>
<td>0.0 (0.0-100.0)</td>
<td>17.5 (0.0 - 100.0)</td>
<td>59.5 (0.0- 100.0)</td>
<td>30</td>
<td>0.03</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>CD56</strong></td>
<td>0.0 (0.0-0.0)</td>
<td>100.0 (0.0 -16.7)</td>
<td>0.0 (0.0-0.0)</td>
<td>24</td>
<td>0.15</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>CD123</strong></td>
<td>0.0 (0.0-90.9)</td>
<td>94.7 (0.0 -100.0)</td>
<td>1.0 (0.0 -33.3)</td>
<td>25</td>
<td>&lt;0.001</td>
<td>0.06</td>
</tr>
</tbody>
</table>

AML, acute myeloid leukemia; BM, bone marrow; NBM, normal bone marrow; n=number of samples. P-values were calculated using the Mann-Whitney U test. Marker positivity was defined using a negative control population (e.g. negative fraction of the lymphocyte population)
Supplementary Figures

FIGURE S1 | Performance of markers for the identification of LSCs and HSCs in the CD34+CD38- compartment

I. Gating of CD34+CD38- blast cells (#1634)
Red blood cells were lysed and subsequently BM cells were labeled with the appropriate antibodies (Table 1). In a FSC/SSC plot remaining erythrocytes and dead cells were excluded (A). WBCs were identified
based on CD45 expression (B) and within the WBC population, blasts cells were gated based on CD45dim expression and SSC<sup>dim</sup> (C). Subsequently, CD34<sup>+</sup> blast cells were gated within this population of blasts cells (D). Within the CD34<sup>+</sup> blast population, CD38<sup>+</sup> cells were gated. After the gating procedure, for each of the following conditions 1 point was given to the particular marker; according to this, the number of points ranged from 0–3 points. A point was given if a) the marker of interest had a clear distinction between CD34<sup>+</sup>CD38<sup>+</sup> marker+ and marker- cells, which can also be a clear over-expression of a particular marker; b) high negative predictive value of the particular marker implying no/little pollution with LSCs in the marker negative fraction (percentage of marker negative cells was ≤10% less than for the best marker used for that particular patient [total CD34<sup>-</sup>CD38<sup>-</sup> compartment set on 100%]) and c) the marker of interest had a high sensitivity (defined if the marker expression on CD34<sup>+</sup>CD38<sup>+</sup> was factor ≤2 different as compared to the best marker measured in that particular AML sample. Figure S1 row II-V only show marker expression on CD38<sup>+</sup> (CD34<sup>+</sup>CD45<sup>dim</sup>) cells. Marker positive and negative CD34<sup>+</sup>CD38<sup>-</sup> cells are shown in red and green, respectively.

**II. Examples of cell surface markers with 3 points (#1742, #1733, #1771 and #1824)**

Four different markers in four different patients (#1742, #1733, #1771 and #1824) are shown whereby all markers, TIM-3 (A), CD22 (B), CLL-1 (C) and CD56 (D) were (one of) the most optimal marker(s) for CD34<sup>+</sup>CD38<sup>-</sup> LSC detection in that particular patient based on the scoring system under I. The criteria for which points were given are listed between parentheses following the number of points. Dashed lines indicate boundaries between negative and positive CD34<sup>+</sup>CD38<sup>-</sup> cells and the percentage indicates the percentage of marker positivity on the stem cells. Note that in two cases (CD22 and CLL-1) the LSC percentage was 100% of the CD38<sup>-</sup> compartment. The level of expression is so high that HSCs would have been easily identified if present. Therefore also 1 point was given for criterion a).

**III. Examples of cell surface markers with 2 points (#1634, #1700)**

Two examples of markers that scored 2 points are shown. In patient #1634, CD33 (column B) scored 3 points and was the best marker. CD96 (A) scored 2 points based on criteria a (clear separation indicated by dashed line) and c (percentage of positivity of CD96 on CD34<sup>+</sup>CD38<sup>-</sup> cells was 67.7% as compared to 80.2% positivity of best marker CD33; difference of expression was thus ≤2 factor different). Since the percentage of marker negative cells for CD96 (100%-67.7%=32.3%) was >10% different than for the best marker CD33 (100%-80.2%=19.8%, no points were given for criterion b).

In patient #1700, CD123 (D) scored 3 points and was the best marker. CD7 (C) scored 2 points based on a) and b) (percentage of CD7<sup>-</sup> CD34<sup>+</sup>CD38<sup>-</sup> cells is 96.1% (100%-3.9%) as compared to 87.4% (100%-12.6%) CD123<sup>-</sup> cells; the CD7 negative fraction is thus ≤10% different as compared to the best marker). Since the percentage of CD7<sup>-</sup> CD34<sup>+</sup>CD38<sup>-</sup> cells is 3.9% as compared to 12.6% CD123<sup>-</sup> cells difference of expression was factor >2 different and thus no points were given for c, sensitivity). Note that the majority of CD34<sup>+</sup>CD38<sup>-</sup> cells are CD123 negative (D) and therefore the majority of CD34<sup>+</sup>CD38<sup>-</sup> cells are presumed HSCs in this patient.

**IV. Examples of cell surface markers with 1 point (#1626, #1776)**

Two examples of markers that scored 1 point are shown. In patient #1626, CD11b (B) was the best marker (3 points). CD7 (A) scored 1 point based on criteria c, high sensitivity (percentage of CD7 positivity on CD34<sup>+</sup>CD38<sup>-</sup> was 74.9% as compared to 99.9% positivity of CD11b; thus factor ≤2 different). There was no clear discrimination between CD7<sup>-</sup> and CD7<sup>+</sup> cells (criterion a) and the CD7 negative fraction was >10% higher than for CD11b; therefore CD7 scored 1 point in total. CLL-1 in #1776 (C) scored 1 point based on criterion a, a clear separation between CLL-1<sup>-</sup> and CLL-1<sup>+</sup> cells. Sensitivity was not high since there was no useful CLL-1 expression as compared to CD7 (D), the best marker (14.4% CLL-1 positivity as compared to 50.0% positivity of CD7; difference >2 factor).

**V. Examples of cell surface markers with 0 points (#1631)**

In patient #1631, plots A-C show expression levels on CD34<sup>+</sup>CD38<sup>-</sup> of three different markers (CD7, CD14 and CD19) that all scored 0 points. In this patient CD44, a marker that is already highly expressed on HSCs, enabled a clear discrimination of HSCs (CD44<sup>+</sup>) and LSCs (CD44<sup>+</sup>).
FIGURE S2 | MFI values of CD34+CD38- HSCs, lymphocytes and the calculated MFI ratios
This figure shows results for the 8 different markers as shown in Table 2 whereby median MFI values, including 95% confidence intervals, of the CD34+CD38- HSCs and the relevant negative lymphocyte sub-populations are shown. The MFI ratios were calculated by dividing the CD34+CD38- HSC population through the MFI of the negative fraction of the lymphocytes. Median MFI ratios as shown in this figure therefore correspond to the ratios as shown in Table 2. Results are shown for the three different conditions: AML diagnosis BM (A), AML follow-up BM (B) and normal BM (C).
I. AML Diagnosis #1441

II. AML Relapse #1441

**FIGURE S3 | Efficacy of the marker cocktail in a relapse sample with an altered immunophenotype**

This figure shows both diagnosis (I) and relapse (II) of a patient (#1441) whereby the CD34<sup>+</sup>CD38<sup>-</sup> immunophenotype changed during disease. At time of diagnosis (I) all individual markers (A-F) have no or very little expression on CD34<sup>+</sup>CD38<sup>-</sup> cells. The marker cocktail, with a very small population (1%) of CD34<sup>+</sup>CD38<sup>-</sup> positive cells, shows the same result (G). At relapse (II) some markers still have no/little expression (A, E, F) whereas others are now expressed on the majority of CD34<sup>+</sup>CD38<sup>-</sup> cells (B-D), suggesting that the small subpopulation of positive CD34<sup>+</sup>CD38<sup>-</sup> cells, present at time of diagnosis (B-D in I), has grown out to relapse. The marker cocktail is also highly expressed on CD34<sup>+</sup>CD38<sup>-</sup> cells (G), indicating that despite marker instability, whereby at time of diagnosis it cannot be predicted which marker will gain expression towards relapse, the marker cocktail anyway identifies the fraction of leukemic CD34<sup>+</sup>CD38<sup>-</sup> cells.
FIGURE S4 | Efficacy of the marker cocktail in a relapsed patient with a stable immunophenotype
This figure shows both diagnosis (I) and relapse (II) of a patient (#1997) whereby the CD34^+CD38^− immunophenotype is relatively stable during disease. At time of diagnosis (I) different markers are expressed on CD34^+CD38^− cells (A-D). The marker cocktail shows an additive effect of combining the 6 different markers since expression of the cocktail (G) is significantly higher as compared to all individual markers (A-F). At relapse (II), expression of the single markers is quite similar as compared to diagnosis (A-F). The marker cocktail is also highly expressed on CD34^+CD38^− cells (G), again even higher as compared to all individual markers measured at time of relapse (A-F).