Despite intensive treatment regimens for AML patients roughly half of the patients who enter complete remission will experience a relapse and eventually die from the disease.\(^1\) Refinement of risk assessment, with the aim to re-direct patients to a more optimal treatment modality, is warranted. The aim of this thesis is to assess the role of minimal/measurable residual disease, and especially leukemia stem cells herein. The focus is on multiparameter flow cytometry (MFC).

Measurable residual disease (MRD) already has proven its prognostic value in prospective studies\(^2-5\) and currently, studies are ongoing in which MRD assessed after chemotherapy, enables to define the most appropriate consolidation therapy. In this thesis our focus is on further improvement of the MRD assay, whereby we explored peripheral blood (PB) as an alternative source for MRD.

It is known that leukemia stem cells (LSC) are at the basis of leukemia initiation and, moreover, may preferentially survive chemotherapy and subsequently grow out to MRD. In a further stage these therapy resistant LSCs may lead to disease relapse. The very low frequencies of these cells, especially after therapy, suggests that the LSC frequency may constitute a separate prognostic entity not only at overt disease but also within an MRD leukemic population. This thesis aimed to answer the question whether the CD34-CD38- LSC frequency would improve MRD detection and therefore might have a future role in clinical decision making in AML.

In **chapter one** we provide a general introduction regarding AML and its clinical features. Hereafter MRD and LSC are discussed as possible prognostic parameters in AML. Flow cytometric strategies, whereby aberrant immunophenotypes are used to define MRD and LSC frequencies, are illustrated and clinical applications of MRD and LSC are outlined.

In **chapter two** we made an inventory of the most prominent markers used to define aberrant expression as used in MRD detection. We aimed to solve controversies that exist about the prognostic role of the immunophenotype of the blasts at diagnosis. We showed that expression of several markers on the blasts cells at time of diagnosis is of limited independent prognostic value. Our data strongly suggest that associations between these immunophenotypic markers and molecular and cytogenetic aberrancies, with their well-known prognostic impact, largely explain the lack of independent prognostic impact. About 90% of AML cases are characterized by one or more of such immunophenotypic aberrancies, and since only 40%-50% of the AML cases are characterized by molecular and/or cytogenetic aberrancies, this shows that roughly half of the patients are characterized by immunophenotypic aberrancies that go along with known these molecular/cytogenetic aberrancies. As such, there is thus no clear bias towards particular prognostic sub-groups while using immunophenotypic MRD.

In **chapter three** the role of LSCs at time of diagnosis and follow-up was investigated in a retrospective study. With a limited number of LSC markers and aberrant flow cytometric scatter properties, CD34+CD38- LSCs could be detected in the majority of AML patients. However, the group of so-called CD34-negative AML patients (about 20% of the patient group), wherein only normal CD34+CD38- cells are present, where excluded in this retrospective study. In this relatively small group of CD34-positive AML patients, we for the first time showed that the CD34+CD38- LSC frequency had prognostic impact at diagnosis, but also at follow-up
(high LSC frequency: poor survival). Moreover, the combination of LSC with MRD revealed 4 sub-groups with different survival of which LSC<sub>low</sub>/MRD<sub>low</sub> had the best prognosis and LSC<sub>high</sub>/MRD<sub>high</sub> the poorest.

Since in chapter three the small subgroup of CD34-negative AML patients was excluded, **chapter four** aims to characterize the clinical properties of this patient group. Prognostic value of CD34 expression in AML has been extensively studied in the past, however results so far were contradictory and overall CD34 was not suggested to have prognostic importance. We used a new definition to define CD34 status whereby CD34-negativity is characterized by the complete absence of leukemic CD34+ blasts. This was proven using the aberrant immunophenotype approach at time of diagnosis and results were validated using the molecular aberrancies that characterized the rest of the blast population (e.g. NPM1<sup>mut</sup>). The thus-defined CD34-negative status turned out to be a prognostic factor at diagnosis: absence of neoplastic CD34+ cells was accompanied by a better prognosis. Consequently, the CD34+CD38- LSC frequency was zero percent in this CD34-negative patient group which was thus accompanied by a good prognosis.

Results of the retrospective studies as described in chapters three and four were prospectively validated in **chapter five**. In this study, wherein both CD34-negative and CD34-positive patients were included, we describe results of the first prospective study on the prognostic value of CD34+CD38-LSCs in AML patients under the age of 60. Flow cytometric results of 594 patients, included in the HOVON102 clinical trial, were available at time of diagnosis. We show the prognostic importance of LSC frequencies status at time of diagnosis: LSC frequency of zero (CD34-negative patients) had better survival than patients with LSC frequencies up to 0.03%, which in turn had better survival than patients with LSC frequencies higher than 0.03%. MRD and CD34+CD38-LSC results were available in 242 patients in morphological complete remission that received induction chemotherapy. We showed that the LSC frequency after therapy, similar to MRD, had prognostic value and even strongly improved prognostic impact of MRD: patients who have high MRD and high LSC levels after induction therapy hereby have a very poor outcome. Even in AML patients with a good or intermediate prognosis based on currently used risk schemes, high MRD and high LSC levels predict an adverse outcome. Multivariate analyses confirmed the independent prognostic value of the CD34+CD38-LSC status.

Since currently used LSC detection methods are quite complex, time consuming and needs large bone marrow samples, whereby often multiple tubes with multiple markers are necessary for specific LSC detection in every AML patient, we aimed in **chapter six** to design a specific 8-color LSC detection single tube. To that end we compared 15 commonly used markers for CD34+CD38-LSC detection for usefulness (stability, specificity compared to normal bone marrow, redundancy compared to other markers) in AML diagnosis and follow-up samples. Validation analyses showed that this one tube LSC approach, characterized by 7 single markers in 7 different fluorescence channels and a combination of 6 markers in the remaining fluorescence channel, gives similar LSC results as compared to our currently used multi tube approach (7 tubes) and, moreover, is suitable to use in a multi-institutional approach.
Although conventional MRD measurement is performed in BM samples of AML patients, *chapter seven* provides data concerning specificity and sensitivity differences between MRD detection in peripheral blood and bone marrow of 114 AML patients. In this study we show that median MRD percentage in PB is factor 4-5 lower as compared to BM MRD and thus, as expected, sensitivity of PB MRD is lower as compared to BM MRD. Interestingly, specificity of PB MRD was higher as compared to BM MRD due to lower background expression of the used aberrancies. Lower sensitivity in PB may therefore be counterbalanced by high specificity, in theory making PB as suitable for MRD detection as BM. High PB MRD levels (cut-off 0.04%) do predict an upcoming AML relapse, even when taken into account other prognostic parameters in a multivariate model (HR 2.94). Although these results have to be validated in larger studies, there may be a role for the very specific PB MRD sampling in future AML clinical trials.

Although flow cytometrically defined MRD and LSC levels appear to be of high importance in relapse prediction in AML patients, further improvement of flow cytometric strategies to further reduce the amount of false-negative MRD and LSC results is of high importance. Tumor heterogeneity and the related possible instability of leukemic cells is common in AML; this process of leukemic instability during disease is reviewed in *chapter eight*. Instability of the leukemic clone/population may be reflected by changes in flow cytometric, cytogenetic and/or molecular changes. We here review the occurrence of such changes during disease whereby disappearance of flow cytometric aberrancies during disease can consequently lead to false-negative flow cytometric MRD results. Approaches how to overcome such false-negative results due to disappearing leukemic populations are further discussed below in ‘future perspectives’.

**Future perspectives**

This thesis should contribute to an approach in AML where MRD and LSC will play a significant role in the improvement of relapse prediction and consequently, determine most optimal post-induction treatment strategy for each AML patient. Moreover, we propose that MRD and LSC results should be used to further refine current definitions of complete remission (CR). In current ELN AML recommendations CR<sub>MRD</sub> is already proposed as a response criterion. In future AML response criteria LSC results, as defined in CR patients, can further refine CR definition.

**Improvement of the MRD assay**

Nowadays, MRD is already implemented in different clinical trials whereby type of consolidation treatment is partly based on the MRD status after induction therapy. Further improvement of specificity of the MRD assay is of high importance to prevent patients with low relapse risk from unjustified toxic treatments due to a false-positive MRD result. In our current detection method we use leukemia associated immunophenotypes (LAIPs) as defined on blast cells, to detect MRD. These LAIPs often consist of a combination of three different markers: a primitive marker, a myeloid marker and an aberrant marker. There are several ways to improve
current performance of LAIP based MRD assessment: 1) further expansion of the number of fluorochromes in one flow cytometric tube (e.g. 18-color panels) would enable involvement of more markers in such a LAIP, e.g. exclusion markers, and thereby possibly improving specificity. 2) different LAIPs have different specificities (as illustrated in chapter 5) and we suggest to use different cut-off values for different LAIPs to define MRD positivity. Of note, for easy clinical applicability such a multi cut-off approach should be reduced to one final parameter for clinical use (e.g. positive/high or negative/low result) for instance combined with the LSC approach as described below. 3) false-negative results may in part be the result of emerging leukemic populations that may be present as (very) small subpopulations at time of AML diagnosis. Therefore, in future clinical studies, we aim to combine the MRD LAIP approach with the different from normal approach. In this latter approach aberrant maturation and/or differentiation profiles are used to detect cells that are different from normal hematopoietic cells. Usage of the different from normal approach could improve MRD detection especially in the cases were no LAIPs are present in the diagnosis BM. Moreover the different from normal approach enables detection of upcoming leukemic clones that were not present at time of diagnosis.\textsuperscript{8,9} This combined approach as a new integrated MRD approach has recently been advocated by the MRD AML Working Party of the European Leukemia Net (Schuurhuis et al., submitted).

To facilitate MRD assessment we showed in this thesis that PB MRD can be of additive value in future MRD determinations, and may ultimately at least partly replace BM MRD. To reach such goal, future studies should reveal if the lower sensitivity of PB MRD is outweighed by its higher specificity. For the moment, PB MRD may spare at least some AML patients from BM sampling, since high MRD levels in PB are also associated with high risk of relapse. This would be of particular value in patients where BM aspiration failed (eg dry tap, small amount of BM). It is of great importance to quickly assess the use of PB for MRD assessment since PB acquisitions are much less traumatic for patients and much easier to perform. Lastly, this would also allow for frequent MRD determination during long-term follow-up.

\textbf{Improvement of the LSC assay}

Further refinement in predicting an AML relapse was accomplished by incorporating the CD34+CD38- leukemia stem cell (LSC) frequency as defined in the same follow-up BM as MRD. Based on the results of our prospective study, it can be suggested that in future studies, clinical decision regarding type of consolidation treatment must also take into account the LSC status (LSC negative or positive) after induction therapy. More specifically, our results suggest that, independent on relapse risk based on pre-treatment prognostic factors, an allogeneic stem cell transplantation should be considered in case of positivity for both MRD and LSC after induction therapy.

However, LSC determination is challenging since CD34+CD38-LSCs are present in a very low frequency and false-negative LSC results still occur. As part of our ongoing efforts to further improve LSC detection several suggestions can be made: 1) flow cytometric analyses of many more white blood cells (WBCs) per tube are required as compared to MRD. We nowadays aim to measure up to 3-4 million white blood cells (WBCs) per LSC tube.
2) reduce the amount of tubes necessary for LSC detection. More than 1 tube is necessary for complete LSC assessment in our currently used 8-color flow cytometric approach. To overcome this disadvantage we have designed a future approach for LSC determination (chapter 6) whereby 13 different immunophenotypic markers are divided over 8 fluorochrome channels present in one tube. This enables an easy to use approach that allows to detect up to 10 million WBCs in a one tube approach. With such an approach total stem cell load including LSC from emerging populations during disease can be detected. Such an easy to use and less time consuming LSC assay significantly contributes to future multi-institutional studies wherein it is feasible to incorporate LSC detection as prognostic factor for relapse prediction. Although not used (yet) for clinical decision making, this LSC tube is used in follow-up BM In the currently ongoing Hovon132 study where it thus enables not only tracking the aberrancies present at diagnosis but also the detection of emerging LSC populations. However, still the combined use of LSC and MRD then requires 5 tubes (1 LSC tube and 4 MRD tubes), and therefore efforts to design a one tube assay for both LSC and MRD are ongoing. 3) although PB MRD is possible to facilitate MRD detection, PB LSC assessment is far from applicable since it would require too many cells (e.g. >20-40 million). 4) focus should be on both vanishing and emerging leukemic blast and LSC populations. One of our future aims is to find out if certain (small) subpopulations, present at time of diagnosis, are more capable of surviving chemotherapy treatment. Since often different leukemic subpopulations are present at time of diagnosis it would be the ultimate aim to be able to predict which subpopulations is/are likely to be chemotherapy resistant and cause(s) a relapse. The future combined one-tube MRD and LSC multi-color flow cytometric approach, should help to solve these questions. However, such a one-tube MRD and LSC approach is not yet ready to be implemented in daily clinical practice.

Incorporation of LSCs in the already more established MRD approach will significantly improve the ability to detect an upcoming relapse. Moreover such a combined MRD and LSC would enable the design of an MRD/LSC algorithm that accounts for the different LAIP specificities as mentioned above, implying that different (>2) prognostic subgroups can be detected (e.g. MRD ≤ 0.01%, MRD > 0.01% - ≤ 0.1%, and MRD>0.1%). This can then be combined with LSC defined sub-groups (≥ 2), with the aim to end up with 2 new sub-groups defining a good and poor performing group of patients. This final parameter (good vs. poor) can subsequently be used in daily clinical practice.

Importantly, within CD34-negative AML patients (around 20% of all AML patients) efforts are ongoing to find cell populations enriched with stem cell properties that would enable LSC tracking also in these patients with only normal CD34+CD38- stem cells at time of diagnosis. For the moment the CD34-negative patient group as a whole can be defined as a good performing sub-group with CD34+CD38- LSC frequency of zero. Our ongoing research suggests that CD133 plays a crucial role in determining prognosis within these CD34-negative patients.
Other techniques to assess residual disease

Already well established for MRD and used in clinical settings, is the use of polymerase chain reaction (PCR) to detect residual cells via the presence of molecular aberrations. A combination of flow cytometrically detected MRD/LSC and molecular MRD is expected to significantly improve relapse prediction in AML. A comparative study between residual disease as defined via presence of the NPM1 mutation and flow cytometrically defined MRD is in progress (analyses ongoing). These results will reveal the possible added value in relapse prediction when combining molecular and flowcytometric techniques. Nevertheless, a disadvantage is that such a combination would only be possible in up to 40%-50% of the AML patients since molecular aberrancies suitable for MRD assessment are present in approximately half of the patients.

Next generation sequencing (NGS) is a second upcoming molecular approach that enabled the expansion of the panel of leukemia associated recurrent gene mutations like IDH1/IDH2 and DNMT3A. Theoretically NGS can be used for molecular MRD detection using all available leukemia associated genetic mutations and this technique is supposed to be highly sensitive for detection of residual mutations. Another flow cytometrically based assay that can be used for residual disease detection is cyTOF. This technique encompasses the use of antibodies labeled with heavy-metal ion. This technique is promising in detection of aberrant populations since it enables the usage of large panels of markers (e.g. 30-40 markers). Future studies should elucidate the exact role of these above mentioned upcoming techniques in MRD/LSC detection.

Overall, based on the results of this thesis, future flow cytometric studies should pursue a combined MRD/LSC approach whereby the focus is on both vanishing and upcoming leukemic populations. We propose to not only include MRD, but also CD34+CD38-LSC frequency in future AML risk schemes.
References