MEASURABLE RESIDUAL DISEASE AND LEUKEMIC STEM CELLS IN ACUTE MYELOID LEUKEMIA

Hanekamp, Dirkje Wilhelmina

2021

document version
Publisher's PDF, also known as Version of record

MEASURABLE RESIDUAL DISEASE AND LEUKEMIC STEM CELLS IN ACUTE MYELOID LEUKEMIA

Diana Hanekamp

MEASURABLE RESIDUAL DISEASE AND LEUKEMIC STEM CELLS IN ACUTE MYELOID LEUKEMIA

Dirkje Wilhelmina Hanekamp

ISBN: 978-94-93197-71-8
Cover, layout and printing: Off Page, Amsterdam

Financial support by a grant from the KWF/ Alpe D'HuZes, for the work within this thesis is gratefully acknowledged.

The printing of this thesis was additionally kindly supported by: ChipSoft, Pfizer, Bristol-Meyer Squibb, Merus and Servier.

VRIJE UNIVERSITEIT

MEASURABLE RESIDUAL DISEASE AND LEUKEMIC STEM CELLS IN ACUTE MYELOID LEUKEMIA

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op woensdag 29 september 2021 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Dirkje Wilhelmina Hanekamp
geboren te Kampen

promotor: prof.dr. J. Cloos

copromotoren: prof.dr. G.J. Ossenkoppele
dr. J.J.W.M. Janssen

Voor mijn ouders

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

LEUKEMIA

Hematopoiesis, the production of all the cellular components of blood and blood plasma, occurs within the hematopoietic system, which includes the liver, spleen and the bone marrow (BM). At the origin of hematopoiesis are the so-called hematopoietic stem cells (HSCs). Normal hematopoiesis, in healthy persons, results in a wide variety of distinct differentiated cells with different functionalities. In leukemia cells, the processes of differentiation and maturation are blocked, and together with their acquired proliferative capacity eventually lead to overcrowding of the BM and consequently the suppression of normal functional blood cells. Leukemias are classified as either myeloid or lymphoid, depending on which types of white blood cells are affected.

ACUTE MYELOID LEUKEMIA

Patients with acute myeloid leukemia (AML) will initially present with symptoms due to reduction of normal hematopoiesis, such as anemia, easy bleedings and infections. The diagnosis is made by the presence of an abnormal high percentage (³20%) of immature myeloid blood cells (blasts) in the peripheral blood or BM, or through presence of unique genetic abnormalities, regardless of blast count (i.e. t(8;21), inv(16) or t(15;17)).[1] Although AML is rare in children, the incidence in infancy shows a relative peak at approximately 2 cases per 100,000, but rates rapidly decline subsequently with a nadir around 9 years, after which the incidence of AML increases gradually with age.[2] The average estimated incidence of AML is 3.7 cases per 100,000 persons per year,[3] increasing to ~5.0-40.0 per 100,000 persons per year for people over 60 years.[2] AML can be divided into various subgroups by integration of morphology (which is also at the basis for the outdated French-American-British (FAB) classification), immunophenotypic, cytogenetic and molecular analysis.[1] Based on both cytogenetics and recognized molecular subsets that are distinct from the contribution of cytogenetic risk,[4,5] patients are further classified in three prognostic risk groups according to the European Leukemia-Net (ELN) 2017 recommendations (i.e. favorable, intermediate and adverse). The prognostic groups predict the response to treatment and overall outcome.

TREATMENT

The treatment of AML usually needs to start as soon as possible after diagnosis (or once patient specific targets are determined for targeted therapies), since AML can progress rapidly. Typically, the treatment of AML involves initial induction chemotherapy and post-remission therapy. The primary goal of induction chemotherapy is to achieve complete remission (CR). By definition, CR is reached when morphological examination of BM reveals no more than 5% blast cells, with evidence of normal maturation of all cellular components

(normal erythropoiesis, granulopoiesis and megakaryopoiesis).[6] The choice of induction treatment depends on the age, the performance status of the patient and comorbidities (assessed by the World Health Organization (WHO)/Eastern Cooperative Oncology Group (ECOG) performance status and comorbidities)[7]. In general, the initial treatment entails cytotoxic chemotherapy, with or without targeted therapies.[8] The commonly used cytotoxic induction therapy (as used by the Hemato Oncology Foundation for Adults in the Netherlands, HOVON) consists of two cycles of chemotherapy containing cytarabine and an anthracyclin (idarubicine, daunorubicine or mitoxantrone). Achievement of CR sets stage for post-remission therapy to maintain the hematological remission of the disease, and to eliminate the more drug-resistant leukemia (stem) cells and to ultimately lead to cure.[9] The post-remission (consolidation) therapy can consist of either one or more additional courses of chemotherapy, an autologous stem cell transplantation (autoHSCT) or an allogeneic stem cell transplantation (alloHSCT). For autoHSCT, during remission, the patient's own (stem) cells are collected from his or her BM or blood and cryopreserved. After high-dose chemotherapy the blood-forming cells are reinfused to restore the BM. AlloHSCT is the most common type of SCT in the treatment of AML. In alloHSCT, the BM stem cells of the patient are replaced by the stem cells of a donor, preferably a sibling, to ensure the closest HLA-match. The closer the match between the donor and recipient, the better chance the transplanted cells will engraft and start repopulating the blood and BM, and the lower the risk of graft-versus-host disease GVHD). AlloHSCT has the highest anti-leukemia effect, however, this GVHD is associated with morbidity and mortality. Since this balance between anti-leukemic effect and therapy- (or transplant) related morbidity is delicate, strict selection of consolidation therapy is necessary. In general, only patients with the highest risk of relapse are considered candidates for alloHSCT.

TREATMENT OUTCOME

For AML patients below 60 years of age, median overall survival (OS) after 3 years is 64% (59-68%) in favorable risk vs. 31% (25-38%) in intermediate risk vs. 19% (14-24%) in adverse risk groups.[10] OS decreases when older patients are included, but the stratification of survival per age group remains constant.[11]

Today, CR is achieved in the majority of AML patients: estimated between 70-90%.[12] However, approximately 50% of these patients will experience a relapse, resulting in dismal outcomes (especially in patients with intermediate or adverse risk). Although there is a major contribution to disease outcome of pre-treatment patient-related and disease-related factors, multiple post-diagnosis factors are linked to treatment failure, such as pharmacokinetic and -dynamic variables.[13] Together, these factors eventually determine whether the AML will be eradicated.

Previous work has shown that at CR, residual leukemic cells may be present at various levels, with higher levels consistently related to a higher relapse risk.[14] This measurable/

minimal residual disease (MRD) is thought to be responsible for the development of relapse (see **Figure 1**). The European Leukemia Net recommendations have therefore incorporated the evaluation of treatment response by assessing measurable/minimal residual disease (MRD) after induction chemotherapy.[15]

MEASURABLE RESIDUAL DISEASE DETECTION IN ACUTE MYELOID LEUKEMIA

Assessment of MRD after induction chemotherapy using more sensitive techniques, allows for the detection and quantification of lower levels of residual leukemia that cannot be assessed by morphology. In this chapter, commonly used modalities for MRD detection (immunophenotypic) multiparameter flow cytometry (MFC), reverse-transcription quantitative polymerase chain reaction (RT-qPCR) and next generation sequencing (NGS)) are discussed.

IMMUNOPHENOTYPIC MULTIPARAMETER FLOW CYTOMETRY MRD DETECTION

In routine morphology, assessment of BM or peripheral blood (PB) relies on the examination of a limited number of cells (200-500). Its reliability, in part, depends on sample quality and hematologists'/pathologists' expertise.[17] The introduction of flow cytometry in routine diagnostics allowed more detailed and more sensitive examination of BM and PB for diagnosis of AML. Since most laboratories have gained experience in MFC analysis of AML, it is one of the most frequently used techniques to assess MRD in AML. MFC MRD detection relies on the presence of immunophenotypic aberrant antigen expression. Most laboratories rely on the identification of aberrant cell surface antigen expression.

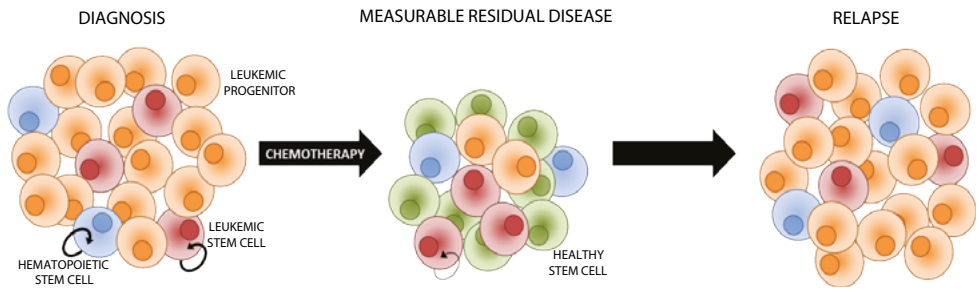


Figure 1. Role of MRD in relapse development. At diagnosis, AML consist of a heterogenous population of leukemic cells and (leukemic) stem cells (LSC). Chemotherapeutic treatment often results in microscopic complete remission (CR). However, small numbers of residual leukemic cells survive therapy and are present within CR bone marrow, termed measurable residual disease (MRD). MRD contains LSC that have the capacity to initiate relapse development. Adapted from Canales et al.[16]

At diagnosis, leukemic cells can be discriminated from normal blood cells using so-called leukemia associated immunophenotypes (LAIP). LAIPs are identified on blast cells (characterized by dim CD45 expression) and are a combination of a myeloid marker (CD13/CD33), a normal progenitor antigen (CD34, CD117 and/or CD133) and (multiple) aberrantly expressed cell surface marker(s). A multitude of LAIPs is recognized,[18] but generally, LAIPs can be divided in four groups: (1) cross-lineage antigen expression (e.g. myeloid blasts with simultaneous expression of lymphoid markers; $CD34^{+}CD13^{+}CD7^{+}$, $CD34^{+}CD13^{+}CD56^{+}$), (2) asynchronous antigen expression (i.e. combination of immature and mature differentiation antigens), (3) lack of antigen expression and (4) overexpression of antigens[19]. The heterogeneity of AML necessitates a broad panel of antibodies, that allows identification of LAIP per individual patient. Such aberrancies can subsequently be used to detect MRD (examples in **Figure 2**).

In MFC MRD, two approaches are known: (1) the LAIP approach; which defines LAIPs at diagnosis which presence is subsequently monitored in follow-up and (2) the different-from-normal (DfN) approach; which screens follow-up material on presence of LAIPs (hence, presence of leukemic cells) in follow-up, irrespective of diagnosis material.[14]

MFC MRD is applicable in a high percentage of AML patients. For the LAIP approach, reported percentages range between 80-95%.[20–23] Expression of LAIPs in normal BM or PB is absent, or at least very low. However, for accurate MRD detection, knowledge of antigen patterns on normal BM or PB cells is needed. Sensitivity of MFC MRD is reported in a range of 10^{-3} to 10^{-5} (i.e. 1 leukemic cell in 1000 to 100,000 cells).[20–22,24,25] The limiting factor for the sensitivity of the approach is the presence of normal cells expressing LAIPs (even when this is at very low percentages). For this, most laboratories make use of a cut-off that aids correct discrimination between true MRD positive patients, and background LAIP expression.[14] Cut-off levels differ between laboratories but 0.1% was proposed by the ELN MRD Working Party[14]; when the cut-off is too high, sensitivity is compromised leading to false negatives, whereas a too low cut-off could classify normal background expression (which could be augmented in regenerating BM after intense chemotherapy, or new therapeutics) as MRD. With the notion that AML is highly heterogeneous and clonal, and that LAIPs at diagnosis are not always present at relapse, the LAIP approach has a possible pitfall.[26–29] Tracking LAIPs identified at diagnosis in follow-up material, immunophenotypic changes (evolution or selection of specific clones under selective pressure of the therapy) can be missed, leading to false MRD-negativity. The DfN approach is capable of identifying these immunophenotypic changes, but due to lower specificity might result in false positivity.[30]

MOLECULAR MRD DETECTION

According to the 2017 ELN recommendations, screening of mutations in NPM1, CEBPA, RUNX1, FLT3, TP53 and ASXL1, together with assessment of fusion genes, is the minimum for accurate risk classification.[15] Similar assays, for at least part of the mutations can be

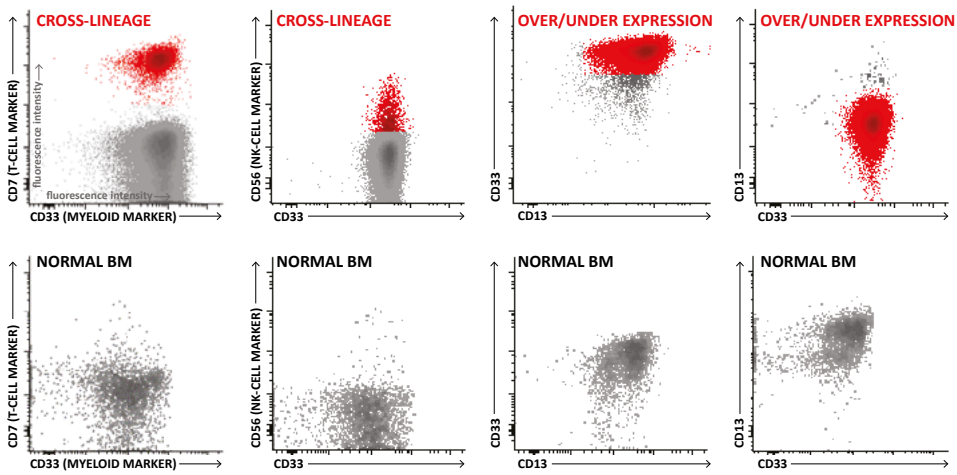


Figure 2. Four examples of leukemia associated immunophenotypes at diagnosis. Leukemic cells can be identified using aberrant cell surface antigen expression. At diagnosis, leukemic cells can be discriminated from normal blood cells using so-called leukemia associated immunophenotypes (LAIP). LAIPs are identified on $CD45^{dim}$ blast cells, and are a combination of a myeloid marker (CD13/CD33), a normal progenitor antigen (CD34, CD133 and/or CD133) and (multiple) aberrantly expressed cell surface marker(s). Shown above are $CD45^{dim}CD34^{+}$ cells from four different AML patients (above) and from normal bone marrow (below). The leukemic subpopulations, identified by LAIPs, are shown in red.

used for molecular MRD detection. Common targets for molecular MRD detection are gene mutations (NPM1[31–33] and FLT3-ITD[34]), over-expression (WT1[35]) and fusion genes (AML1-ETO[36], CBFb-MYH11[37]) and can be best detected using RT-qPCR.

Until recently, RT-qPCR was the cornerstone of molecular MRD detection, which can be used when a suitable target is present. RT-qPCR sensitivities were reported between 10^{-4} to 10^{-6} (i.e. 1 leukemic cell in 10,000 to 1 cell in 1,000,000 cells).[37–39] The best example of RT-qPCR MRD detection is for NPM1. Mutations in NPM1 are specific and stable during the disease course[39,40], allowing sensitive detection of presence of MRD. However, RT-qPCR based MRD detection is limited to patients harboring a specific mutation. Furthermore, in contrast to NPM1, some aberrancies are unstable during disease (e.g. FLT3-ITDs[39,40]) and less qualified for MRD detection. MRD detection using overexpression (e.g. WT1) is restricted to cut-offs with backgrounds variable during therapy, and thereby also less suitable for MRD.[41,42] Next-generation sequencing (NGS), which allows detection of all relevant recurrent mutations in AML is becoming an important tool in the diagnostic AML workup, as well as for the molecular detection of MRD, but sensitivity issues still need to be solved.[43–45]

1

CLINICAL APPLICATION OF MRD

As stated above, assessment of MRD in CR is of prognostic importance. In addition to known risk factors (e.g. age, number of cycles to achieve CR, (cyto)genetics), assessment of MRD could aid in the decision what consolidation treatment to give. The prognostic impact of MRD is shown by studies that correlated MRD to outcome parameters as cumulative incidence of relapse (CIR), leukemia-free survival (LFS) or OS.[14,18,46] In most studies, MRD is regarded as binary outcome (i.e. $MRD^{negative}$ vs. $MRD^{positive}$/MRD^{low} vs. MRD^{high}). For MFC-MRD, where the outcome of percentage LAIP/WBC is continuous, thresholds are needed. Different laboratories use different optimal cut-offs values after initial chemotherapy, ranging between 0.035-0.2%.[14,47] Since these cut-offs all gave statistically significant differences in outcomes, it is evident that there is not a single best cut-off, but a range of cut-offs that significantly predict clinical outcome. For molecular methods, sensitivity is limited to reduction of MRD target below the lower limit of detection. Following this, the recommended method of MRD assessment in patients with suitable targets is RT-qPCR, which reaches up to 100- to 1000-fold greater sensitivity than MFC or NGS.[14,32] In patients without suitable mutations, MFC-MRD detection can identify patients with high risk of relapse.[14]

In 2016-2017, 69% of the surveyed leukemia treating physicians in the United States reported routine use of MRD in management of AML, in 90% to guide therapy.[48] The HOVON/SAKK cooperative group supports the implementation of rapidly available MRD detection for clinical decision making for patients within current and future HOVON protocols. In the HOVON132 study, risk categorization at diagnosis is provisional, and detection of MRD in intermediate and poor risk reclassifies patients in either poor risk or very poor risk respectively. Especially for the intermediate risk patients, measurement of MRD has effect on post-remission treatment choice; intermediate risk patients without MRD would be assigned to autoHSCT, whereas patients with MRD would be assigned to alloHSCT (**Figure 3**).

Still, although MRD status has strong prognostic value, about 30% of MRD negative patients develop a relapse. Several initiatives are therefore currently ongoing to improve the predictive value of MRD for the individual patient. One such effort is the detection of the relapse initiating cells that may be present in low levels at MRD, below the positivity threshold, but that are still capable of repopulating a relapse. The cells that have features like self-renewal, high level detoxification, proficient DNA repair and very immature (non-functional) phenotype (e.g. stem cells) are hypothesized to be the most relevant relapse initiating cells.

LEUKEMIC STEM CELLS

In normal hematopoiesis, stemness features allow normal hematopoietic stem cells (HSCs) to self-renew to support the long-term process of cell differentiation. The organization of

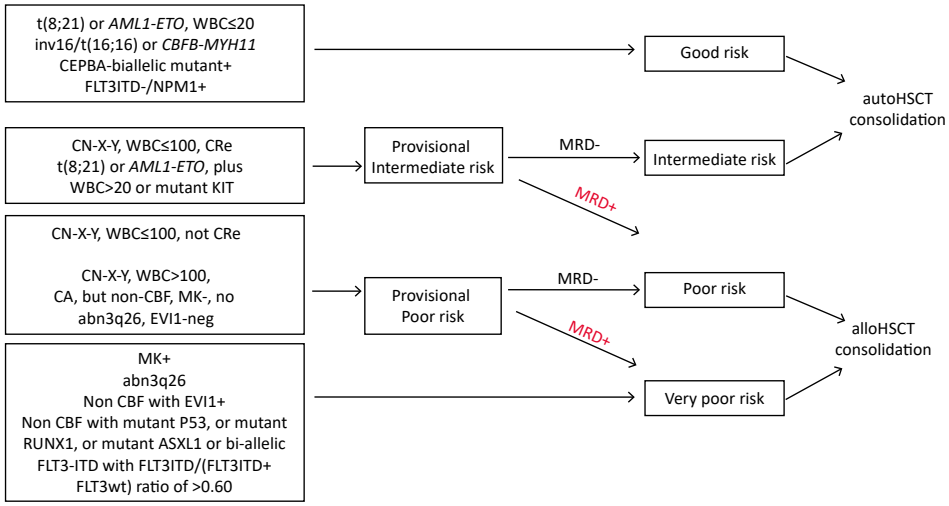


Figure 3. Risk classification of patients with AML (as used in HOVON132). Risk classification scheme based on features at diagnosis and the presence of MRD. Presence or absence of certain features or mutations at diagnosis stratify patients in either good, intermediate, poor or very poor risk category. Good risk patients are scheduled for an autologous stem cell transplantation. Patients within the poor and very poor groups are scheduled for an allogeneic stem cell transplantation. Presence of MRD in patients with provisional intermediate risk are moved to poor, and thus scheduled for a more intense alloHSCT. MRD+; MRD positive after cycle II either by flow cytometry or molecular(NPM1) CBF, core binding factor; MK, monosomal karyotype CRe – early CR (i.e. CR after cycle I) High EVI1 expression is defined as EVI1 expression above 0.1x EVI1 expression in the cell line SKOV3 (reference gene normalized (Groschel et al., JCO 2010, 12 (28) p. 2101-07)) MRD is considered positive whenever residual disease is demonstrated by any assay, whether it is by flow cytometry or molecular analysis (i.e. NPM1mutant).

AML is similarly believed to be clonal and hierarchical, originating from mutations that occur in a multipotent progenitor or a HSC.[49–51] The concept that leukemia originates from cells, distinct from the bulk of tumor cells, originates from decades ago.[52,53] Initial proof was found in experiments demonstrating the presence of leukemic cells showing stemness features as drug resistance, self-renewal and an undifferentiated state.[52,54] Cells that met these criteria, and, moreover, were capable to generate leukemia in immunodeficient mice[55,56] were identified within the $CD34^+CD38^-$ compartment, that also contains the HSCs.

Within the $CD34^+CD38^-$ compartment (with the population frequency of 1:5000 cells) the discrimination between LSCs and HSCs is based on the principle that HSCs only express antigens from the myeloid lineage at a certain level. Aberrant antigen expression on LSCs is heterogenous, but include, amongst others: CD2, CD7, CD11b, CD22, CD33, CD44, CD45RA, CD56, CD123, CD366 (TIM3) and CD371 (CLEC12A).[57–59]

Today, the existence of LSCs is more commonly accepted, and further research has revealed that LSCs can also be present (at lower frequencies) in $CD34^+CD38^+$ or $CD34^-$ immunophenotypes.[60–63] Extensive research into the specifics and differences between

these types of LSCs is lacking, as well as the existence of possible competition between these compartments in leukemogenesis. However, it was shown that the $CD34^+CD38^-$ compartment is the least immunogenic compared to other compartments[64], and therefore could possibly best engraft in NOD/SCID mice which have little residual immunity.[52] The frequency of these $CD34^+CD38^-$ LSCs at diagnosis and after chemotherapy treatment is an important prognostic factor for OS and relapse.[65,66] Whereas LSCs are also undoubtedly present in CD34- cells in patients with CD34-negative AML , no markers are known for the identification of these cells.[67] Flow cytometric assessment of LSC frequency at diagnosis, and after chemotherapy is therefore focused on the $CD34^+CD38^-$ compartment.[19]

MRD AND LSC AS SURROGATE ENDPOINT FOR NEW THERAPEUTICS IN AML

Currently, OS is most frequently used as endpoint to evaluate clinical benefit of new treatments in AML. As a consequence, this requires a long follow-up time and introduces confounding factors as post-remission therapies, which differ among patients. Using MRD as an endpoint has the potential to more quickly demonstrate the potential benefit of a given therapy. In 2018, the U.S. Food and Drug Administration (FDA) published a draft guidance for the use of MRD as assessment of treatment responses, in regulatory submissions or to support marketing approval of new therapeutics for hematologic malignancies[68]. Although MRD in CR has major prognostic implications at the level of groups of AML patients, its predictive value for individual patient treatment response and outcome remains to be established. This thesis tries to shed more light on this issue.

INTRODUCTION TO THE CHAPTERS

The overall aim of the studies described in this thesis is to investigate the role of MRD and LSC, and several initiatives to improve the MRD assessment to be used for relapse prediction for the individual patient.

Chapter 2 covers a review on several aspects of LSCs in AML and its considered role in relapse progression. Moreover, it discusses how these relatively rare cells can be detected by flow cytometry, and furthermore discusses how this detection is currently used in clinical application.

In chapter 3-4 we investigated if the LSC frequency harbors prognostic information for improved relapse prediction for AML. In chapter 3 we present the clinical significance of the presence and frequency of CD34+CD38- LSCs at time of diagnosis and in remission bone marrow in adult AML. In addition, the prognostic relevance of the combination of LSC-MRD and MFC-MRD is investigated. In chapter 4 we investigated whether detection of CD34+CD38- LSCs in BM of newly diagnosed pediatric AML bears similar prognostic relevance as shown in adult AML.

In chapter 5-6 we elaborate on the importance of standardization of the flow cytometric MRD and LSC detection approaches. In **chapter 5** we evaluated the technical and analytical feasibility of the previously designed eight-color LSC single tube assay, as well as standardization of the process. In **chapter 6** we present a new flow cytometric model for standardized and objective MRD calculation, retrospectively applied in a large clinical study. For this, we evaluate if the balance between neoplastic and normal progenitors in CR bone marrow has prognostic relevance.

In **chapter 7** we evaluate whether next-generation sequencing has clinical value for the prediction of relapse. Since measurements were simultaneously evaluated for MFC-MRD, we investigated whether NGS and MFC-MRD have independent and additive prognostic value.

In addition, we studied whether MRD and LSC-MRD is a valid? surrogate endpoint in AML. As shown in a recent clinical trial, the new therapeutic clofarabine has clinical beneficial effect in a subgroup of patients. In **chapter 8** we investigated whether the prospectively defined MRD and LSC-MRD frequencies were different between patients with clofarabine and patients without clofarabine, and whether MRD levels mirrored the clinical outcome within this subgroup.

Finally, in **chapter 9** we summarize the results of this thesis and which implications these results may have for future AML relapse prediction. Furthermore, we evaluate the different techniques used in this thesis, discuss how each technique can be further optimized and elaborate on the optimal use for future clinical trials.

1

REFERENCES

1. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).

2. Ursula *et al.* Acute myelogenous leukemia in adolescents and young adults. *Pediatr. Blood Cancer* **65**, e27089 (2018).

3. Visser, O. *et al.* Incidence, survival and prevalence of myeloid malignancies in Europe. *Eur. J. Cancer* **48**, 3257–3266 (2012).

4. Grimwade, D., Ivey, A. & Huntly, B. J. P. P. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood* **127**, 29–41 (2015).

5. Grimwade, D. *et al.* The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322–33 (1998).

6. Larson, R. A. Remission criteria in acute myeloid leukemia and monitoring for residual disease. *UpToDate* 1–5 (2019).

7. Østgård, L. S. G. *et al.* Comorbidity and performance status in acute myeloid leukemia patients: a nation-wide population-based cohort study. *Leukemia* **29**, 548–555 (2015).

8. Walter, R. B. & Estey, E. H. Selection of initial therapy for newly-diagnosed adult acute myeloid leukemia: Limitations of predictive models. *Blood Reviews* 100679 (2020). doi:10.1016/j.blre.2020.100679

9. Derman, B. A. & Larson, R. A. Post-remission therapy in acute myeloid leukemia: Are we ready for an individualized approach? *Best Practice and Research: Clinical Haematology* **32**, 101102 (2019).

10. Bill, M. *et al.* Prognostic impact of the 2017 European LeukemiaNet (ELN) classification of younger adult patients with acute myeloid leukemia (AML) treated on cancer and leukemia group B (CALGB)/ALLIANCE protocols. *HemaSphere* **3**, 68–69 (2019).

11. Hansen, D. K. *et al.* ELN 2017 Risk Classification Predicts Survival of AML Patients Receiving Allogeneic Hematopoietic Stem Cell Transplantation. *Biol. Blood Marrow Transplant.* **25**, S112 (2019).

12. Versluis, J. *et al.* Acute myeloid leukemia in adults. in *The EBMT Handbook: Hematopoietic Stem Cell Transplantation and Cellular Therapies* 507–521 (Springer International Publishing, 2018). doi:10.1007/978-3-030-02278-5_69

13. Ossenkoppele, G. J. & Schuurhuis, G. J. MRD in AML: It is time to change the definition of remission. *Best Pract. Res. Clin. Haematol.* **27**, 265–271 (2014).

14. Schuurhuis, G. J. *et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* **131**, 1275–1291 (2018).

15. Döhner, H. *et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017).

16. Canales, T. M., de Leeuw, D. C., Vermue, E., Ossenkoppele, G. J. & Smit, L. Specific depletion of leukemic stem cells: Can microRNAs make the difference? *Cancers* **9**, (2017).

17. DeAngelo, D. J., Stein, E. M. & Ravandi, F. Evolving Therapies in Acute Myeloid Leukemia: Progress at Last? *Am. Soc. Clin. Oncol. Educ. book. Am. Soc. Clin. Oncol. Annu. Meet.* **35**, e302-12 (2016).

18. Terwijn, M. *et al.* High Prognostic Impact of Flow Cytometric Minimal Residual Disease Detection in Acute Myeloid Leukemia: Data From the HOVON/SAKK AML 42A Study. *J. Clin. Oncol.* **31**, 3889–3897 (2013).

19. Zeijlemaker, W., Kelder, A., Cloos, J. & Schuurhuis, G. J. Immunophenotypic Detection of Measurable Residual (Stem Cell) Disease Using LAIP Approach in Acute Myeloid Leukemia. *Curr. Protoc. Cytom.* **91**, (2019).

20. Buccisano, F. *et al.* Monitoring of minimal residual disease in acute myeloid leukemia. *Current Opinion in Oncology* **21**, 582–588 (2009).

21. San Miguel, J. F. *et al.* Immunophenotyping investigation of minimal residual disease is a useful approach for predicting relapse in acute myeloid leukemia patients. *Blood* **90**, 2465–70 (1997).

22. Kern, W. *et al.* Determination of relapse risk based on assessment of minimal residual disease during complete remission by multiparameter flow cytometry in unselected patients with acute myeloid leukemia. *Assessment* **104**, 3078–3085 (2004).

23. Al-Mawali, A., Gillis, D. & Lewis, I. The use of receiver operating characteristic analysis for detection of minimal residual disease using five-color multiparameter flow cytometry in acute myeloid leukemia identifies patients with high risk of relapse. *Cytom. Part B - Clin. Cytom.* **76**, 91–101 (2009).

24. Venditti, a *et al.* Level of minimal residual disease after consolidation therapy predicts outcome in acute myeloid leukemia. *Blood* **96**, 3948–3952 (2000).

25. Feller, N. *et al.* MRD parameters using immunophenotypic detection methods are highly reliable in predicting survival in acute myeloid leukaemia. *Leukemia* **18**, 1380–1390 (2004).

26. Langebrake, C. *et al.* Immunophenotypic differences between diagnosis and relapse in childhood AML: Implications for MRD monitoring. *Cytom. Part B - Clin. Cytom.* **63**, 1–9 (2005).

27. Zeijlemaker, W., Gratama, J. W. & Schuurhuis, G. J. Tumor heterogeneity makes AML a "moving target" for detection of residual disease. *Cytom. Part B Clin. Cytom.* **86**, 3–14 (2014).

28. Baer, M. R. *et al.* High frequency of immunophenotype changes in acute myeloid leukemia at relapse: Implications for residual disease detection (Cancer and Leukemia Group B Study 8361). *Blood* **97**, 3574–3580 (2001).

29. Macedo, A. *et al.* Phenotypic changes in acute myeloid leukaemia: implications in the detection of minimal residual disease. *J. Clin. Pathol.* **49**, 15–8 (1996).

30. Wood, B. L. Acute Myeloid Leukemia Minimal Residual Disease Detection: The Difference from Normal Approach. *Curr. Protoc. Cytom.* **93**, (2020).

31. Schnittger, S. *et al.* Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML. *Blood* **114**, 2220–2231 (2009).

32. Dillon, R. *et al.* Molecular MRD status and outcome after transplantation in NPM1-mutated AML. *Blood* **135**, 680–688 (2020).

33. Gorello, P. *et al.* Quantitative assessment of minimal residual disease in acute myeloid leukemia carrying nucleophosmin (NPM1) gene mutations. *Leukemia* **20**, 1103–1108 (2006).

34. Blätte, T. J. *et al.* getITD for FLT3-ITD-based MRD monitoring in AML. *Leukemia* **33**, 2535–2539 (2019).

35. Cilloni, D. *et al.* Real-time quantitative polymerase chain reaction detection of minimal residual disease by standardized WT1 assay to enhance risk stratification in acute myeloid leukemia: A European LeukemiaNet Study. *J. Clin. Oncol.* **27**, 5195–5201 (2009).

36. Zhang, L. *et al.* Monitoring of minimal residual disease in acute myeloid leukemia with t(8;21)(q22;q22). *Int. J. Hematol.* **97**, 786–792 (2013).

37. Guerrasio, A. *et al.* Assessment of minimal residual disease (MRD) in CBFbeta/MYH11-positive acute myeloid leukemias by qualitative and quantitative RT-PCR amplification of fusion transcripts. *Leukemia* **16**, 1176–1181 (2002).

38. Sugimoto, T. *et al.* Quantitation of minimal residual disease in t(8;21)-positive acute myelogenous leukemia patients using real-time quantitative RT-PCR. *Am. J. Hematol.* **64**, 101–106 (2000).

39. Ivey, A. *et al.* Assessment of Minimal Residual Disease in Standard-Risk AML. *N. Engl. J. Med.* **374**, 422–433 (2016).

40. Bachas, C. *et al.* High-frequency type I/II mutational shifts between diagnosis and relapse are associated with outcome in pediatric AML: Implications for personalized medicine. *Blood* **116**, 2752–2758 (2010).

41. Selim, A. G. & Moore, A. S. REVIEW Molecular Minimal Residual Disease Monitoring in Acute Myeloid Leukemia Challenges and Future Directions. *J. Mol. Diagnostics* **20**, (2018).

42. Voso, M. T. *et al.* MRD in AML: The Role of New Techniques. *Frontiers in Oncology* **9**, (2019).

43. Getta, B. M. *et al.* Multicolor Flow Cytometry and Multi-Gene Next Generation Sequencing Are Complimentary and Highly Predictive for Relapse in Acute Myeloid Leukemia Following Allogeneic Hematopoietic Stem Cell Transplant. *Present. 58th Am. Soc. Hematol. Annu. Meet. Expo.* **Dec**, San Diego, CA (2016).

44. Bibault, J.-E. *et al.* Next-generation sequencing of FLT3 internal tandem duplications for minimal residual disease monitoring in acute myeloid leukemia. *Oncotarget* **6**, (2015).

45. Salipante, S. J., Fromm, J. R., Shendure, J., Wood, B. L. & Wu, D. Detection of minimal residual disease in NPM1-mutated acute myeloid leukemia by next-generation sequencing. *Mod Pathol.* **27**, 1438–1446 (2014).

46. Ravandi, F., Walter, R. B. & Freeman, S. D. Evaluating measurable residual disease in acute myeloid leukemia. *Blood Adv.* **2**, 1356–1366 (2018).

47. Ossenkoppele, G., Schuurhuis, G. J., van de Loosdrecht, A. & Cloos, J. Can we incorporate MRD assessment into clinical practice in AML? *Best Pract. Res. Clin. Haematol.* **32**, 186–191 (2019).

48. Epstein-Peterson, Z. D., Devlin, S. M., Stein, E. M., Estey, E. & Tallman, M. S. Widespread use of measurable residual disease in acute myeloid leukemia practice. *Leuk. Res.* **67**, 92–98 (2018).

49. Krivtsov, A. V. *et al*. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).

50. Shlush, L. I. *et al*. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104–108 (2017).

51. Thomas, D. & Majeti, R. Biology and relevance of human acute myeloid leukemia stem cells. *Blood* **129**, 1577–1585 (2017).

52. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737 (1997).

53. McCulloch, E. A. Stem cells in normal and leukemic hemopoiesis (Henry Stratton Lecture, 1982). *Blood* **62**, 1–13 (1983).

54. Gupta, P. B., Chaffer, C. L. & Weinberg, R. A. Cancer stem cells: mirage or reality? *Nat. Med.* **15**, 1010–1012 (2009).

55. Vormoor, J. *et al*. Immature human cord blood progenitors engraft and proliferate to high levels in severe combined immunodeficient mice. *Blood* **83**, 2489–2497 (1994).

56. Lapidot, T. *et al*. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).

57. Taussig, D. C. *et al*. Hematopoietic stem cells express multiple myeloid markers: Implications for the origin and targeted therapy of acute myeloid leukemia. *Blood* **106**, 4086–4092 (2005).

58. Kersten, B. *et al*. CD45RA, a specific marker for leukaemia stem cell sub-populations in acute myeloid leukaemia. *Br. J. Haematol*. **173**, 219–235 (2016).

59. Zeijlemaker, W. *et al*. A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. *Leukemia* **30**, 439–46 (2016).

60. Sarry, J. *et al*. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J. Clin. Invest.* **121**, 384–395 (2011).

61. Quek, L. *et al.* Genetically distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *J. Exp. Med.* **213**, 1513–35 (2016).

62. Bachas, C. *et al.* The role of minor subpopulations within the leukemic blast compartment of AML patients at initial diagnosis in the development of relapse. *Leukemia* **26**, 1313–1320 (2012).

63. Ng, S. W. K. W. *et al*. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).

64. Costello, R. T. *et al.* Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. *Cancer Res.* **60**, 4403–4411 (2000).

65. van Rhenen, a *et al.* Aberrant marker expression patterns on the CD34+CD38− stem cell compartment in acute myeloid leukemia allows to distinguish the malignant from the normal stem cell compartment both at diagnosis and in remission. *Leukemia* **21**, 1700–1707 (2007).

66. Terwijn, M. *et al.* Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. *PLoS One* **9**, e107587 (2014).

67. Zeijlemaker, W. *et al.* Absence of leukaemic CD34+ cells in acute myeloid leukaemia is of high prognostic value: A longstanding controversy deciphered. *Br. J. Haematol.* **171**, 227–238 (2015).

68. Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment | FDA. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hematologic-malignancies-regulatory-considerations-use-minimal-residual-disease-development-drug-and. (Accessed: 23rd April 2020)

CHAPTER 2

LEUKEMIC STEM CELLS; IDENTIFICATION AND CLINICAL APPLICATION

Diana Hanekamp, Jacqueline Cloos and Gerrit Jan Schuurhuis

International Journal of Hematology 2017

INTRODUCTION

Acute myeloid leukemia (AML) is a heterogeneous disease characterized by a multitude of molecular abnormalities. Despite increasing understanding of the pathogenesis no significant changes in treatment have been achieved as yet. With current treatment strategies, the percentage of adult patients achieving complete remission has increased, mainly due to better risk classifications and improvements in supportive care. However, with still roughly half of these patients relapsing, survival rates remain low. A rare population of therapy resistant cells is believed to be at the origin of the relapse. Since these cells have the self-renewal capacity to repopulate a leukemia despite their low frequency, they are considered leukemic stem cells (LSC), also referred to as the leukemia initiating cells (LIC). Whether these cells originate from normal hematopoietic stem cells (HSC) or from more mature progenitors that gained stemness features remains elusive and may differ among patients. Currently, many studies reveal the importance of estimating LSC burden for prognostic purposes and strategies to eradicate these cells in order to completely eliminate the leukemia. In this review we will focus mainly on the identification of these LSC using flow cytometry and summarize the current knowledge on LSC genetics and novel opportunities for elimination of these LSC.

IDENTIFICATION OF LEUKEMIC STEM CELLS

In order to identify LSC, knowledge of their specific characteristics is essential. The recognition of stemness features (e.g. drug resistance, self-renewal and undifferentiated state)[1] alone is not sufficient since those features are also characteristics for HSC coexisting in the bone marrow (BM)[2]. The existence of leukemic cells that meet these criteria was first demonstrated over twenty years ago. In the nineties, Dick and colleagues demonstrated that a rare fraction of AML cells (i.e. cells with the CD34+/CD38- immunophenotype, similar to HSC) were capable of generating leukemia in immunodeficient mice[3,4]. Since this discovery, these putative LSC have been the focus of extensive research. Golden standard property of LSC populations is the ability to engraft and initiate leukemia in a recipient mouse (initiation), to grow out after re-transplantation into secondary recipients (self-renewal) and preferable in tertiary recipients. Since normal CD34+/CD38- cells possess similar features as LSC and the design of new therapies require the specific eradication and monitoring of CD34+CD38- LSC it is crucial to specifically discriminate LSC containing fractions from HSC using cell surface markers. Table 1 presents a summary of markers that are commonly used to distinguish LSC from HSC, thereby allowing to define the contribution of both the LSC and the HSC to the total CD34+/CD38- compartment. However, each marker allowed identification of LSC only in part of the AML patient population and often identified only part of the total LSC population in a particular patient[5,6]. Combining all markers and other properties that distinguish LSC from HSC[5] allows to robustly identify

Table 1. Distinct leukemic stem cell markers

Marker	Identified as	Expression				Ref.
		Normal	in AML	HSC	CD34+CD38- LSC	
IL1RAP	IL1R3	T cells	79%	-	+	[82-84]
CLL-1	CLEC12A, MICL, DCAL-2	myeloid cells	70%	-	+	[6]
TIM-3	T-cell Ig Mucin 3	activated T cells, NK cells	91%	-	+	[85]
CD2	SRBC, LFA2, T11	T cells, NK cells	87%	-	+	[14]
CD7	GP40, TP41, LEU-9	T cells	43%	-	+	[6]
CD11b	Integrin alpha M, Mac-1	myeloid cells	55%	-	+	[6]
CD22	BL-CAM, Siglec-2	B cells	51%	-	+	[6]
CD25	IL2RA, TAC	activated B and T cells	25%	-	+	[86]
CD33	P67, Siglec-3	myeloid cells, NK cells	82%	+	++	[6][87]
CD44	Adhesion molecule	ubiquitously	100%	+	++	[6]
CD45RA	Tyrosine phosphatase receptor type C	T cells, myeloid cells	65%	-	+	[88]
CD47	Integrin associated protein (IAP)	ubiquitously	100%	+	++	[89]
CD56	N-CAM, MSK39	NK cells, activated T cells	32%	-	+	[6]
CD96	TACTILE	activated T cells	33%	-	+	[6]
CD99	MIC2, single-chain type-1 glycoprotein	myeloid cells	83%	-	+	[90]
CD123	IL3R	myeloid cells	82%	+	++	[6][60][91]

the LSC and to estimate its frequency. Since the use of all markers to identify CD34+/CD38- LSC fraction of the total CD34+/CD38- compartment in each patient would require a dramatic amount of work and money, we developed a simplified comprehensive panel of markers that included only non-redundant LSC markers[6].

The existence of LSC outside the CD34+CD38- fraction has been proposed as early as 1996, when CD34- cells were shown initiating leukemia in immune deficient mice[6,7]. Indeed, in less immunodeficient mice strains, both CD34+ and CD34- populations engraft[8-10]. Besides the influence of distinct properties of the mouse models used on engraftment, important studies of Bonnet et al. showed that there is another phenomenon that determines engraftment: they showed that the anti-CD38 antibody (used to purify CD34+CD38-, CD34+CD38+, CD34-CD38- and CD34-CD38+ cells prior to transplantation) inhibited subsequent engraftment[11,12]. This is suggesting that earlier results could be based on technical interference rather than true biological differences of the CD38- and CD38+ populations.

Taken together, these studies advocate that LSC might co-exist in all CD34/CD38 defined subpopulations. In a recent study of Ng et al., the four CD34/CD38 defined cell populations of AML patients were sorted and were subsequently injected into mice and screened for their leukemia-initiating ability[13]. This exquisite approach confirmed that LSC activity was detected in all fractions, however with a majority of CD34+ fractions, especially CD34+/CD38-, and minority of CD34- fractions containing LSC. The fact that there were hardly cases in which leukemia initiating cells originated from CD34- and/or CD34+CD38+ without concomitant activity in CD34+CD38- suggests that the CD34+/CD38- fraction contains the most important leukemia initiating cells when the other fractions are concomitantly present. This hypothesis is confirmed by other observations: in transplantation experiments using NOD/SCID mice of unfractionated AML, engraftment correlated only with the CD34+/CD38- frequency in the original sample, but not with the CD34+/CD38+ or CD34 frequency[14]. In addition, in line with the finding of in vitro and in vivo therapy resistance[15,16], it was found that it is only the CD34+/CD38- LSC frequency that correlates with therapy outcome and minimal residual disease (MRD) levels, i.e. number of leukemic blasts detected after therapy[5]. This suggests that it is this fraction, and not the CD34+/CD38+ and CD34- fractions, that preferentially survives therapy and recapitulate leukemia.

The frequency of LSC within all mononuclear cells is shown to vary widely between patients (1 in $1.6x10^3$ – 1 in $1.1x10^6$)[10]. Since the CD34+CD38- population frequency is 1 in $5x10^3$,[17] we need, at least in part of the patients, assays to identify the smaller subpopulation of LSC within this population. Exploiting other (non-immunophenotypical) features of the LSC allows this.

About 20%[18] of AML cases are characterized by absence of neoplastic CD34+ cells[18,19]. In these cases the commonly small CD34+ (<1%) blast population does not contain leukemic cells[18,20,21]. By definition, these CD34negative patients lack CD34+/CD38- or CD34+/CD38+ leukemic populations, however, a potential LSC population should be

found within the remaining CD34- fraction. Apparently, there are small and yet unidentified subpopulations to consider as most leukemogenic and therapy resistant in these AML cells.

Since LSC are supposed to be relatively chemotherapy resistant, the finding of a very small cellular compartment that is defined by high ABC drug transporter activity is of particular interest. Indeed a specific cell population (i.e. side population (SP)) could be identified using flow cytometry analysis in which the specific Hoechst dye 33342[22,23] is extruded efficiently by these drug pumps. These SP cells are resistant to AML therapies that include drugs that are used for treatment of AML patients like anthracyclines[24]. Purified SP cells were shown to have leukemic initiating capacity in NOD/SCID mouse models[24,25] and contained both CD34+ cells and CD34- cells[26,27] which are indeed in part neoplastic[28]. Although this suggest that a small part of the CD34- cells are therapy resistant, it remains to be established whether the SP cells are candidates for the leukemia initiating cells in so called CD34negative leukemia.

Since the SP population can contain both HSC and LSC, inclusion of LSC specific surface markers should aid in distinction between the LSC and HSC within the SP.

Next, stem cells are known to protect themselves by high expression of aldehyde dehydrogenase (ALDH), which is a cytosolic enzyme involved in retinoic acid metabolism maintaining cellular homeostasis. ALDH is shown to protect against DNA damage induced by reactive oxygen species and reactive aldehydes. In normal BM CD34+CD38- HSC display high levels ALDH activity (ALDHhigh)[29]. In both normal BM and in the majority of AML BM cells, the CD34+CD38-ALDHhigh population is considered to contain only HSC[21,29,30]. In contrast to normal BM, in AML a second population can be discriminated with cells having intermediate ALDH expression[29]. When purified, this population was most potent in AML engraftment in immunodeficient mice and was generally found positive for leukemic cytogenetic markers[29]. Furthermore, presence of this population after therapy was highly predictive for relapse[29]. In conclusion, ALDH activity can be used as a functional stem cell marker, identifying HSC population and LSC population in AML. Validity to therapeutically target ALDH in AML treatment is controversial; a recent paper showed that in vitro and in vivo inhibition of ALDH selectively eradicates CD34+CD38-ALDH+ cells[31]. In this study, the authors used the CD34+CD38-ALDH+ phenotype to describe LSC, which is distinct from most other studies that define CD34+CD38-ALDH+ to reflect HSC. To reveal whether these CD34+CD38-ALDH+ cells that are targeted, are indeed neoplastic cells, additional genetic characterization might be insightful.

LSC HETEROGENEITY OF LSC WITHIN A PATIENT

Recently it was shown that the constitution of AML at relapse may differ from diagnosis due to clonal changes including clonal evolution, clonal regression and clonal selection, with possible changes on immunophenotypic[32,33], cytogenetic[34], genetic[34,35] and epigenetic[35] level. Detailed whole genome sequencing studies, analyzing paired diagnosis-relapse

samples, showed that at time of diagnosis, patients could present with a wide array of small subclones of which some remained in relapse[36]: indicative for clonal selection under therapy pressure. In 2012, we showed that immunophenotypically defined subpopulations of cells prominent at relapse could be traced back as very minor immature (CD34+/CD38-/dim) subpopulations of cells at diagnosis[32], again suggesting the importance of the CD34+/CD38- leukemic stem cell fraction. Since LSC are currently followed during therapy as biomarker of treatment efficacy and as prognostic factor for relapse, it is of great relevance to identify all (possibly minor) LSC populations that are potentially capable of causing relapse[37].

CLINICAL RELEVANCE OF LSC LOAD FOR PROGNOSIS

The CD34+CD38- burden of AML patients is of strong prognostic value. In adult AML, patients with CD34+CD38- frequencies higher than 3.5% at diagnosis had a median relapse-free survival of 5.6 months, compared to 16 months in those with lower CD34+CD38- frequencies[14]. These results were later confirmed in other studies in adult AML[14] and in pediatric AML[38]. As knowledge on the makeup of the CD34+/CD38- fraction increased, other markers and properties were included anticipating better selectivity in defining LSC as previously summarized[5,29]. In studies on the prognostic impact of CD34+CD38- LSC on disease outcome, the prognostic influence of complete absence of this fraction was also discovered: CD34negative status, characterized by the complete absence of neoplastic CD34+ cells[18], turned out to be an independent prognostic factor identifying patients with better prognosis in adult[17] and pediatric AML[39] compared to patient with high or low CD34+CD38- LSC frequencies.

Despite the accumulating evidence of the prognostic relevance of LSC load at diagnosis, this feature is currently not included in risk group stratification. It is our assumption that implementation of flow cytometric quantification of LSC could be implemented without great effort using our one-tube assay[6]. Since prediction of outcome also greatly depends on many different factors during therapy, including LSC measurements during therapy (for instance at MRD time points) is warranted[40].

IMPACT OF LSC FREQUENCY DURING THERAPY

Assessment of the frequency of remaining leukemic cells present during and after therapy (measurable/minimal residual disease, MRD) is increasingly used as an early read-out of therapy efficacy[6,41]. MRD frequency measurement has been shown to have independent prognostic impact across different cytogenetic and molecular subgroups[42-45], and is currently used to refine risk group classifications after induction therapy. In particular, MRD is implemented in the HOVON/SAKK H132 study to guide decisions for transplantation

type in intermediate risk patients. In this study, immunophenotypic MRD measurements are complemented with mutation analysis in NPM1 mutated patients, in which NPM1 status at MRD is leading for the clinical decision. In fact, many MRD studies are currently being performed, which use (or include) molecular assays[57].

It is remarkable that in all immunophenotype and/or molecular MRD studies still a proportion of MRD-negative patients develop a relapse. There are multiple possible reasons for this, e.g. low assay sensitivity, occurrence of mutational/immunophenotypic shifts or different kinetics of MRD disappearance. There may, however, also be a biological explanation: it may not only be the number of leukemic blast cells, reflecting MRD, that defines the risk of relapse, but also the number of LSC present within this blast cell population. As we argued earlier, stem cells have been demonstrated to be more therapy resistant than leukemic blast cells. The MRD population is thus likely enriched with LSC, but these are too low in frequency to contribute significantly to the total frequency of MRD cells. Indeed, when the number of CD34+/CD38- LSC after therapy was determined, LSC load was an independent predictive factor for patient survival[5]. Such was found by others too, be it with different assays and different immunophenotypical and functional definitions of stem cells[29,46]. Remarkably, assessment of both LSC and MRD led to better separation of patients risk group classification than either MRD or LSC alone[5,17].

For newly tested therapies, survival end point is the most important determinant of the therapeutic effectiveness. However, large clinical trials are needed with high numbers of included patients. At best, it than takes approximately 2-3 years to predict survival[47]. With increasing numbers of tested therapies, specifically targeting LSC, usage of LSC frequencies as surrogate intermediate endpoint for survival would be highly beneficial[48].

THERAPEUTIC OPPORTUNITIES ELIMINATING LSC

GENERAL PRINCIPLES AND CHALLENGES FACED BY TARGETING LSC

With the poor prognosis of AML and only little improvements in therapeutic options, there is a pressing need for novel therapies. Therapies targeting LSC offer hope for such improvement. Fundamental to LSC therapy is the selection of the target and the timing of the therapy. Ideally, the target is highly expressed by LSC, highly selective, i.e. absence of expression on other cells in particular HSC and no circulating antigens, and preferentially expressed by high numbers of patients. Acknowledging the many similarities that LSC and HSC share, it is not surprising that current treatment approaches are limited. As the specific identification and, with that, the characterization of LSC has become more detailed, therapies directed to LSC, while sparing normal HSC, are becoming reality and are currently investigated as delineated below.

Distinct cell-surface markers have been proposed as potential LSC specific targets (table 1) and several approaches targeting some of these LSC surface markers are currently

in clinical trials (see table 2). Of these markers, therapies targeting CD33 are possibly the most studied in AML patients. Although targeting CD33 was originally not meant as an anti-LSC therapy, it turned out that CD33 was overexpressed in LSC compared to HSC[6]. Treatment with Gemtuzumab ozogamicin (GO) treatment was associated with reduced relapse risk and improved overall survival in patient subgroups[49,50]. Whether GO targets CD33+ LSC, causing the reduction in relapse risk, remains unclear[50] as higher numbers of CD34+/CD38-/CD33+ cells and high CD33 expression levels decreased GO sensitivity in vitro[51]. One lesson that can be learned from GO treatment is clear: high specificity of the therapy is important. CD33 is, next to leukemic cells, also present on most HSC, on mature and immature myeloid cell and on various progenitors[52], possibly underlying toxicities as found in earlier studies[53]. Anti CD123 therapy may have similar disadvantages[6], while results of clinical trials targeting newer discovered surface markers more specific for LSC (including CCL-1[54,55], TIM3[55-57], CD96[58]), will provide important insights in validity of therapies targeting immunophenotypic markers.

Next to specificity, the design of the antibody in terms of conjugates is of importance for effectiveness. Novel engineering of antibodies has potential to improve efficacy and reducing immunogenicity (mechanisms and constructs are reviewed by Scott et al.[59] and Tiller and Tessier[60])

One alternative way of direct LSC targeting is with the use of small molecule inhibitors interfering in key signaling pathways altered in LSC (see table 2). Using this strategy the leukemic progenitor cells are also targeted since mutations found in signaling pathways in AML are not limited to the LSC, but are inherited by their progeny. Recent studies have also indicated the relevance of splicing on signaling pathways[61], therefore, small molecules that affect the spliceosome are also investigated as novel therapeutics to eradicate LSC[62].

FUTURE PERSPECTIVES

LSC maintenance and functioning is related, at least in part, to signals from the BM microenvironment[63-65]. Therapeutic targeting is therefore not only directed to LSC. Initial studies inhibiting factors necessary for LSC homing (e.g. CXCR4, CXCL12) have shown to abrogate chemoresistance[66], suggesting combination therapies with LSC specific targets. Clinical trials targeting the LSC niche are in progress[67] (see table 2).

In this review we have conveyed the important role of LSC in AML with emphasis on the identification of LSC using flow cytometry. As the identification of CD34+/CD38- LSC allows for the identification of patients with a poor prognosis, we consider LSC measurements as valuable asset for clinical decision making. This concerns both risk group classification at diagnosis or definition of risk groups after therapy (in a MRD situation). Seen the large heterogeneity of LSC within and among patients, the identification of all specific LSC would be too costly in terms of AML cells, time and money. For that reason, a broadly applicable simple one-tube approach has been developed, which can easily be

Table 2. Evaluation of novel leukemic stem cell directed drugs

Target	Antibody/small molecule	Efficacy	Trials	Ref.
Therapy targeting stem cell-specific surface markers				
CD33	AMG330 (CD33-CD3 BiTE)	Reduced in vitro CFU	I	[92]
	Gemtuzumab ozogamicin	Selective kill of CD34+CD38-CD123+ LSC, sparing HSC	I - III	[63]
	SGN-CD33A	Activity requires CD33 expression, but does not correlate with expression levels	I - III	[93]
IL1RAP	IgG mAb 81.2	Selective kill of IL1RAP positive leukemic blasts and LCS-enriched populations	N/A[1]	[82]
TIM3	ATIK2a	Selective block of LSC engraftment/ development, sparing HSC	N/A	[68]
CLL-1	CLL1-CD3 BiTE	Internalization leads to stem cell death, induction of CDC and ADCC activity	I	[94]
CD123	SL-101	Selective CFU suppression of leukemic progenitors	N/A	[60]
	SGN-CD123A	Anti-leukemic activity in preclinical AML models	I	[95]
CD44	IgG1 H90	Specificity towards leukemic cells over normal CD34+ cells, inhibits mTOR	I	[96]
Therapy targeting LSC-related molecular pathways				
AKT	MK-2206	Impaired leukomogenesis and reduced LIC frequency in vivo	I - II	[97]
	Perifosine	Reduced clonogenic activity, sparing normal CD34+ cells	I	[98]
mTOR	Torkinib, PP242	Reduced proportion of CD34+ cells *in vivo*	N/A	[99]
	MLN0128	CFU inhibition in LSC isolated from primary and secondary xenograft	N/A	[100]
BCL-2	ABT-263	Selective targets LSC mitochondrial energy generation, induced cell death	N/A[2]	[101]
XPO1	KPT-8602	Selective kill of blasts and LCS in AML patient-derived xenograft models	N/A[3]	[102]
	Selinexor, KPT-330	Selective d ecrease of LIC frequency in AML cells isolated from xenografts	I - II	[103]

Table 2. (continued)

Target	Antibody/small molecule	Efficacy	Trials	Ref.
NF-kB	Parthenolide	Preferentially target AML progenitors (in vitro CFU) and stem cell in SCID xenografts	N/A	[104]
Smoothened	PF-913	Reduced fraction of CD34+CD38- cells, sensitized AML cells to cytosine arabinoside	N/A	[105]
Proteasome	Carfilzomib	Reduced long-term survival of AML CD34+ cells	I	[106]
	Bortezomib	Bortezomib-treated mice showed significant deacrease in LIC-enriched populations	I - III	[107]
Histone deacetylase	Chidamide	Induced apoptosis in LSC-like cells and primary AML CD34+ cells	I - II	[108]
DOT1L	EPZ004777	CFU inhibition in primary samples with DNMT3A mutation, not affecting cells without this mutation	N/A	[109]
Therapy targeting the LSC microenvironment				
CXCR4	Plerixafor, AMD3100	Decreasing BM homing	I - II	[78]
	AMD3465		N/A	[78]
	BMS-936564		I	[110]
VLA4	Natalizumab		II	[111]

CDC; cell dependent cytotoxicity, ADCC; antibody-dependent cell-mediated cytotoxicity, CFU; colony forming unit, BM; bone marrow, [1] in chronic myeloid leukemia, [2] in chronic lymphoid leukemia, [3] in multiple myeloma

implemented in routine diagnostics[6]. Additionally, screening for CD34+/CD38- LSC also enables identification of CD34-negative patients, who generally have a better prognosis[18]. Furthermore, as LSC specific therapies –targeting LSC specific surface markers– become available, individualized therapy may come in view. To select the most effective marker-directed therapy, the LSC phenotype of the individual patients needs to be determined. With increasing numbers of markers becoming available, innovations in flow cytometers will continue to support a growing number of channels/colors available in simultaneous measurements. The currently available multicolor flow cytometry approach used in AML does not exceed ten colors[68]. While this allows a universal screening, for precise characterization of the most pure (very minor) LSC population, multiple markers are needed. Current technological advances will come from high-number-multicolor flow cytometry or Cytoff approaches[69] in which an extensive panel of LSC markers will be available.

ACKNOWLEDGEMENTS

D. Hanekamp receives a grant from the Dutch Cancer Society (KWF)/Alpe d'Huez; Contract Grant Number: ALPE 2013-6371.

REFERENCES

1. Gupta PB, Chaffer CL, Weinberg RA. Cancer stem cells: mirage or reality? Nat. Med. 2009; 15: 1010–1012.

2. Vormoor J, Lapidot T, Pflumio F, Risdon G, Patterson B, Broxmeyer HE, et al. Immature human cord blood progenitors engraft and proliferate to high levels in severe combined immunodeficient mice. Blood 1994; 83:2489–2497.

3. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 1994; 367:645–648.

4. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. 1997; 3:730–737.

5. Terwijn M, Zeijlemaker W, Kelder A, Rutten AP, Snel AN, Scholten WJ, et al. Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. PLoS One 2014; 9:e107587.

6. Zeijlemaker W, Kelder A, Oussoren-Brockhoff YJM, Scholten WJ, Snel AN, Veldhuizen D, et al. A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. Leukemia 2016; 30:439–46.

7. Terpstra W, Prins A, Ploemacher RE, Wognum BW, Wagemaker G, Löwenberg B, et al. Long-term leukemia-initiating capacity of a CD34- subpopulation of acute myeloid leukemia. Blood 1996; 87:2187-2194.

8. Sarry JE, Murphy K, Perry R, Sanchez PV, Secreto A, Keefer C, et al. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD / SCID / IL2R γ c-deficient mice. J Clin Invest. 2011; 121:384-395.

9. Ishikawa F, Yasukawa M, Lyons B, Yoshida S, Miyamoto T, Yoshimoto G, et al. Development of functional human blood and immune systems in NOD/SCID/IL2 receptor γ chainnull mice. Blood 2005; 106:1565–1573.

10. Eppert K, Takenake K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat. Med. 2011; 17:1086–1093.

11. Taussig DC, Miraki-Moud F, Anjos-Afonso F, Pearce DJ, Allen K, Ridler C, et al. Anti-CD38 antibody mediated clearance of human repopulating cells masks the heterogeneity of leukemia initiating cells Anti-CD38 antibody mediated clearance of human repopulating cells masks the heterogeneity of leukemia initiating cells. Blood 2008; 112:568–576.

12. Taussig DC, Vargaftig J, Miraki-Moud F, Griessinger E, Sharrock K, Luke T, et at. Leukemia initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34 - fraction Leukemia initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34 fraction. Blood 2010; 115:1976–1985.

13. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016; 540:433–437.

14. van Rhenen A, Feller N, Kelder A, Westra AH, Rombouts E, Zweegman S, et al. High Stem Cell Frequency in Acute Myeloid Leukemia at Diagnosis Predicts High Minimal Residual Disease and Poor Survival. Clin. Cancer Res. 2005; 11:6520–6527.

15. Costello RT, Mallet F, Gaugler B, Sainty D, Arnoulet C, Gastaut JA, et al. Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. Cancer Res. 2000; 60:4403–4411.

16. Ishikawa F, Yoshida S, Saito Y, Hijikata A, Kitamura H, Tanaka S, et al. Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat. Biotechnol. 2007; 25:1315–1321.

17. Zeijlemaker W, Meijer R, Kelder A, Ham J, Oussoren-Brockhoff YJM, Snel AN, et al. manuscript submitted.

18. Zeijlemaker W, Kelder A, Wouters R, Valk PJM, Witte BI, Cloos J, et al. Absence of leukaemic CD34 + cells in acute myeloid leukaemia is of high prognostic value: a longstanding controversy deciphered. Br. J. Haematol. 2015; 171:227–238.

19. Zhu H, Liu Y, Jiang H, Lu J, Qin Y, Jiang Q, et al. CD34 expression on bone marrow blasts is a novel predictor of poor prognosis independent of FlT3-ITD in acute myeloid leukemia with the NPM1-mutation. Leuk. Res. 2013; 37:624–630.

20. van der Pol MA, Feller N, Roseboom M, Moshaver B, Westra G, Broxterman HJ, et al. Assessment of the normal or leukemic nature of CD34+ cells in acute myeloid leukemia with low percentages of CD34 cells. Haematologica 2003; 88:983–993.

21. Schuurhuis GJ, Meel MH, Wouters F, Min LA, Terwijn M, de Jonge NA, et al. Normal hematopoietic stem cells within the AML bone marrow have a distinct and higher ALDH activity level than co-existing leukemic stem cells. PLoS One 2013; 8:e78897.

22. Boesch M, Reimer D, Rumpold H, Zeimet AG, Sopper S, Wolf F, et al. DyeCycle Violet Used for Side Population Detection is a Substrate of P-Glycoprotein. Cytometry Part A 2012; 81A:517–522.

23. Richard V, Nair MG, Santhosh Kumar TR, Pillai MR. Side population cells as prototype of chemoresistant, tumor-initiating cells. Biomed Res. Int. 2013.

24. Wang Y, Yin C, Feng L, Ma L, Wei Y, Sheng Q. Sorting, identification and enrichment of side population cells in THP-1 acute monocytic leukemia cells. Oncol. Rep. 2013; 29:1923–1931.

25. Roshal M, Chien S, Othus M, Wood BL, Fang M, Appelbaum FR, et al. The proportion of CD34(+)CD38(low or neg) myeloblasts, but not side population frequency, predicts initial response to induction therapy in patients with newly diagnosed acute myeloid leukemia. Leukemia 2013; 27: 728–731.

26. Feuring-Buske M, Hogge DE. Hoechst 33342 efflux identifies a subpopulation of cytogenetically normal CD34+CD38- progenitor cells from patients with acute myeloid leukemia. Blood 2001; 97:3882–3890.

27. Wulf GG, Wang R, Kuehnle I, Weidner D, Marini F, Brenner MK, et al. A leukemic stem cell with intrinsic drug efflux capacity in acute myeloid leukemia. Blood 2001; 98:1166–1173.

28. Moshaver B, van Rhenen A, Kelder A, van der Pol M, Terwijn M, Bachas C, et al. Identification of a small subpopulation of candidate leukemia-initiating cells in the side population of patients with acute myeloid leukemia. Stem Cells 2008; 26:3059–3067.

29. Gerber JM, Smith BD, Ngwang B, Zhang H, Vala MS, Morsberger L, et al. A clinically relevant population of leukemic CD34+CD38- cells in acute myeloid leukemia. Blood 2012; 119:3571–3577.

30. Hoang VT, Buss EC, Wang W, Hoffmann I, Raffel S, Zepeda-Moreno A, et al. The rarity of ALDH+ cells is the key to separation of normal versus leukemia stem cells by ALDH activity in AML patients. Int. J. Cancer 2015; 137:525–536.

31. Venton G, Pérez-Alea M, Baier C, Fournet G, Quash G, Labiad Y, et al. Aldehyde dehydrogenases inhibition eradicates leukemia stem cells while sparing normal progenitors. Blood Cancer J. 2016; 6:e469.

32. Bachas C, Schuurhuis GJ, Assaraf YG, Kwidama ZJ, Kelder A, Wouters F, et al. The role of minor subpopulations within the leukemic blast compartment of AML patients at initial diagnosis in the development of relapse. Leukemia 2012; 26:1313–1320.

33. Ho T, LaMere M, Stevens BM, Ashton JM, Myers JR, O'Dwyer KM, et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. Blood 2016; 128:1671–1679.

34. Parkin B, Ouillette P, Li Y, Keller J, Lam C, Roulston D, et al. Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia. Blood 2013; 121:369–377.

35. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat. Med. 2016; 22:792–799.

36. 36Ding L, Ley TJ, Larson DE, Miller CA, Kobolt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukemia revealed by whole genome sequencing. Nature 2012; 481:506–510.

37. Krönke J, Bullinger L, Teleanu V, Tschürtz F, Gaidzik VI, Kühn MWM, et al. Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. Blood 2013; 122:100–109.

38. Witte KE, Ahlers J, Schäfer I, André M, Kerst G, Scheel-Walter HG, et al. High Proportion of Leukemic Stem Cells at Diagnosis Is Correlated with Unfavorable Prognosis in Childhood Acute Myeloid Leukemia. Pediatr. Hematol. 2011; 28:91–99.

39. Hanekamp DW, Denys B, Kaspers GJL, te Marvelde JG, Schuurhuis GJ, de Haas V, et al. The relevance of stem cell load at diagnosis for the development of relapse in pediatric acute myeloid leukemia. Blood 2015; 126:2584.

40. Schuurhuis GJ. A Gene-expression Profile for Leukaemia. Nature 2016; 540:346–348.

41. Buldini B, Rizzati F, Masetti R, Fagioli F, Menna G, Micalizzi C, et al. Prognostic significance of flow-cytometry evaluation of minimal residual disease in children with acute myeloid leukaemia treated according to the AIEOP-AML 2002/01 study protocol. Br. J. Haematol. 2017; 1–11.

42. Terwijn M, van Putten WLJ, Kelder A, van der Velden VHJ, Brooimans RA, Pabst T, et al. High Prognostic Impact of Flow Cytometric Minimal Residual Disease Detection in Acute Myeloid Leukemia: Data From the HOVON/SAKK AML 42A Study. J. Clin. Oncol. 2013; 31:3889–3897.

43. Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S, et al. Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. JAMA 2015; 314:811–822.

44. Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A, et al. Assessment of Minimal Residual Disease in Standard-Risk AML. N. Engl. J. Med. 2016; 374:422–433.

45. Freeman SD, Virgo P, Couzens S, Grimwade D, Russell N, Hills RK, et al. Prognostic Relevance of Treatment Response Measured by Flow Cytometric Residual Disease Detection in Older Patients With Acute Myeloid Leukemia. J. Clin. Oncol. 2013; 31:4123–4131.

46. Khan N, Freeman SD, Virgo P, Couzens S, Richardson P, Thomas I, et al. An immunophenotypic pre-treatment predictor for poor response to induction chemotherapy in older acute myeloid leukaemia patients: blood frequency of CD34+ CD38 low blasts. Br. J. Haematol. 2015; 170:80–84.

47. Estey E, Othus M, Lee SJ, Appelbaum FR, Gale RP. New drug approvals in acute myeloid leukemia: what's the best end point? Leukemia 2015; 30:521-525.

48. Han L, Jorgensen JL, Brooks C, Shi C, Zhang Q, Nogueras GM, et al. Anti-leukemia efficacy and mechanisms of action of SL-101, a novel anti-CD123 antibody-conjugate in acute myeloid leukemia. Clin. Cancer Res. 2017.

49. Jedema I, Barge RMY, van der Velden VHJ, Nijmeijer BA, van Dongen JJM, Willemze R, et al. Internalization and cell cycle-dependent killing of leukemic cells by Gemtuzumab Ozogamicin: rationale for efficacy in CD33-negative malignancies with endocytic capacity. Leukemia 2003; 18: 316–325.

50. Khan N, Hills RK, Virgo P, Couzens S, Clark N, Gilkes A, et al. Expression of CD33 is a predictive factor for effect of Gemtuzumab Ozogamicin at different doses in adult acute myeloid leukemia. Leukemia 2016.

51. Jawad M, Seedhouse C, Mony Y, Grundy M, Russell NH, Pallis M. Analysis of factors that affect in vitro chemosensitivity of leukaemic stem and progenitor cells to gemtuzumab ozogamicin (Mylotarg) in acute myeloid leukaemia. Leukemia 2010; 24:74–80.

52. Giles F, Estey E, O'Brien S. Gemtuzumab ozogamicin in the treatment of acute myeloid leukemia. Cancer 2003; 98:2095–2104.

53. de Witte T, Amadori S. The optimal dosing of gemtuzumab ozagamicin: Where to go from here? Haematologica 2016; 101:653–654.

54. van Rhenen A, van Dongen GAMS, Kelder A, Rombouts EJ, Feller N, Moshaver B, et al. The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. Blood 2007; 110:2659–2666.

55. Darwish NHE, Sudha T, Godugu K, Elbaz O, Abdelghaffar HA, Hassan EEA, et al. Acute myeloid leukemia stem cell markers in prognosis and targeted therapy: potential impact of BMI-1, TIM-3 and CLL-1. Oncotarget 2016; 7:57811-57820.

56. Kikushige Y, Shima T, Takayanagi S, Urata S, Miyamoto T, Iwasaki H, et al. TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. Cell Stem Cell 2010; 7:708–717.

57. Kikushige Y, Miyamoto T. Identification of TIM-3 as a Leukemic Stem Cell Surface Molecule in Primary Acute Myeloid Leukemia. Oncology 2015; 89(suppl 1):28–32.

58. Du W, Hu Y, Lu C, Li J, Liu W, He Y, et al. Cluster of differentiation 96 as a leukemia stem cell-specific marker and a factor for prognosis evaluation in leukemia. Mol. Clin. Oncol. 2015; 3:833–838.

59. Scott AM, Wolchok JD, Old LJ. Antibody therapy of cancer. Nat. Rev. 2012; 12:278–287.

60. Tiller KE, Tessier PM. Advances in Antibody Design. Annu. Rev. Biomed. Eng 2015; 17:191–216.

61. Wojtuszkiewicz A, Assaraf YG, Maas MJP, Kaspers GJL, Jansen G, Cloos J. Pre-mRNA splicing in cancer: the relevance in oncogenesis, treatment and drug resistance. Expert Opin. Drug Metab. Toxicol. 2015; 11:673–689.

62. Lee SCW, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. Nat. Med. 2016; 22: 976–986.

63. Lee GY, Kim JA, Oh IH. Stem cell niche as a prognostic factor in leukemia. BMB Rep. 2015; 48: 427–428.

64. Konopleva MY, Jordan CT. Leukemia stem cells and microenvironment: biology and therapeutic targeting. J. Clin. Oncol. 2011; 29:591–599.

65. Sánchez-Aguilera A, Méndez-Ferrer S. The hematopoietic stem-cell niche in health and leukemia. Cell. Mol. Life Sci. 2017; 74:579-590.

66. Rashidi A, DiPersio, JF. Targeting the leukemia-stroma interaction in acute myeloid leukemia: rationale and latest evidence. Ther. Adv. Hematol. 2016; 7:40–51.

67. Barbieri F, Bajetto A, Thellung S., Würth R, Florio T. Drug design strategies focusing on the CXCR4/CXCR7/CXCL12 pathway in leukemia and lymphoma. Expert Opin. Drug Discov. 2016; 11: 1093–1109.

68. Getta BM, Devlin S, Maloy MA, Mohanty A, Arcila M, Tallman MS, et al. Multicolor Flow Cytometry and Multi-Gene Next Generation Sequencing Are Complimentary and Highly Predictive for Relapse in Acute Myeloid Leukemia Following Allogeneic Hematopoietic Stem Cell Transplant. Blood 2016; 128:834.

69. Han L, Qiu P, Zeng Z, Jorgensen JL, Mak DH, Burks JK, et al. Single-cell mass cytometry reveals intracellular survival/ proliferative signaling in FLT3-ITD-mutated AML stem/ progenitor cells. Cytometry A 2015; 87:346–356.

70. Askmyr M, Ågerstam H, Hansen N, Gordon S, Arvanitakis A, Rissler M, et al. Selective killing of candidate AML stem cells by antibody targeting of IL1RAP. Blood 2013; 121:3709–3713.

71. Ågerstam H, Karlsson C, Hansen N, Sandén C, Askmyr M, von Palffy S, et al. Antibodies targeting human IL1RAP (IL1R3) show therapeutic effects in xenograft models of acute myeloid leukemia. Proc. Natl. Acad. Sci. 2015; 112:10786–10791.

72. Barreyro L, Will B, Bartholdy B, Zhou L, Todorova TI, Stanley RF, et al. Overexpression of IL-1 receptor accessory protein in stem and progenitor cells and outcome correlation in AML and MDS. Blood 2012; 120:1290–1298.

73. Jan M, Chao MP, Cha AC, Alizadeh AA, Gentles AJ, Weissman IL, et al. Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. Proc. Natl. Acad. Sci. 2011; 108:5009–5014.

74. Saito Y, Kitamura H, Hijikata A, Tomizawa-Murasawa M, Tanaka S, Takagi S, et al. Identification of Therapeutic Targets for Quiescent, Chemotherapy-Resistant Human Leukemia Stem Cells Yoriko. Sci Transl Med 2010; 2:17ra9.

75. Ehninger A, Kramer M, Röllig C, Thiede C, Bornhäuser M, von Bonin M, et al. Distribution and levels of cell surface expression of CD33 and CD123 in acute myeloid leukemia. Blood Cancer Journal 2014; 4:e218.

76. Kersten B, Valkering M, Wouters R, van Amerongen R, Hanekamp D, Kwidama Z, et al. CD45RA, a specific marker for leukaemia stem cell sub-populations in acute myeloid leukaemia. Br. J. Haematol. 2016; 173:219–235.

77. Ponce LP, Fenn NC, Moritz N, Krupka C, Kozik JH, Lauber K, et al. SIRPα-antibody fusion proteins stimulate phagocytosis and promote elimination of acute myeloid leukemia cells. Oncotarget 2017; 8:11284–11301.

78. Chung SS, Tavakkoli M, Klimek VM, Park CY. CD99 Is a Therapeutic Target On Disease Stem Cells In Acute Myeloid Leukemia and The Myelodysplastic Syndromes. Blood 2014; 124:3760.

79. van Rhenen A, Moshaver B, Kelder A, Feller N, Nieuwint AWM, Zweegman S, et al. Aberrant marker expression patterns on the CD34+CD38− stem cell compartment in acute myeloid leukemia allows to distinguish the malignant from the normal stem cell compartment both at diagnosis and in remission. Leukemia 2007; 21:1700–1707.

80. Harrington KH, Gudgeon CJ, Laszlo GS, Newhall KJ, Sinclair AM, Frankel SR, et al. The broad anti-AML activity of the CD33/CD3 BiTE antibody construct, AMG 330, is impacted by disease stage and risk. PLoS One 2015; 10:1–13.

81. Sutherland MSK, Walter RB, Jeffrey SC, Burke PJ, Yu C, Kostner H, et al. SGN-CD33A: a novel CD33-targeting antibody-drug conjugate using a pyrrolobenzodiazepine dimer is active in models of drug-resistant AML. Blood 2013; 122:1455–1463.

82. Noordhuis P, Terwijn M, Rutten AP, Smit L, Ossenkoppele GJ, Schuurhuis GJ. Targeting of CLEC12A In Acute Myeloid Leukemia by Antibody-Drug-Conjugates and Bispecific CLL-1×CD3 BiTE Antibody. Blood 2010; 116:2890.

83. Sutherland MSK, Yu C, Walter RB, Westendorf L, Valliere-Douglass J, Pan L, et al. SGN-CD123A, a Pyrrolobenzodiazepine Dimer Linked Anti-CD123 Antibody Drug Conjugate, Demonstrates Effective Anti-Leukemic Activity in Multiple Preclinical Models of AML. Blood 2015; 126:330.

84. Gadhoum SZ, Madhoun NY, Abuelela AF, Merzaban JS. Anti-CD44 antibodies inhibit both mTORC1 and mTORC2: a new rationale supporting CD44-induced AML differentiation therapy. Leukemia 2016; 30:2379-2401.

85. Blackburn JS, Liu S, Wilder JL, Dobrinski KP, Lobbardi R, Moore FE, et al. Clonal evolution enhances leukemia propagating cell frequency in T-cell acute lymphoblastic leukemia through akt/mtorc1 pathway activation. Cancer Cell 2014; 25:366–378.

86. Papa V, Tazzari PL, Chiarini F, Cappellini A, Ricci F, Billi AM, et al. Proapoptotic activity and chemosensitizing effect of the novel Akt inhibitor perifosine in acute myelogenous leukemia cells. Leukemia 2008; 22:147–160.

87. Zeng Z, Shi YX, Tsao T, Qiu Y, Kornblau SM, Baggerly KA, et al. apoptosis in AML cells under conditions mimicking the bone marrow microenvironment Targeting of mTORC1 / 2 by the mTOR kinase inhibitor PP242 induces apoptosis in AML cells under conditions mimicking the bone marrow microenvironment. Blood 2012; 120:2679–2689.

88. Zeng Z, Wang Y, Qiu YH, Mak DH, Coombes K, You SY, et al. MLN0128, a novel mTOR kinase inhibitor, disrupts survival signaling and triggers apoptosis in AML and AML stem/ progenitor cells. Oncotarget 2016; 7:55083-55097.

89. Lagadinou ED, Sach A, Callahan K, Rossi RM, Neering SJ, Minhajuddin M, et al. BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. Cell Stem Cell 2013; 12:329–341.

90. Etchin J, Berezovskaya A, Conway AS, Galinsky IA, Stone RM, Baloglu E, et al. KPT-8602, a second-generation inhibitor of XPO1-mediated nuclear export, is well tolerated and highly active against AML blasts and leukemia-initiating cells. Leukemia 2017; 31:143–150.

91. Etchin J, Montero J, Berezovskaya A, Le BT, Kentsis A, Christie AL, et al. Activity of a selective inhibitor of nuclear export, selinexor (KPT-330), against AML-initiating cells engrafted into immunosuppressed NSG mice. Leukemia 2016; 30:190–199.

92. Guzman ML, Rossi RM, Karnischky L, Li x, Peterson DR, Howard DS, et al. The sesquiterpene lactone parthenolide induces apoptosis of human acute myelogenous leukemia stem and progenitor cells. Blood 2005; 105:4163–4169.

93. Fukushima N, Minami Y, Kakiuchi S, Kuwatsuka Y, Hayakawa F, Jamieson C, et al. Small-molecule Hedgehog inhibitor attenuates the leukemia-initiation potential of acute myeloid leukemia cells. Cancer Sci. 2016; 107:1422–1429.

94. van der Helm LH, Bosman MCJ, Schuringa JJ, Vellenga E. Effective targeting of primitive AML CD34+ cells by the second-generation proteasome inhibitor carfilzomib. Br. J. Haematol. 2015; 171: 652–655.

95. Kagoya Y, Yoshimi A, Kataoka K, Nakagawa M, Kumano K, Arai S, et al. Positive feedback between NF-kappaB and TNF-alpha promotes leukemia-initiating cell capacity. J. Clin. Invest. 2014; 124:528–542.

96. Li Y, Chen K, Zhou Y, Xiao Y, Deng M, Jiang Z, et al. A New Strategy to Target Acute Myeloid Leukemia Stem and Progenitor Cells Using Chidamide, a Histone Deacetylase Inhibitor. Curr Cancer Drug Targets 2015; 15:493–503.

97. Rau RE, Rodriguez BA, Luo M, Jeong M, Rosen A, Rogers JH, et al. DOT1L as a therapeutic target for the treatment of DNMT3A-mutant acute myeloid leukemia. Blood 2016; 128:971-981

98. Kuhne MR, Mulvey T, Belanger B, Chen S, Pan C, Chong C, et al. BMS-936564/MDX-1338: A fully human anti-CXCR4 antibody induces apoptosis in vitro and shows antitumor activity in vivo in hematologic malignancies. Clin. Cancer Res. 2013; 19:357–366.

99. Hsieh YT, Jiang E, Pham J, Kim HN, Abdel-Azim H, Khazal S, et al. VLA4 Blockade In Acute Myeloid Leukemia. Blood 2013; 122:3944.

CHAPTER 3

CD34+CD38- LEUKEMIC STEM CELL FREQUENCY TO PREDICT OUTCOME IN ACUTE MYELOID LEUKEMIA

Wendelien Zeijlemaker, Diana Hanekamp,
Tim Grob, Rosa Meijer,
Angèle Kelder, Jannemieke C Carbaat-Ham,
Yvonne JM Oussoren-Brockhoff, Alexander N Snel,
Dennis Veldhuizen, Willemijn J Scholten,
Johan Maertens, Dimitri A Breems,
Thomas Pabst, Markus G Manz,
Vincent HJ van der Velden, Jennichjen Slomp,
Frank Preijers, Jacqueline Cloos,
Arjan A van de Loosdrecht, Bob Löwenberg,
Peter JM Valk, Mojca Jongen-Lavrencic,
Gert J Ossenkoppele, Gerrit J Schuurhuis

Leukemia 2019 May;33(5):1102-1112

ABSTRACT

Current risk algorithms are primarily based on pre-treatment factors and imperfectly predict outcome in acute myeloid leukemia (AML). We introduce and validate a post-treatment approach of leukemic stem cell (LSC) assessment for prediction of outcome.LSC containing CD34+CD38- fractions were measured using flow cytometry in an add-on study of the HOVON102/SAKK trial. Predefined cut-off levels were prospectively evaluated to assess CD34+CD38- LSC levels at diagnosis (n=594), and, to identify LSC^{low}/LSC^{high} (n=302) and MRD^{low}/MRD^{high} patients (n=305) in bone marrow in morphological complete remission (CR). In 242 CR patients combined MRD and LSC results were available. At diagnosis the CD34+CD38-LSC frequency independently predicts overall survival (OS). After achieving CR, combining LSC and MRD showed reduced survival in MRD^{high}/LSC^{high} patients (hazard ratio [HR] 3.62 for OS and 5.89 for cumulative incidence of relapse [CIR]) compared to MRD^{low}/LSC^{high}, MRD^{high}/LSC^{low}, and especially MRD^{low}/LSC^{low} patients. Moreover, in the NPM1mutant positive sub-group, prognostic value of golden standard NPM1-MRD by qPCR can be improved by addition of flow cytometric approaches. This is the first prospective study demonstrating that LSC strongly improves prognostic impact of MRD detection, identifying a patient subgroup with an almost 100% treatment failure probability, warranting consideration of 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 can be detected during therapy within the white blood cell (WBC) compartment of the bone marrow (BM). 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 assessments. 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] These leukemia stem cell like populations (further referred to as LSC) may have different immunophenotypes ($CD34^+CD38^+$, $CD34^+CD38^-$, $CD34^-$).[9–13] However, $CD34^+CD38^-$ stem cells seem to be, in vitro and in vivo, most therapy resistant and least immunogenic.[7,14,15] Furthermore, in a study wherein engraftment potential of CD34/CD38 defined sub-populations was assessed and correlated with gene expression profiles to provide a stem cell signature, a closer look revealed that in almost all AML cases with engraftment, CD34+CD38- sub-populations, but not always CD34+CD38+ and CD34- sub-populations engrafted.[13] Moreover, the CD34+CD38-/dim LSC population is of particular interest because of its independent prognostic impact[10,16,17], in contrast to CD34+CD38+ and CD34- populations.[10] Since the CD34+CD38- LSC cells seem to have the highest leukemogenic ability and therapy resistance, in this paper we focus on the CD34+CD38-LSC containing fraction.

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 cytrometric MRD[2,18], we prospectively validated predefined LSC threshold levels[10] 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. Moreover, combined multiparameter flow cytometry (MFC), using LSC-MRD or MFC-MRD and molecular approach (using the NPM1 qPCR) is most promising to predict relapse in AML.

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.

All of these patients received standard treatment consisting of two cycles of induction chemotherapy followed by either a third cycle, autologous- or allogeneic stem cell transplantation, dependent on pre-treatment risk assessment. In the HOVON102 study 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 HOVON102 clinical trial.[19] 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 for MFC when BM was not available to define leukemia associated immunophenotypes (LAIPs). At follow-up, only BM was used. Sampling procedures are outlined in the supplementary file. MRD assessment was performed as described in detail before.[2] 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.[2,18] For the retrospective combined molecular and MFC analyses, patients from the HOVON42a (n=37, 2001-2008) and HOVON92 (n=9, 2008-2010) were also included.

For detailed information regarding the retrospective molecular MRD analyses we would like to refer to our earlier published paper in JCO.[2]

CD34/LSC status

CD34 status was defined at diagnosis as described by Zeijlemaker et al.[20] 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 LSC^{low} and LSC^{high} cases was used.[10] CD34+CD38-LSCs were defined

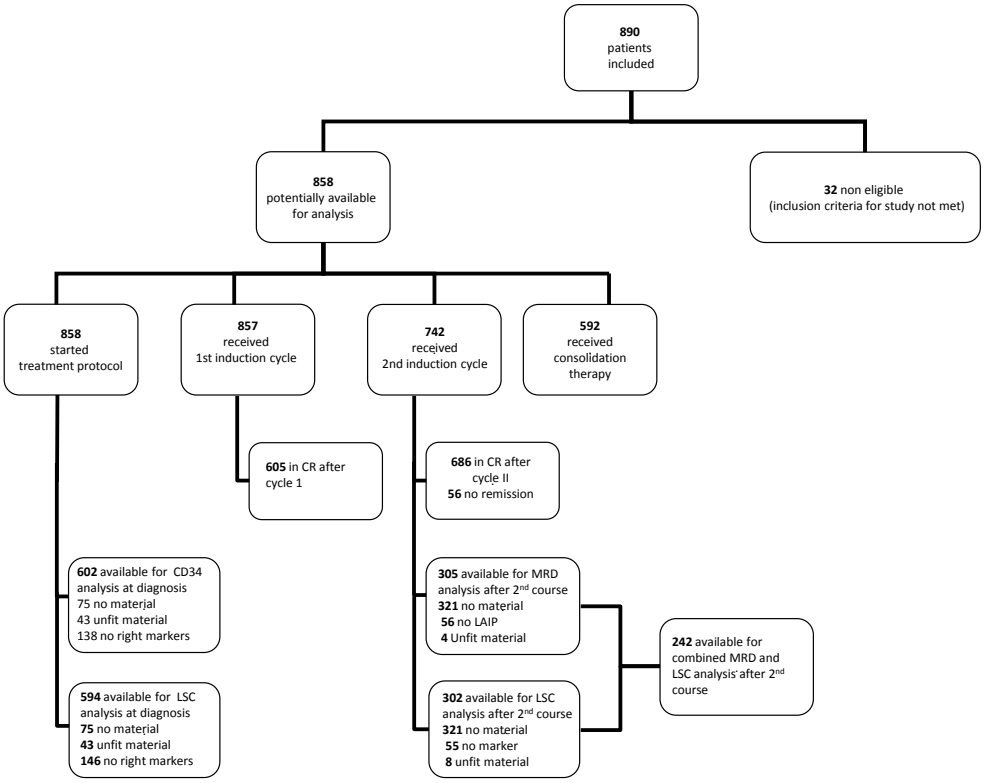


Figure 1. Consort flow chart for sample availability of patients included in the hovon/sakk 102 study. In total 890 patients were included, of which in total 594 patients were available for combined CD34/LSC analyses at time of diagnosis. After the second cycle there were 305 and 302 samples available 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.

at diagnosis and follow-up as described in our earlier paper by Terwijn et al (Figure 1 in that paper).[10] Sampling procedures are outlined in the supplementary file.

The antibody panels used at diagnosis are shown in Table S2. Gating strategy was performed as outlined.[21] Depending on the aberrant immunophenotype(s) as found at AML diagnosis and on the availability of 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 0.0000% (% 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. Table S3 shows the robustness of the data since a range of different cut-off values post-remission showed significant differences between the LSC^{low} and LSC^{high} group.

Table 1. Patient characteristics of CR patients

	Patients with MRD &LSC results	%	Other Patients	%
Total	242		444	
Sex				
Male	119	49	250	56
Female	123	51	194	44
Age, median	52		54	
Range	19-66		18-65	
WBC at diagnosis $x10^9$/L				
Median	8.4		6.5	
Range	0.6-229		0.2-341	
AML type				
De Novo	202	83	358	81
Secondary	24	10	33	7
MDS	16	7	53	12
CD34 status				
CD34 negative	35	14	38	9
CD34 positive LSC ≤0.03%	134	55	138	31
CD34 positive LSC >0.03%	64	26	62	14
Missing	9	4	206	46
MRD / LSC status after 2nd course			NA	
MRD neg LSC neg	136			
MRD neg LSC pos	58			
MRD pos LSC neg	28			
MRD pos LSC pos	20			
Karyotype classification				
Core binding factor	36	15	34	8
Normal karyotype	119	49	238	54
Monosomal karyotype	20	8	44	10
Cytogenetic rest abnormalities	59	24	111	25
Missing	8	3	17	4
Mutated CEBPα				
Neg	171	71	300	68
Pos	7	3	18	4
Missing	64	26	126	28
FLT3/NPM1 status				
$FLT3^{wt}$ / $NPM1^{wt}$	122	50	221	50
$FLT3^{wt}$ / $NPM1^{mut}$	45	19	82	18
$FLT3^{ITD}$ / $NPM1^{wt}$	26	11	35	8
$FLT3^{ITD}$ / $NPM1^{mut}$	26	11	49	11
Missing	23	10	57	13
EVI1				
Neg	171	71	320	72
Pos	19	8	38	9
Missing	52	21	86	19

Table 1. (continued)

	Patients with MRD &LSC results	%	Other Patients	%
IDH				
IDH1^{neg} / IDH2^{neg}	154	64	250	56
IDH1^{pos} / IDH2^{neg}	18	7	35	8
IDH1^{neg} / IDH2^{pos}	27	11	39	9
IDH1^{pos} / IDH2^{pos}	1	0	0	0
Missing	42	17	120	27
DNMT3a				
Neg	139	57	230	52
Pos	61	25	94	21
Missing	42	17	120	27
First consolidation therapy				
None	31	13	87	20
3rd cycle	90	37	134	30
Autologous SCT	32	13	37	8
Allogeneic SCT	89	37	186	42
CR reached				
After cycle 1	207	86	352	79
After cycle 2	35	14	92	21
HOVON risk group				
Good	80	33	123	28
Intermediate	60	25	122	27
Poor	67	28	126	28
Very poor	35	14	73	16
Treatment group				
Standard treatment	129	53	222	50
Clofarabine 10mg	106	44	203	46
Clofarabine 15mg	7	3	19	4

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 both the Fine and Gray model[22] and the Kaplan Meier method 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. For multivariate analyses a combined MRD/LSC parameter was used instead of assuming an additive effect model using MRD and LSC as single parameters. This was based on statistical grounds, since there is an important interaction between MRD and LSC and using an additive effect model could give unreliable results. Multivariate analyses for CIR were calculated using the Fine and Gray model. All variables that were 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

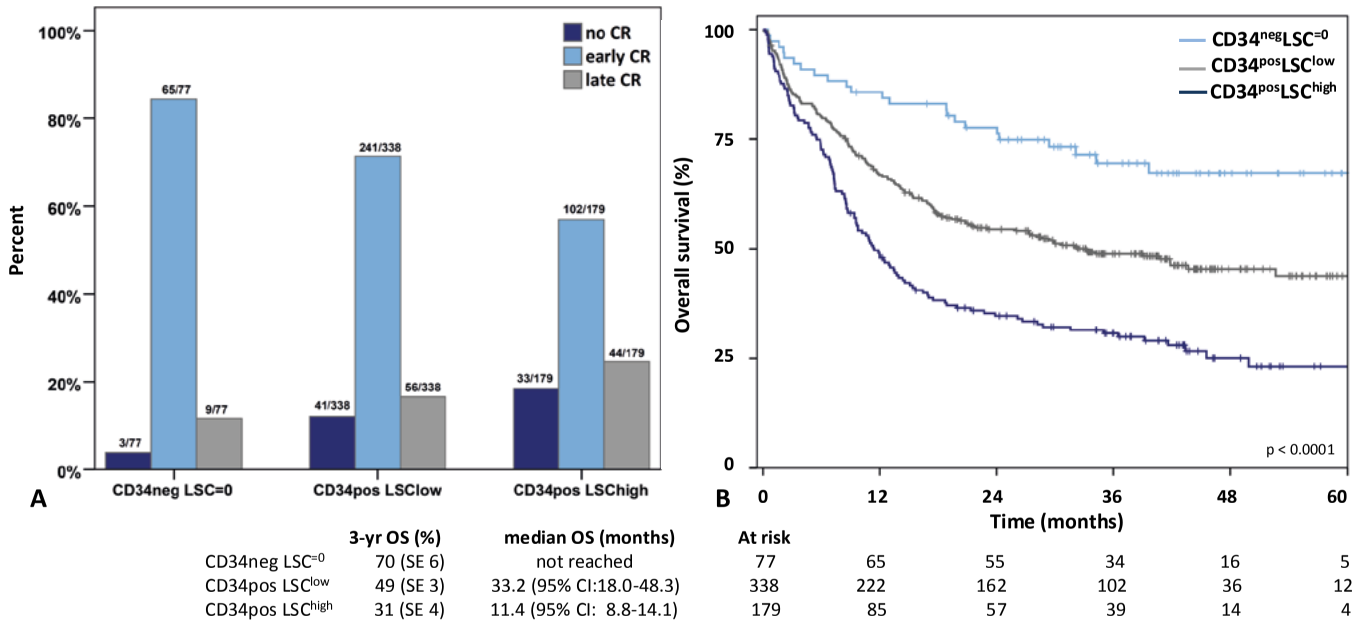
CD34/LSC status at diagnosis

Baseline characteristics

Baseline characteristics of the in total 594 patients with a sample available at time of diagnosis are shown in Table S4. Of these 594 samples, BM was used in 495 cases and PB in 97 cases (data concerning origin of the material was missing in 2 cases). 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^{low}, and 179 (30%) CD34-positive cases with high CD34+CD38- LSC levels (≥0.03%), further referred to as LSC^{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. In 434 of these 517 CD34-positive cases BM was used (median CD34+CD38-LSC percentage 0.0064%, range 0.0000%-19.8761%) and in 81 cases PB was used (median LSC percentage 0.0324%, range 0.0000%-16.8174%). No difference in outcome (OS/EFS) was found regarding patients with a BM or PB sample available (results not shown).

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



	3-yr OS (%)	median OS (months)	At risk					
CD34neg LSC$^{=0}$	70 (SE 6)	not reached	77	65	55	34	16	5
CD34pos LSC^{low}	49 (SE 3)	33.2 (95% CI:18.0-48.3)	338	222	162	102	36	12
CD34pos LSC^{high}	31 (SE 4)	11.4 (95% CI: 8.8-14.1)	179	85	57	39	14	4

Figure 2. Effects CD34/LSC status at diagnosis on complete remission rates and overall survival. Figure 2A illustrates, in the total group of 594 evaluable patients, the difference in remission status between 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 in the figure. Clearly, with increasing CD34+CD38- LSC frequencies (0, LSC^{low}, LSC^{high}), 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 both CD34-positive LSC^{low} and CD34-positive LSC^{high} patients, while, in turn, CD34-positive LSC^{low} have a better survival than CD34-positive LSC^{high} patients. 3-Yrs OS and median OS for the three groups are indicated in the text at the bottom of the figure (left part).

independent prognostic factor for OS: HR in LSC^{low} was 1.94 (95%CI 1.11-3.39) and in LSC^{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^{low} group and 18% (33/179) of LSC^{high} group ($p<0.001$, Figure 2A). In multivariate analyses, odds ratios for achieving CR were 0.65 for LSC^{low} and 0.45 for LSC^{high}; differences were not significant (Table S5).

At diagnosis associations exist with molecular aberrancies as shown in Table S6. As expected[23,24], associations were found between CD34-negative AML (CD34+CD38- LSC=0) and the good prognostic $FLT3^{wt}/NPM1^{mut}$ profile (Table S6A, left part). Moreover, the percentage of $FLT3^{wt}/NPM1^{mut}$ patients was higher in LSC^{low} than in LSC^{high} patients (Table S6A, right part). In addition, the percentage of the poor prognostic $FLT3^{mut}/NPM1^{wt}$ increased with increasing LSC frequency (LSC^{high}> LSC^{low}>CD34 negative). Table S6B shows the association of LSC frequency with the poor prognostic presence of EVI1 over-expression: LSC^{high}> LSC^{low}>CD34 negative. Moreover double mutations of CEBPA are known to be associated with strong CD34 expression.[25] We indeed found this good prognostic marker exclusively in CD34-positive patients, however especially in the LSC^{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.

Diagnosis CD34/LSC results were available for 277/302 patients (92%). Table S7 shows results concerning the occurrence of changes of LSC status between diagnosis and follow-up. These preliminary results show that of in total 40 CD34-negative AML patients at diagnosis 20% (8/40) convert to the CD34-positive status during follow-up. Moreover, it appears that CD34-negative and CD34-positive patients with low LSC levels that convert to LSC positivity after treatment have a worse outcome as compared to the patients that remain LSC negative after treatment (Table S7).

For follow-up analyses, 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. As expected, 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).

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^{neg}/LSC^{neg} (n=136; 56%); 2. MRD^{pos}/LSC^{neg} (n=28; 12%); 3. MRD^{neg}/LSC^{pos} (n=58; 24%); 4. MRD^{pos}/LSC^{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^{neg}/LSC^{neg} , 43% for MRD^{pos}/LSC^{neg}, 53% for MRD^{neg}/LSC^{pos} and 100% for MRD^{pos}/LSC^{pos} patients. Similar results were found for OS with 3-year OS of 66% for MRD^{neg}/LSC^{neg} , 68% for MRD^{pos}/LSC^{neg}, 53% for MRD^{neg}/LSC^{pos} and 0% for MRD^{pos}/LSC^{pos} patients. These results show that double positivity is associated with an excessively poor outcome. Multivariate analyses showed that the combined MRD^{pos}/LSC^{pos} status has profound predictive significance for both OS (HR 3.62), CIR (HR 5.89) and EFS (HR 4.24, Table S8).

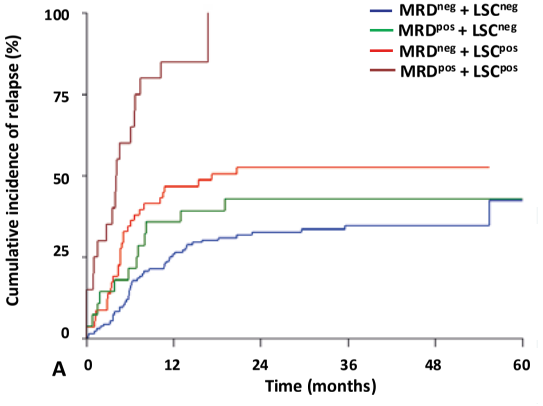
MRD/LSC results in distinct risk categories

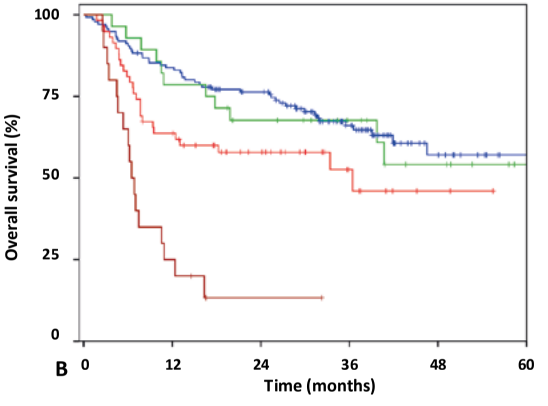
In Table S9 median MRD and LSC levels are shown for the 4 different MRD/LSC subgroups while also taken into account the CD34/LSC status as defined at diagnosis. Although some subgroups are very small, these results show that patients within the $MRD^{pos}LSC^{pos}$ subgroup appear to have higher MRD and LSC levels. Table S10 shows that MRD^{neg}/LSC^{neg} patients are present in all four different risk groups, although the frequency of MRD^{neg}/LSC^{neg} patients becomes lower with increasing risk. Moreover, the frequency of MRD^{pos}/LSC^{pos} patients was highest in the very poor risk group (15%), although MRD^{pos}/LSC^{pos} patients were also present in the other risk groups (e.g 4% in the good risk group). 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 S5 shows MRD/LSC results for CIR (Figure S5.I.A-D) and for OS (Figure S5.II.A-D). Although patient numbers for some subgroups are small, these results show that MRD^{pos}/LSC^{pos} patients have a (very) poor prognosis, even when present in the good or intermediate 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.

LSC-MRD and MFC-MRD as compared to molecular MRD detection

Since molecular MRD is the golden standard (Figure 4A, NPM1 MRD alone) for AML patients with mutated NPM1, we subsequently combined NPM1 qPCR with LSC MRD (n=64) and Flow MRD (n=95) measurements in a subset of patients. When comparing NPM1 MRD with LSC MRD, we find congruent results (either double positive or negative) in 65% of patients. The combined use of both techniques express improved prognostic value for the risk of relapse compared to the individual techniques separately (Figure 4B,





	At risk						3-yr CIR (%)	At risk						3-yr OS (%)	Median OS (months)
MRD^{neg} LSC^{neg}	136	96	77	46	15	5	35 (SE 4)	136	113	95	52	16	6	66 (SE 4)	not reached
MRD^{pos} LSC^{neg}	28	17	13	9	6	1	43 (SE 9)	28	22	17	12	6	1	68 (SE 9)	not reached
MRD^{neg} LSC^{pos}	58	28	20	7	2	0	53 (SE 7)	58	36	22	8	2	0	53 (SE 8)	36.4 (95% CI -)
MRD^{pos} LSC^{pos}	20	3	0	0	0	0	100 (-)	20	5	1	0	0	0	0 (-)	6.5 (95% CI 5.1-7.9)

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^{neg}/LSC^{neg} and MRD^{pos}/LSC^{neg} > MRD^{neg}/LSC^{pos} >> MRD^{pos}/LSC^{pos}. At the bottom CIR, 3-years OS and median OS, are summarized for the different groups.

NPM1 MRD combined with LSC; p<0.001). Similar findings were observed when NPM1 MRD was combined with Flow MRD. In these analyses concordant results were found in 58% of AML patients and the combination of both approaches results in enhanced relapse prediction (Fig 4C and D, NPM1 alone and NPM1 combined with Flow MRD, p=0.001). These results require validation in larger cohorts since flow cytometric approaches might improve current relapse prediction of molecular NPM1 MRD.

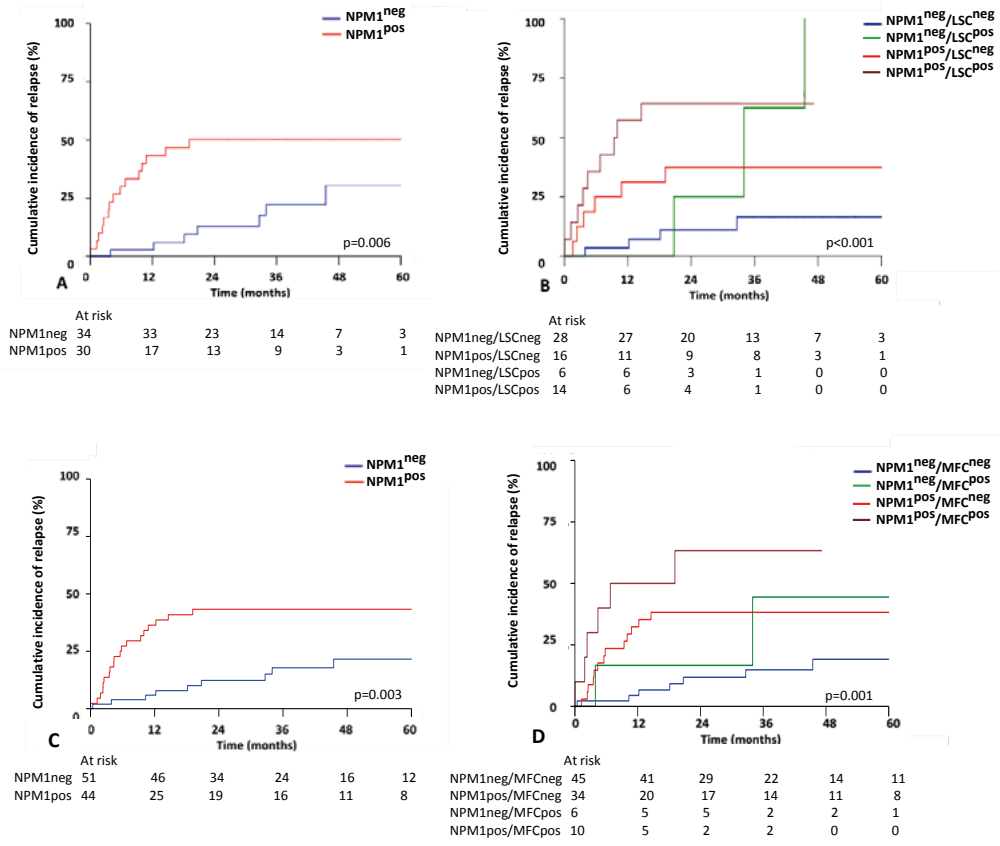


Figure 4. Prognostic value of LSC, MFC-MRD and molecular MRD for NPM1 positive patients. Cumulative incidence of relapse is shown for 64 NPM1 positive AML patients of which NPM1-MRD and LSC results after treatment were available. Figure A shows relapse incidence based on NPM1-MRD measurement as a single parameter for these 64 patients. In Figure B combined results for LSC and NPM1-MRD are shown. Figure C shows cumulative incidence of relapse for 95 NPM1 positive AML patients of which MFC-MRD and NPM1-MRD results were available after treatment. In Figure D combined results for MFC-MRD and NPM1-MRD are shown.

DISCUSSION

MRD as defined by flow cytometry or molecular assays has been proven to be a reproducible marker for relapse prediction.[2,6,26,27] 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.[10,16,28] Support for this notion comes from the prognostic value of LSC frequencies in preliminary retrospective studies.[10,16,28] 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 ($FLT3^{wt}/NPM1^{mut}$) and poor ($FLT3^{mut}/NPM1^{wt}$, EVI1 over-expression) molecular aberrancies suggests that these and probably other genetic aberrations may translate into differences in frequencies of CD34+CD38- LSC. Therefore, CD34+CD38-LSC may represent the total of effects of all mutations that influence therapy resistance, not only at diagnosis, but also at follow up. Our own research group first described the prognostic importance of CD34+CD38- in 2005.[29] Similar results were found by others[17,30]: In these papers focus was on the total CD34+CD38- cell compartment and thus no distinction was made between normal and leukemic CD34+CD38-cells. Hwang et al also demonstrated the prognostic importance of the CD34+CD38-LSC frequency as determined at diagnosis, however the panel of aberrant markers to define LSCs was limited to CD44, CD123 and CD184 and therefore CD34+CD38-LSCs that express other aberrant markers might be missed.[31]

Until recently, risk schemes in AML were primarily based on pre-treatment risk factors, where our approach is also based on post-treatment assessments that account for the occurrence of therapy resistance. It is known that CD34+CD38- LSCs are therapy resistant and immune-evasive[14,15], and, during leukemic growth, may outcompete other stem cell immune-phenotypes.[10] Therefore, CD34+CD38-LSCs, as defined after remission, takes in to account all pre- and post-therapy effects on therapy resistance. In this study we were able to prospectively confirm that the frequency of CD34+CD38-LSCs in patients in remission is an independent prognostic factor for patient outcome. 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^{pos}/LSC^{pos} patients have a high risk of relapse and poor OS (Figure S5.I.A and S5.II.A). Especially for intermediate risk patients the combined MRD/LSC results could be of importance for decisions regarding consolidation treatment.

Although patient numbers are sometimes small, our data strongly suggest that MRD^{pos}/LSC^{pos} patients, independent on risk category, should be considered as poor/very poor risk patients, preferably to be treated as such.

However, there are still MRD^{neg}/LSC^{neg} patients who will relapse and therefore further improvement of this flow cytometric assay is warranted. Especially, to better discriminate

risk of relapse between MRD^{neg}/LSC^{neg}, MRD^{pos}/LSC^{neg} and MRD^{neg}/LSC^{pos} patients. Improvements concerning sensitivity/specificity of the MRD and LSC assay, can be expected when using a smart combination of flow cytometric MRD and molecular MRD.

Although NPM1 offers a very specific, sensitive, and stable target for MRD in AML[6] , our results show that for proper prognostics the addition of MFC-MRD and/or LSC-MRD may be opportune. In the combined LSC and NPM1 analyses, results show (Figure 4) the potential additive value of LSC in NPM1 positive patients. As to the combination of MFC-MRD and NPM1-MRD, it appears that for the best prognostic impact both NPM1-MRD and MFC-MRD are necessary. Of course MFC-MRD and/or LSC-MRD offer the method of choice for the majority of non-NPM1 patients (70.7%). Overall, the additive prognostic value of both LSC-MRD and MFC-MRD in NPM1 positive patients should be validated in future larger studies. Moreover, the exact role of Next-Generation Sequencing (NGS) based MRD as compared to LSC-MRD should be clarified in future studies. For MFC-MRD it recently has been shown that it is supplementary as compared to NGS-MRD.[32]

Different studies have shown prognostic significant of MRD levels below the 0.1% cut-off value.[3,33,34] Therefore, lowering the cut-off value for MRD and combine these results with LSC levels may define patients with a particularly good outcome (we have found this for a cut-off of 0.01%; results not shown).

Further improvement is possible when using the combination of MRD/LSC assessment after different courses of therapy (for instance after course 1 and 2 or sequential monitoring after remission), and finally, the detection of upcoming MRD and LSC populations is of high importance for further improvement: 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.[35–39] Table S7 shows results concerning change of LSC status between diagnosis and follow-up. In the present study, however, mainly aberrancies defined at diagnosis were used for MRD and, to a lesser extent, for LSC assessment. Thus, for future studies, including upcoming populations could reduce frequencies of false negative MRD/LSC results. Lastly, LSC tracking for individual patients is challenging due to the very low numbers of LSC in remission BM. Therefore, for future LSC tracking, 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. An 8- color "one-tube" LSC approach has already been developed[40] whereas an 18-color "one-tube" approach is currently being tested in our institute.

Overall, we show that the LSC frequency at baseline and after chemotherapy is an independent prognostic factor, both in $NPM1^{mutant}$ and $NPM1^{wt}$ patients 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 HOVON/SAKK study centers, and the patients herein.
Conflict of interest: The authors declare no competing financial interests.

FURTHER SUPPLEMENTARY TEXT/TABLES/FIGURES

Further supplemental data can be found online at https://www.nature.com/articles/s41375-018-0326-3.

REFERENCES

1. Hokland P, Ommen HB, Mulè MP, Hourigan CS. Advancing the Mrd Concept in Acute Myeloid. Semin Hematol 2015; 52: 184–92.

2. Terwijn M, van Putten WLJ, Kelder A, van der Velden VHJ, Brooimans RA, Pabst T, et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. J Clin Oncol 2013; 31: 3889–97.

3. Freeman SD, Virgo P, Couzens S, Grimwade D, Russell N, Hills RK, et al. Prognostic relevance of treatment response measured by flow cytometric residual disease detection in older patients with acute myeloid leukemia. J Clin Oncol 2013; 31: 4123–31.

4. Loken MR, Alonzo T a, Pardo L, Gerbing RB, Raimondi SC, Hirsch B a, et al. Residual disease detected by multidimensional flow cytometry signifies high relapse risk in patients with de novo acute myeloid leukemia: a report from Children's Oncology Group. Blood 2012; 120: 1581–8.

5. Zhu H, Zhang X, Qin Y, Liu D, Jiang H, Chen H, et al. MRD-directed risk stratification treatment may improve outcomes of t(8 ;21) AML in the first complete remission: results from the AML05 multicenter trial. Blood 2013; 121: 4056–62.

6. Ivey A, Hills RK, Simpson M a, Jovanovic J V, Gilkes A, Grech A, et al. Assessment of Minimal Residual Disease in Standard-Risk AML. N Engl J Med 2016; 374: 422–33.

7. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 1997; 3: 730–7.

8. Becker MW, Jordan CT. Leukemia stem cells in 2010: current understanding and future directions. Blood Rev 2011; 25: 75–81.

9. Taussig DC, Vargaftig J, Miraki-Moud F, Griessinger E, Sharrock K, Luke T, et al. Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. Blood 2010; 115: 1976–84.

10. Terwijn M, Zeijlemaker W, Kelder A, Rutten AP, Snel AN, Scholten WJ, et al. Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. PLoS One 2014; 9: e107587.

11. Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell 2011; 19: 138–52.

12. Sarry J-E, Murphy K, Perry R, Sanchez P V, Secreto A, Keefer C, et al. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. J Clin Invest 2011; 121: 384–95.

13. Ng SWK, Mitchell A, Kennedy J a., Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016; 540: 433–7.

14. Ishikawa F, Yoshida S, Saito Y, Hijikata A, Kitamura H, Tanaka S, et al. Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat Biotechnol 2007; 25: 1315–21.

15. Costello RT, Mallet F, Gaugler B, Sainty D, Arnoulet C, Gastaut J a, et al. Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy

and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. Cancer Res 2000; 60: 4403–11.

16. Bradbury C, Houlton a E, Akiki S, Gregg R, Rindl M, Khan J, et al. Prognostic value of monitoring a candidate immunophenotypic leukaemic stem/ progenitor cell population in patients allografted for acute myeloid leukaemia. Leukemia 2014; 9: 1–4.

17. Jentzsch M, Bill M, Nicolet D, Leiblein S, Schubert K, Pless M, et al. Prognostic impact of the CD34+/CD38- cell burden in patients with acute myeloid leukemia receiving allogeneic stem cell transplantation. Am J Hematol 2017; 92: 388–96.

18. Feller N, van der Pol M a, van Stijn A, Weijers GWD, Westra a H, Evertse BW, et al. MRD parameters using immunophenotypic detection methods are highly reliable in predicting survival in acute myeloid leukaemia. Leukemia 2004; 18: 1380–90.

19. Löwenberg B, Pabst T, Maertens J, van Norden Y, Biemond BJ, Schouten HC, et al. Therapeutic value of clofarabine in younger and middle aged (18 – 65 yrs) adults with newly diagnosed AML. Blood 2017; 129: 1636–45.

20. Zeijlemaker W, Kelder A, Wouters R, Valk PJM, Witte BI, Cloos J, et al. Absence of leukaemic CD34+ cells in acute myeloid leukaemia is of high prognostic value: a longstanding controversy deciphered. Br J Haematol 2015; 171: 227–38.

21. Cloos J, Harris J., Janssen JJ, Kelder A, Huang F, Sijm G, et al. Comprehensive Protocol to Sample and Process Bone Marrow for Measuring Measurable Residual Disease and Leukemic Stem Cells in Acute Myeloid Leukemia. J Vis Exp 2018; 133. doi:10.3791/56386.

22. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc 1999; 94: 496–509.

23. Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N Engl J Med 2005; 352: 254–66.

24. Tsykunova G, Reikvam H, Hovland R, Bruserud Ø. The surface molecule signature of primary human acute myeloid leukemia (AML) cells is highly associated with NPM1 mutation status. Leukemia 2012; 26: 557–9.

25. Zeijlemaker W, Gratama JW, Schuurhuis GJ. Tumor heterogeneity makes AML a 'moving target' for detection of residual disease. Cytometry B Clin Cytom 2014; 86: 3–14.

26. Chen X, Xie H, Wood BL, Walter RB, Pagel JM, Becker PS, et al. Relation of clinical response and minimal residual disease and their prognostic impact on outcome in acute myeloid leukemia. J Clin Oncol 2015; 33: 1258–64.

27. Schuurhuis GJ, Ossenkoppele G. Minimal residual disease in acute myeloid leukemia: already predicting a safe haven? Expert Rev Hematol 2010; 3: 1–5.

28. Khan N, Freeman SD, Virgo P, Couzens S, Richardson P, Thomas I, et al. An immunophenotypic pre-treatment predictor for poor response to induction chemotherapy in older acute myeloid leukaemia patients: blood frequency of CD34+ CD38 low blasts. Br J Haematol 2015; 170: 80–4.

29. Van Rhenen A, Feller N, Kelder A, Westra AH, Rombouts E, Zweegman S, et al. High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. Clin Cancer Res 2005; 11: 6520–7.

30. Witte K-E, Ahlers J, Schäfer I, André M, Kerst G, Scheel-Walter H-G, et al. High

proportion of leukemic stem cells at diagnosis is correlated with unfavorable prognosis in childhood acute myeloid leukemia. Pediatr Hematol Oncol 2011; 28: 91–9.

31. Hwang K, Park C-J, Jang S, Chi H-S, Kim D-Y, Lee J-H, et al. Flow cytometric quantification and immunophenotyping of leukemic stem cells in acute myeloid leukemia. Ann Hematol 2012; 91: 1541–6.

32. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, al Hinai A, Zeilemaker A, et al. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. N Engl J Med 2018; 378: 1189–99.

33. Quek L, Otto GW, Garnett C, Lhermitte L, Karamitros D, Stoilova B, et al. Genetically distinct leukemic stem cells in human CD34 − acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. J Exp Med 2016; 213: 1513–35.

34. Walter RB, Buckley SA, Pagel JM, Wood BL, Storer BE, Sandmaier BM, et al. Significance of minimal residual disease before myeloablative allogeneic hematopoietic cell transplantation for AML in fi rst and second complete remission. Blood 2013; 122: 1813–22.

35. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukemia revealed by whole genome sequencing. Nature 2012; 481: 506–10.

36. Bachas C, Schuurhuis GJ, Assaraf YG, Kwidama ZJ, Kelder A, Wouters F, et al. The role of minor subpopulations within the leukemic blast compartment of AML patients at initial diagnosis in the development of relapse. Leukemia 2012; 26: 1313–20.

37. Bachas C, Schuurhuis GJ, Hollink IHIM, Kwidama ZJ, Goemans BF, Zwaan CM, et al. High-frequency type I/II mutational shifts between diagnosis and relapse are associated with outcome in pediatric AML: implications for personalized medicine. Blood 2010; 116: 2752–8.

38. Wang ES, Sait SNJ, Gold D, Mashtare T, Starostik P, Ford LA, et al. Genomic, immunophenotypic, and NPM1/FLT3 mutational studies on 17 patients with normal karyotype acute myeloid leukemia (AML) followed by aberrant karyotype AML at relapse. Cancer Genet Cytogenet 2010; 202: 101–7.

39. Ottone T, Zaza S, Divona M, Hasan SK, Lavorgna S, Laterza S, et al. Identification of emerging FLT3 ITD-positive clones during clinical remission and kinetics of disease relapse in acute myeloid leukaemia with mutated nucleophosmin. Br J Haematol 2013. doi:10.1111/bjh.12288.

40. Zeijlemaker W, Kelder A, Oussoren-Brockhoff YJM, Scholten WJ, Snel a N, Veldhuizen D, et al. A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. Leukemia 2016; 30: 439–46.

SUPPLEMENTARY DATA

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]). The HOVON 42a and HOVON 92 studies were also reviewed and approved (HOVON 42a number 2000-220; HOVON 92 number 2008/216). On the HOVON website (www.hovon.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/m^2, days 1-3) and cytarabine (200mg/m^2, days 1-7), and the second induction cycle contained amsacrine (120 mg/m^2, days 4-6) and cytarabine (1000mg/m^2, 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.[1] The experimental arm in the HOVON 42a consisted of standard induction therapy combined with granulocyte-colony-stimulating factor (G-CSF). In the HOVON 92 study patients received laromustine besides standard induction therapy. 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/m^2, days 1-5) and etoposide (100mg/m^2, days 1-5). In general good risk patients received a third cycle of chemotherapy, intermediate risk patients mainly received an autologous or an allogeneic transplantation dependent on the expected treatment related mortality, poor risk and very poor patients in general proceeded to an allogeneic transplantation.

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 $2x10^6$ 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 this[3]: 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.[4] and LSC status at diagnosis and follow-up by Terwijn et al.[5] CD34 data of 173 patients (of the 594 patients in total), as defined at time of diagnosis, have been previously published.[4] Flow cytometric data concerning LSC values at diagnosis and LSC/MRD values as defined at follow-up were not previously published. In our previous study[5], 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.[5]), 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).

Table S2. Antibody panels

6-Color antibody panel				
Tube	FITC	PE	PerCP	PC7
1	PBS	PBS	CD45 (2D1, BD)	CD34 (581, BC)
2	CD2 (MT910, DC)	CD7 (M-T701, BD)	CD45	CD34
3	CD36 (CLB-IVC7, Sanquin)	CD133 (AC133/1, MB)	CD45	CD34
4	CD15 (MMA-BD)	CD33 (P67.6, BD)	CD45	CD34
5	CD13 (WM-47, DC)	CD56 (My31, BD)	CD45	CD34
6	CD7 (M-T701, BD)	TIM-3 (344823, R&D)	CD45	CD34
7	CD11b (Bear1)	CLL-1 (687317, R&D)	CD45	CD34
8-Color antibody panel				
Tube	FITC	PE	PerCP-CY5.5	PC7
1	HLA-DR (L243, BD)	CD33 (P67.6, BD)	CD13(WM15, BD)	PBS
2	CD44 (J173, BC)	CLL-1 /CLEC12a (50C1, BD)	CD13	CD56 (N901, BC)
3	CD7 (M-T701, BD)	TIM-3(344823, R&D)	CD13	CD117 (104D2D1, BC)
4	CD2 (MT910, DC)	CD133 (AC133/1, MB)	CD13	CD117
5	CD36 (CLB-IVC7, Sanquin)	CD123 (9F5, BD)	CD13	CD33 (D3HI60,251, BC)
6	CD11b (Bear1, BC)	CD96 (6F9, BD)	CD13	CD117
7	CD15 (MMA, BD)	CD22 (S-HCL-1, BD)	CD13	CD117

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

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

Cut-off %	*0*	*0,000*	*0,000*	*0,000*	*0,000*	*0,000*
Patients (n) above cutoff	78	78	64	53	48	34
Patients (n) below cutoff	164	164	178	189	194	208
p-value	0.000	0.000	0.000	0.000	0.000	0.000

APC	APC-H7
PBS	PBS
CD13 (WM15, Pharmigen)	HLA-DR (L243, BD)
CD22 (S-HCL-1, BD)	CD19 (SJ25C1, BD)
CD11b (D12, BD)	CD14 (MoP9, BD)
CD117 (104D2, DC)	HLA-DR
CD38 (HB7, BD)	HLA-DR
CD38	CD19

APC	APC-H7	HV450	KO
CD14 (Mop9, BD)	PBS	CD34 (8G12,BD)	CD45 (J.33, BC)
CD38 (HB7, BD)	HLA-DR (L243, BD)	CD34	CD45
CD38	CD19 (SJ25C1, BD)	CD34	CD45
CD38	CD19	CD34	CD45
CD38	CD14 (MoP9, BD)	CD34	CD45
CD38	CD14	CD34	CD45
CD38	HLA-DR	CD34	CD45

tubes 2-7. Associated clones are given in brackets. BC, Beckman Coulter; BD, Becton Dickinson; DC, DakoCytomation; MB, Miltenyi Biotec; PBS, phosphate buffered saline; R&D, R&D systems.

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

Table S5. Multivariate analyses for therapy response, overall survival, and event free survival at time of diagnosis

		Multivariate analysis			Multivariate analysis			Multivariate analysis		
		CR			Death			Event*		
Variable	No.	OR	95% CI	p	HR	95% CI	p	HR	95% CI	p
Total	518									
CD34 status										
CD34-negative	64									
CD34-positive LSC^{low}	294	0.65	0.13-3.19	0.591	1.94	1.11-3.39	0.02	1.36	0.87-2.14	0.183
CD34-positive LSC^{high}	160	0.45	0.09-2.25	0.329	2.38	1.34-4.25	0.003	1.81	1.13-2.90	0.014
Age (continuous)	518	0.94	0.91-0.97	<0.001	1.03	1.02-1.04	<0.001	1.02	1.01-1.03	<0.001
AML type										
De Novo	424									
Secondary	54	0.67	0.30-1.45	0.307	1.12	0.78-1.62	0.547	1.04	0.72-1.49	0.849
MDS	40	2.00	0.64-6.23	0.23	1.08	0.71-1.64	0.734	1.00	0.67-1.49	0.999
Risk group ELN										
Good	140									
Intermediate-I	175	0.48	0.17-1.37	0.171	1.78	1.12-2.82	0.014	2.00	1.32-3.01	0.001
Intermediate-II	94	1.39	0.40-4.80	0.605	1.32	0.81-2.15	0.268	1.76	1.14-2.74	0.011
Adverse	109	0.33	0.12-0.94	0.037	3.07	1.93-4.89	<0.001	3.67	2.41-5.60	<0.001
FLT3/NPM1 status										
$FLT3^{wt}$ / $NPM1^{wt}$	315									
$FLT3^{wt}$ / $NPM1^{mut}$	90	4.78	0.85-26.71	0.075	0.61	0.35-1.07	0.086	0.57	0.35-0.93	0.024
$FLT3^{mut}$/ $NPM1^{wt}$	54	1.41	0.53- 3.75	0.494	0.76	0.51-1.14	0.189	0.83	0.57-1.20	0.318
$FLT3^{mut}$/ $NPM1^{mut}$	59	2.94	0.93- 9.25	0.066	0.89	0.57-1.40	0.627	0.77	0.50-1.17	0.214
WBC at diagnosis (x 10^9/L)										
<20	351									
20-100	130	0.63	0.30-1.31	0.212	1.63	1.19-2.23	0.002	1.67	1.25-2.21	<0.001
>100	37	0.58	0.17-2.02	0.391	1.96	1.20-3.20	0.007	2.34	1.51-3.64	<0.001

Table S5. (continued)

		Multivariate analysis			Multivariate analysis			Multivariate analysis		
		CR			Death			Event		
Variable	No.	OR	95% CI	p		95% CI	p		95% CI	p
First consolidation therapy										
None	166									
Cycle 3	144	NA	NA	NA	0.45	0.28-0.70	0.001	0.7	0.46-1.06	0.092
Autologous SCT	48				0.49	0.28-0.85	0.011	0.67	0.40-1.10	0.111
Allogeneic SCT	160				0.47	0.34-0.65	<0.001	0.44	0.30-0.64	<0.001
Treatment arm										
control group	254									
clofarabine	264	1.53	0.87-2.72	0.142	0.79	0.62-1.01	0.055	0.75	0.60-0.95	0.015

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 S7. Changes of LSC status between diagnosis and follow-up for 277 AML patients

LSC status at diagnosis	LSC status after 2nd cycle	N (number)	% of total	3yr OS % (SE)	3yr EFS % (SE)
CD34-negative	LSCneg	32	12	65 (9)	56 (9)
CD34pos LSC^{low}	LSCneg	119	43	69 (5)*	60 (5)*
CD34pos LSC^{high}	LSCneg	40	14	49 (8)$^{\#}$	40 (8)$^{\#}$
CD34-negative	LSCpos	8	3	0 (0)	0 (0)
CD34pos LSC^{low}	LSCpos	45	16	44 (8)*	39 (8)*
CD34pos LSC^{high}	LSCpos	33	12	19 (10)$^{\#}$	21 (7)$^{\#}$
Total		277	100%		

This table shows the different CD34/LSC categories at diagnosis and LSC status after the 2nd cycle. Diagnosis LSC results were available in 277/302 (92%) patients. This table shows that 8 CD34-negative patients (with no CD34+CD38-LSC at diagnosis) convert to LSC^{pos} after the 2nd cycle. Moreover, 45 patients with low LSC levels at diagnosis convert to LSC^{pos} after treatment. These patients seem to have a worse outcome as compared to CD34-negative and CD34-positive patients with low LSC levels who are LSC^{neg} after treatment. For both the CD34-pos LSC^{low} and CD34-pos LSC^{high} group a significant difference based on LSC status after treatment was found (*p=0.001 and p=0.002, $^{\#}$p=0.009 and p=0.044 for both OS and EFS respectively. EFS, event free survival; OS, overall survival; SE, standard error.

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 S8). 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.

Table S8. Multivariate analyses for relapse incidence, overall survival, and event free survival for CR patients

		Cumulative relapse incidence			Death incidence			Event incidence		
Variable	No.	HR	95% CI	p	HR	95% CI	p	HR	95% CI	p
Total	212									
CD34 status										
CD34-negative	30									
CD34-positive LSC low	120	0.67	0.34-1.32	0.242	1.32	0.54-3.22	0.543	0.81	0.41-1.60	0.538
CD34-positive LSC high	62	1.10	0.55-2.17	0.789	1.65	0.65-4.17	0.294	1.14	0.55-2.34	0.727
Age	212	1.01	1.00-1.03	0.138	1.02	1.00-1.05	0.035	1.02	1.00-1.04	0.034
AML type										
De Novo	179									
Secondary	21	1.10	0.55-2.22	0.783	1.30	0.65-2.61	0.464	0.96	0.49-1.87	0.898
MDS	12	0.91	0.41-2.01	0.811	0.94	0.39-2.29	0.897	1.02	0.45-2.30	0.960
CR reached										
Early CR#	180									
Late CR	32	1.44	0.77-2.67	0.251	1.60	0.86-2.99	0.139	1.66	0.92-2.99	0.092
MRD/LSC status										
MRD^{neg} LSC^{neg}	121									
MRD^{pos} LSC^{neg}	23	2.00	1.04-3.87	0.039	0.98	0.47-2.03	0.960	1.46	0.77-2.76	0.245
MRD^{neg} LSC^{pos}	51	1.61	0.87-2.97	0.127	1.11	0.63-1.97	0.719	1.34	0.80-2.26	0.269
MRD^{pos} LSC^{pos}	17	5.89	3.32-10.47	<0.001	3.62	1.86-7.04	<0.001	4.24	2.24-8.02	<0.001
HOVON risk group										
Good	75									
Intermediate	43	2.99	1.39-6.44	0.005	1.62	0.71-3.68	0.250	1.84	0.90-3.75	0.092
Poor	60	2.95	1.29-6.74	0.010	2.30	0.98-5.42	0.057	1.76	0.83-3.72	0.14
Very poor	34	8.37	3.92-17.83	<0.001	8.40	3.58-19.71	<0.001	5.63	2.68-11.84	<0.001

Table S8. (continued)

		Cumulative relapse incidence			Death incidence			Event incidence		
Variable	No.	HR	95% CI	p	HR	95% CI	p	HR	95% CI	p
FLT3/NPM1 status										
FLT3^{wt} / NPM1^{wt}	116									
FLT3^{wt} / NPM1^{mut}	45	0.51	0.23-1.13	0.099	0.75	0.29-1.95	0.559	0.46	0.21-1.03	0.060
FLT3^{mut}/ NPM1^{wt}	26	1.12	0.53-2.37	0.762	1.47	0.73-2.95	0.276	1.09	0.56-2.09	0.806
FLT3^{mut}/ NPM1^{mut}	25	0.86	0.37-2.00	0.721	1.04	0.47-2.32	0.918	0.81	0.40-1.67	0.577
WBC at diagnosis (x 10^9/L)										
<20	141									
20-100	52	3.07	1.82-5.20	<0.001	2.62	1.53-4.48	<0.001	2.79	1.72-4.53	<0.001
>100	19	2.66	1.32-5.35	0.006	1.84	0.82-4.10	0.138	2.92	1.46-5.83	0.002
First consolidation therapy										
None	26									
Cycle 3	81	0.96	0.47-1.97	0.909	0.37	0.17-0.78	0.009	0.83	0.40-1.71	0.611
Autologous SCT	26	0.53	0.25-1.13	0.101	0.33	0.14-0.78	0.012	0.57	0.25-1.29	0.179
Allogeneic SCT	79	0.22	0.11-0.44	<0.001	0.28	0.15-0.52	<0.001	0.33	0.16-0.66	0.002
Treatment arm										
control group	112									
clofarabine	100	0.70	0.45-1.10	0.120	0.61	0.38-0.97	0.035	0.68	0.45-1.04	0.077

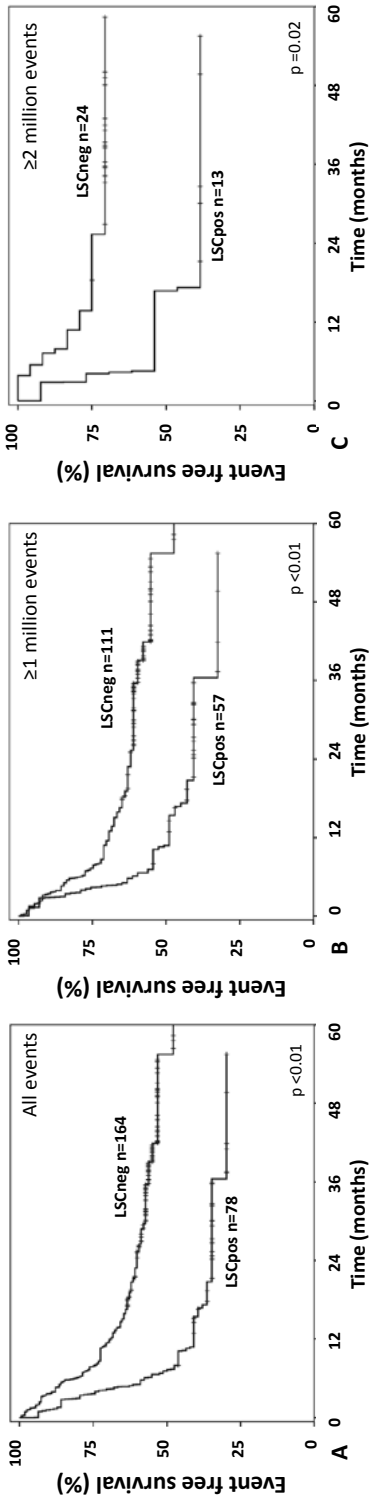


Figure S2. LSC results for event free survival depending on the number of WBC events measured. Figure S2A shows EFS for the group of 242 AML patients when all patient samples are included, independent on the number of WBCs acquired using flow cytometry (A). Results are similar when only patient samples in which ≥ 1 million WBCs were acquired, were included (n=168). Similar results were also found when including only cases with ≥ 2 million WBCs acquired, 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.

REFERENCE LIST

1. Löwenberg B, Pabst T, Maertens J, van Norden Y, Biemond BJ, Schouten HC, *et al*. Therapeutic value of clofarabine in younger and middle aged (18 – 65 yrs) adults with newly diagnosed AML. *Blood* 2017; **129**: 1636–45.

2. Terwijn M, van Putten WLJ, Kelder A, van der Velden VHJ, Brooimans RA, Pabst T, *et al*. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J Clin Oncol* 2013; **31**: 3889–97.

3. Feller N, van der Velden VHJ, Brooimans R a, Boeckx N, Preijers F, Kelder A, *et al*. Defining consensus leukemia-associated immunophenotypes for detection of minimal residual disease in acute myeloid leukemia in a multicenter setting. *Blood Cancer J* 2013; **3**: e129.

4. Zeijlemaker W, Kelder A, Wouters R, Valk PJM, Witte BI, Cloos J, *et al*. Absence of leukaemic CD34+ cells in acute myeloid leukaemia is of high prognostic value: a longstanding controversy deciphered. *Br J Haematol* 2015; **171**: 227–38.

5. Terwijn M, Zeijlemaker W, Kelder A, Rutten AP, Snel AN, Scholten WJ, *et al*. Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. *PLoS One* 2014; **9**: e107587.

CHAPTER 4

LEUKEMIC STEM CELL LOAD AT DIAGNOSIS PREDICTS THE DEVELOPMENT OF RELAPSE IN YOUNG ACUTE MYELOID LEUKEMIA PATIENTS

Diana Hanekamp, Barbara Denys, Gertjan J.L. Kaspers, Jeroen G. te Marvelde, Gerrit Jan Schuurhuis, Valerie De Haas, Barbara De Moerloose, Eveline S. de Bont, C. Michel Zwaan, Anja de Jong, Barbara Depreter, Tim Lammens, Jan Philippé, Jacqueline Cloos, Vincent H.J. van der Velden

Br J Haematol 2018 Nov;183(3):512-516

ABSTRACT

The most effective strategy to improve the dismal outcome of pediatric acute myeloid leukemia (AML) is to prevent relapse. In this study we explored whether leukemic stem cells (LSC) in pediatric AML samples taken at time of diagnosis can identify patients at high risk of relapse. LSC, defined as CD34+CD38-/low cells with aberrant expression of CD123, CD7, CD56 and/or CD2 were characterized for 86 patients. Sixteen per cent of the patients were found CD34-negative and therefore by definition LSC negative. CD34-negative patients showed the highest 5-year relapse free survival (75%). In contrast, LSC were detected in 93% of CD34-positive patients. High LSC burden, determined by receiver operating characteristic analysis as ≥17.2% of the CD34+ population, was significantly associated with the occurrence of relapse (43% 5-year relapse probability in LSClow vs. 78% 5 year relapse probability in LSChigh, p=0.05). After multivariate adjustment LSC frequency, white blood cell count and core binding factor AML were significant predictors of relapse (HR 4.1, 95% CI 1.1-15.2, p=0.04, HR 2.5, 95% CI 1.1-5.8, p=0.03 and HR 0.3, 95% CI 0.0-0.8, p=0.03 respectively). These results warrant the incorporation of this relatively simple method to determine CD34+/CD38-/low LSC frequency in future risk group stratification.

BACKGROUND

Pediatric acute myeloid leukemia (AML) is an aggressive haematological malignancy comprising 15-20% of all acute leukemia's in children, adding up to an average of 40 children per year in the Netherlands and Belgium combined (population of 28 million people)ze[1]. Despite the fact that prognosis has improved significantly over the past decades, the overall survival (OS) of paediatric AML remains relatively low (~70%)[2]. Nowadays 90-95% of patients achieve complete remission (CR)[3,4]. Nonetheless, up to 40% of the patients eventually relapse due to the outgrowth of persisting leukemic cells which are not eradicated by therapy. Since currently therapy options to eliminate these resistant cells are limited, the most effective strategy to improve the dismal outcome of AML patients is to refine risk group stratification in order to prevent relapse.

As AML is a heterogeneous disease in terms of chromosomal aberrancies and gene mutations, clinical prognosis is extremely variable. In most leukemia study groups, risk group stratified therapy is based on a combination of cytogenetic and molecular aberrations[5,6], as well as on response to treatment including the detection of minimal residual disease (MRD)[7,8]. However, it is well recognized that these (cyto)genetic signatures cannot always predict the outcome in individual patients[9]. For this reason, new stratification parameters are needed to allow more accurate therapeutic decision-making.

Over the last decades, accumulating evidence suggests that AML develops in a hierarchical structure: it originates from hematopoietic stem cells (HSCs) which are transformed to leukemia-initiating cells (LICs), often referred to as leukemic stem cells (LSCs)[10–12]. In many aspects, LSC and HSC are much alike, sharing features such as quiescence, self-renewal capacity and the expression of drug efflux proteins[13,14]. LSCs are of fundamental interest for AML patients at risk for relapse as these cells are thought to be resistant to current standard treatments and therefore supposed to be responsible for initiation of the relapse. In adult AML, the frequency of LSCs at diagnosis has shown to be of importance for clinical outcome[14].

Although LSCs reside in different compartments, the $CD34^+/CD38^-$ or $CD34^+/CD38^{dim}$ fractions are considered to be most enriched for LSCs[15]. The association between these fractions and risk of relapse and OS in paediatric AML was shown in one previous study with a limited number of patients (n=17)[16]. Since $CD34^+/CD38^-$ comprise both HSC and LSC, little is known about the prognostic impact of the LSC pool within this immature compartment. The LSC can be distinguished from normal $CD34^+/CD38^-$ cells on the basis of aberrant expression of cell surface molecules using multicolour flow cytometry[14,17–19].

In this study we determined LSC burden at diagnosis by defining aberrantly expressed cell surface markers in $CD34^+/CD38^{-/low}$ cells using a single 8-color LSC tube. LSC measurements were implemented in routine diagnostics using a minimal combination of markers. This specific identification of LSC is used to give further insight in the prognostic role of LSC burden in paediatric AML patients at time of diagnosis.

PATIENTS AND METHODS

PATIENTS

From the >100 pediatric AML patients diagnosed between 2010-2014 and treated according to the Dutch-Belgian paediatric AML-01 (DB AML-01) protocol[20] (EudraCT 2009-014462-26), 86 patients were analysed with an additional stem cell identification tube at diagnosis. Next to stem cell identification, standard morphological, immunological, cytogenetic and molecular criteria according to the WHO 2008 classification was included in immunophenotypic characterization at diagnosis. Exclusion criteria were AML with PML-RARA, secondary AML, untreated refractory anaemia with excess blasts (RAEB) and RAEB in transformation.

PATIENT MATERIALS

Bone marrow (BM) samples were taken at diagnosis after informed consent was obtained from patients, their parents or guardians. Samples from AML patients diagnosed at different Belgian were directly transported to the reference laboratory at the Ghent University Hospital (Ghent, Belgium). Samples from Dutch paediatric AML patients (n=69) were either analysed at the department of Immunology, Erasmus MC, Rotterdam or sent to the Dutch reference centre; the Dutch Childhood Oncology Group (DCOG, The Hague, The Netherlands).

FLOW CYTOMETRY

Flow cytometric LSC characterization was performed on BM from AML patients collected at diagnosis. Fifty microliter of BM cell suspension was incubated with monoclonal antibody combinations for 15 minutes in the dark, followed by red blood cell lysis using FACSLysing solution (Becton Dickinson, BD, San Jose, CA, U.S.A.). Data were acquired on a FACSCanto-II flow cytometer with EuroFlow instrument settings[21], aiming to measure a minimum of 1 million WBC (median: 803963, range: 16414-1396231). All protocols and antibody panels were standardized between the three participating laboratories, with the exception of CD45 and HLADR, which were conjugated with HV450 and HV500 in the Belgium laboratory. Samples were both analysed by the different laboratories and centrally by the laboratory of the VUMC for blast percentage, CD34 percentage and LSC percentage (Suppl. Fig.1). To identify CD34-negative patients, the more stringent definition published last year[22] was adopted. Patients with <1% CD34 expression on blasts and no marker-positive $CD34^{+}/CD38^{-}$ cells, were defined CD34-negative; all other patients were classified as CD34-positive. LSC were defined as $CD34^{+}/CD38^{-}/CD45^{low}$ cells with aberrant expression of the most commonly found leukemia specific markers[19] CD123 and/or CD7 and/or CD56/CD2 (example shown in Figure 1) with at least three specific events. Antibodies for CD56 and CD2 were used in the same fluorescence channel and therefore differences in expression of these two markers could not be determined. Details

on antibodies are shown in Suppl. Table I. Analysis was performed using Infinicyt software (Cytognos SL, Salamanca, Spain).

STATISTICAL ANALYSIS

Statistical analyses were performed using IBM SPSS Statistics software, version 20 (IBM Netherlands BV, Amsterdam, The Netherlands) and GraphPad prism version 6.02 (GraphPad Software, La Jolla California USA). Chi-square test was used when comparing categorical variables. Non-parametric Mann-Whitney U test was used for continuous variables. Receiver operating characteristic (ROC) analysis was performed to find the best cut-off value defining LSC^{high}. Cut-off value defining CD34-negative[22] patients and cut-offs

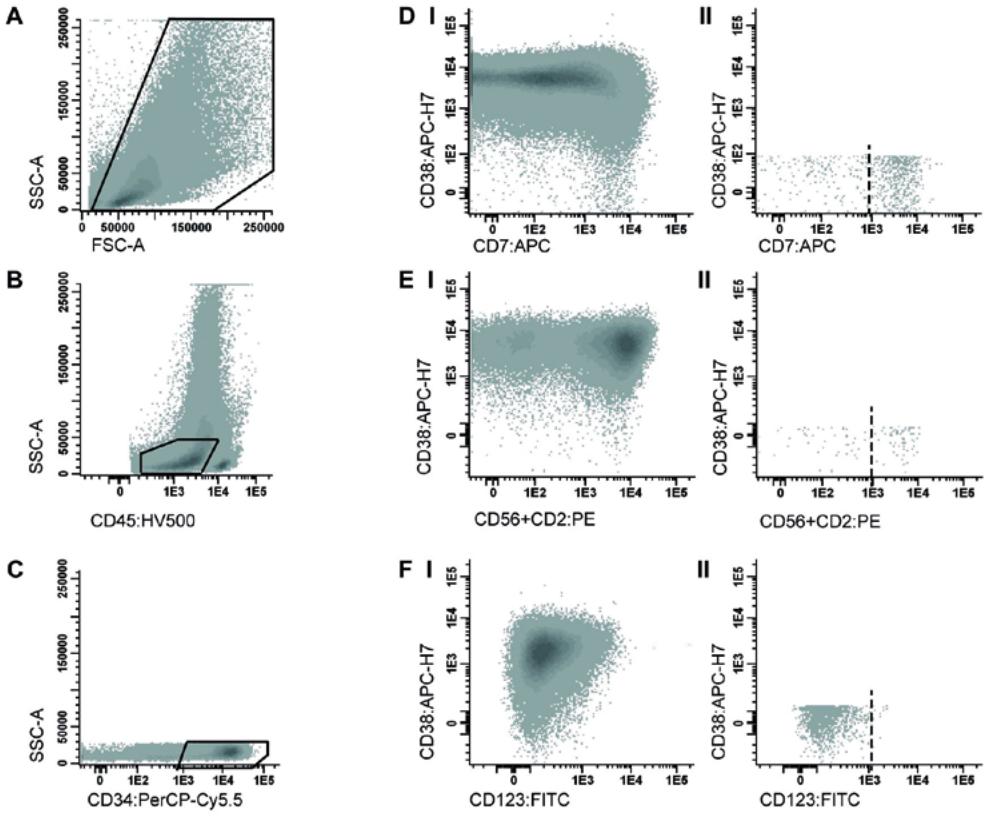


Figure 1. Examples of marker defined CD34$^+$/CD38$^-$ leukemic stem cells (LSC) in three patient samples at time of diagnosis. After labelling diagnostic samples with the appropriate antibody combination, (A) the white blood cells are gated. Subsequently the (B) CD45^{dim} blasts and (C) CD45^{low}/CD34$^+$ compartment are gated. (D) Patient had a clear CD38$^-$ 'tail' (D II) with a distinct aberrant population of CD7 expressing LSCs. (E I) Patient with a notably less prominent CD38$^-$ compartment, (E II) but clear separation between marker positive-LSCs and marker negative cells based upon aberrant CD2/CD56 expression. (F) Patient in which no LSC based upon any of the aberrant markers (here CD123 is shown) could be defined.

defining patients with either high or low $CD34^+/CD38^-$ fraction[16] were based on literature. For Kaplan-Meier analysis, OS was calculated from date of diagnosis to death or date to last follow-up, relapse free survival (RFS) from first remission to date of first relapse or date to last follow-up. Cumulative incidence of developing a relapse (CIR) and cumulative incidence of death (CID) due to high frequency LSC, were estimated and compared between groups using Gray's test[23]. Log-Rank test was performed comparing survival distributions between groups. In univariate analysis for RFS, the following variables were tested: gender, age, WBC count, proportion of blasts cells in BM and PB, time to achieve first CR, FAB classification, cytogenetics and molecular aberrations. Subsequently, in multivariate analysis, using Cox regression, the impact of LSC burden was weighed with respect to other known prognostic risk factors with significant influence on RFS found by univariate analysis.

RESULTS

COHORT STATISTICS

Available patients' characteristics are summarized in Table I. In this cohort CR was achieved in 80 patients (93%) and 5-year relapse-free survival (RFS) was 54%. A second CR was achieved in 54% of relapsed patients finally resulting in 5-year OS of 84%. Six patients (7%) never reached CR, with one early death after only 18 days post diagnosis. The blast percentage in the with flow cytometry evaluated samples ranged from 1% to 95% (median: 61%).

ABSENCE OF LSC IN CD34 NEGATIVE PATIENTS

In our cohort, median $CD34^+$ fraction within the blast compartment was 17.9% (range, 0%-99%) (Suppl. Fig. 1b). Based on expression of CD34 of the blast cells, AML can be divided into two broad subtypes: CD34-negative and CD34-positive AML patients. CD34-negative patients are characterized by <1% of $CD34^+$ blast cells with absence of leukemic cells within this population[22], and are usually associated with a better prognosis[24,25]. In our cohort, 13 patients (15%) were found CD34-negative ($CD34^+$ blasts median 0.08%, 0%-0.96%). Median CD34 percentage of blasts determined in the CD34-positive group was 38% (range 0.02%–99.12%).

IDENTIFICATION AND PREVALENCE OF LSC

LSC were observed in 68 of the 73 CD34+ patients (93%), with percentages between $1x10^{-4}$% and 27% of all leukocytes (median: 0.1%). The median ratio LSC, calculated as the percentage aberrant $CD34^+/CD38^-$ of the complete $CD34^+/CD38^-$ compartment, was 68% (range, 0.3-100%). Moreover, when calculating this LSC-load as a percentage of total $CD34^+$ cells, the median was 0.9% (range, $4x10^{-4}$ - 86%). In our cohort we found both patients with complete marker-negative $CD34^+/CD38^-$ compartments (n=5) as well

Table 1. Patient's characteristics

		N	Median (Range)
Age (years)		86	6.72 (0-16)
White blood cell count (x 10^9 /l)		81	24.30 (2-449)
Blast in bone marrow at diagnosis (percentage)		81	74.50 (0-98)
CD34 expressing cells (percentage)			
White blood cells		86	5.90 (0-93)
Blast cells		86	17.85 (0-99)
Time (days) to			
First complete remission		80	31.50 (12-400)
First relapse		80	53.00 (2-127)
		N	%
Gender (N=85)	Male	42	48.8
	Female	43	50.0
Mutation status (N=86)	FLT3	16	18.6
	NPM1	7	8.1
	t(8;21)	9	10.5
	Inv(16)	7	8.2
Karyotype (N=85)	MML	3	3.6
	Complex	2	2.3
Cytogenetic risk group (N=86)	Good	2	2.3
	Intermediate	65	75.6
	Poor	12	13.9
French American British-type (N=83)	M0	3	3.5
	M1	13	15.1
	M2	14	16.3
	M4	18	20.9
	M5	19	22.1
	M6	1	1.2
	M7	5	5.8
	unknown	10	11.6

as patients with complete LSC $CD34^+/CD38^-$ compartments (n=4). Patients with a low proportion of $CD34^+CD38^-$ cells (cut off: <0.68%)[16] had significantly less LSC than patients with a high proportion $CD34^+CD38^-$ cells (median 25% (0%-100%) vs. 95% (0%-100%), respectively; P= 0.001). In 14% of the patients CD7 gave the best separation between HSC and LSC. In 19% of the patients the combination CD56-CD2 was most discriminative. In the majority (67%) LSC were best defined by CD123 overexpression (data not shown).

LSC-LOAD AT DIAGNOSIS AS PROGNOSTIC FACTOR

We first evaluated the prognostic significance of LSC frequencies. Results of the ROC analysis showed an optimal cut-off value for 'high' LSC frequency at 17.2%. Patients were

thereafter defined as either LSChigh (≥17.2% LSC at diagnosis) or LSClow (<17.2% LSC at diagnosis). Kaplan-Meier survival analysis, as depicted in Figure 2, showed a significant association between a high LSC-load and impaired RFS (39.4% relapses in LSClow vs. 77.8% relapses in LSChigh, Plogrank= 0.026). Competing risk analysis for CIR showed a trend towards a significant difference between LSC^{low} and LSC^{high} (Gray's test, P= 0.058) but not for CID (Gray's test, P= 0.79) (Suppl. Figure 3). Univariate Cox regression analysis showed that high LSC frequency is associated with increased risk of relapse (Hazard ratio [HR] 2.5, 95% confidence interval [CI] 1.08-5.73) (Suppl. Table III).

CD34- NEGATIVITY IS ASSOCIATED WITH HIGHER RFS

We examined the prognostic relevance of the CD34 status. Kaplan-Meier analysis showed enhanced, but not significant, RFS for the CD34-negative group as compared to the CD34-postive group ($p_{logrank}$= 0.20) (Figure 3a). When stratifying patients in CD34-negative (no LSC), CD34-positive/LSC^{low} and CD34-positve/LSC^{high} groups, the CD34-negative patient group showed a better RFS compared to both CD34-positive/LSC^{low} and CD34-positve/LSC^{high} patients (Figure 3b; $P_{logrank}$= 0.05). Therefore, combining CD34-status and LSC-frequencies allows identification of patients with different risks of relapse.

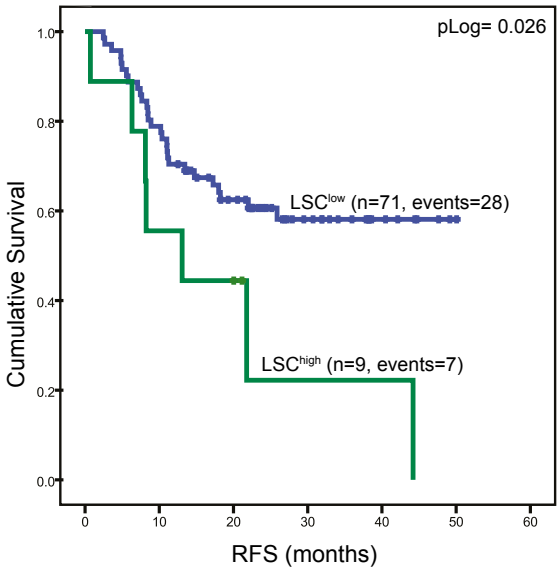


Figure 2. Correlation between leukemic stem cell (LSC) burden and relapse-free survival (RFS). Kaplan-Meier analysis showing patients grouped in low or high LSC burden, LSClow/LSChigh respectively. Cutoff used is 17.2%. Patients with high number of leukemic stem cells at time of diagnosis had significant more risk of developing relapse compared to patients with less LSC (77.8% vs. 39.4%) (Plogrank= 0.026).

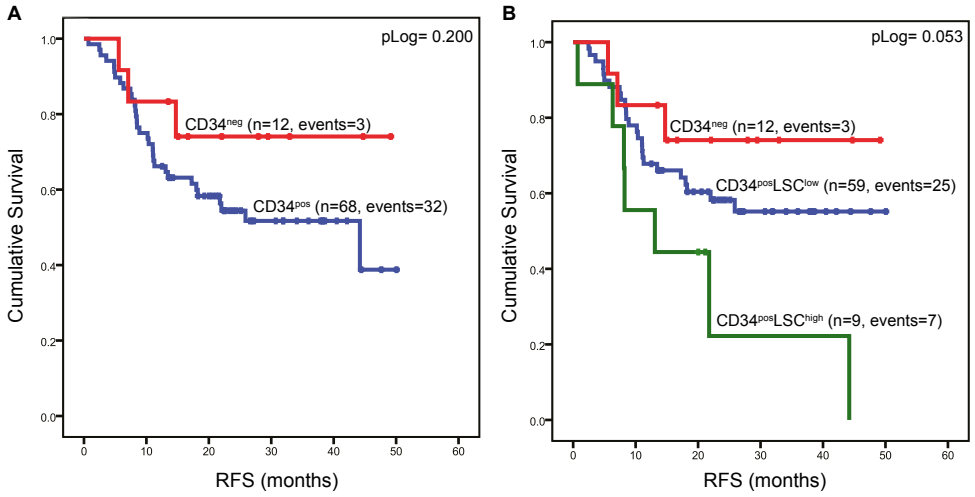


Figure 3. Kaplan-Meier analyses showing prognostic impact of CD34 expression on RFS. (A) Kaplan-Meier plot showing enhanced RFS in patients with low number of CD34+ blasts (<1%) and, more importantly without presence of LSC within this compartment ($P_{logrank}$= 0.200)[22]. (B) Subsequently, as CD34-negative patients previously were regarded as LSC^{low}, combining CD34-status with LSC burden will divide patients into three groups. Using this method, it is possible to identify patients with higher risk of relapse and moreover patients at distinct less risk ($P_{logrank}$= 0.053). LSC, leukemic stem cell; RFS, relapse-free survival.

OTHER FACTORS ASSOCIATED WITH HIGHER RFS

Additional univariate analyses of prognostic factors for relapse (summarized in Suppl. Table III) showed that, in addition to LSC load, high WBC count at diagnosis (P= 0.03), core binding factor (CBF) AML (P=0.05) and percentage blasts in PB (P= 0.06) had impact on RFS. Multivariate analysis showed both LSC load (P= 0.04), CBF-AML (P= 0.03) and high WBC count (P= 0.03) to have independent prognostic influence on RFS (Suppl. Table IV). Probably due to relatively low patient numbers, previously defined risk classifiers, such as FAB classification, different karyotype classes and other molecular aberrations, were not significant in our analysis. Differences in patient characteristics between LSC^{low} and LSC^{high} groups are summarized in table II.

DISCUSSION

In the current study we show that the frequency of LSC at diagnosis in paediatric AML patients can distinguish patients more likely to fail current treatment regimens as these patients develop significantly more relapses. In paediatric AML, a single study has been published about the role of LSC[16]. Witte et al. have shown that in a small cohort of paediatric AML patients (n=17) with a higher proportion of $CD34^{+}/CD38^{-}/CD45^{low}$ cells corresponded with lower EFS and a trend towards a lower OS[16]. When applying the same

Table 2. Patient's characteristics per group

		LSC^{low}		LSC^{high}		
		N	Median (Range)	*N*	Median (Range)	P-value
Age (years)		77	7.05 (0-20)	9	9.58 (1-16)	0.838
White blood cell count (x 10^9 /l)		73	37.00 (2-449)	8	14.50 (2-136)	0.217
Blast in bone marrow at diagnosis (percentage)		72	74.50 (0-98)	9	71.00 (12-95)	0.652
CD34 expressing cells (percentage)						
White blood cells		77	5.46 (0-93)	9	8.40 (2-50)	0.521
Blast cells		77	17.46 (0-99)	9	26.04 (3-83)	0.667
Time (days) to						
First complete remission		71	32.00 (13-400)	9	42.00 (12-68)	0.322
First relapse		28	22.08 (6-66)	7	20.92 (2-112)	0.793
		N	%	*N*	%	
Gender	Male	36	46.75	6	66.67	0.274
	Female	40	51.95	3	33.33	
Mutation status	FLT3	14	18.18	2	22.22	0.768
	NPM1	7	9.09	0	0.00	0.345
	t(8;21)	8	10.39	1	11.11	0.947
	Inv(16)	5	6.49	2	22.22	0.106
Karyotype	MML	2	2.60	0	0.00	0.677
	Complex	2	2.60	0	0.00	
Cytogenetic risk group	Good	2	2.60	0	0.00	0.590
	Intermediate	57	74.03	8	88.89	
	Poor	12	15.58	0	0.00	
French American British-type	M0	3	3.90	0	0.00	0.273
	M1	12	15.58	1	11.11	
	M2	10	12.99	4	44.44	
	M4	15	19.48	3	33.33	
	M5	19	24.68	0	0.00	
	M6	1	1.30	0	0.00	
	M7	5	6.49	0	0.00	
	unknown	9	11.69	1	11.11	

method to our cohort, we observed that patients with higher proportion of $CD34^+/CD38^-$ cells have a worse OS compared to patients with less $CD34^+/CD38^-$ cells (Suppl. Figure 2b), while this was not statistically significant for RFS (Suppl. Figure 2a).

For this study there were no selection criteria other than the availability of stem cell flow cytometry data. The cohort statistics of the study population versus the total DB-AML01 study population are CR rate: 93% versus 90%; 4-year survival probability: 84% versus 78%. As the additional flow cytometry measurement acquires extra cells, we could reason

that only samples with enriched cell number are included. However, a high WBC count at diagnosis is considered an prognostic factor for poor outcome[26] whereas our cohort had a better outcome.

The $CD34^{+}/CD38^{-}/CD45^{-/low}$ compartment reflects the most immature cell fraction and it contains both normal HSCs and LSCs. In this study we have shown that with the addition of four surface markers (which could be added to the necessary backbone markers CD45, CD38, CD34 in an 8-color flow panel), LSCs could be identified based on aberrant expression.

As LSC have heterogeneous marker expressions, it is suggested that multiple LSC markers are necessary to discriminate LSC from normal HSC[14,17,19,27]. Fifteen LSC markers were tested by Zeijlemaker et al. ranking them on their performance in distinguishing LSC from HSC with high specificity and high sensitivity, concluding that 13 markers are needed for adequate identification[19]. In the present study, 8-color flow cytometry was used. As a result, present research is limited in sensitivity: 14 patients (16%) were identified with HSC with absence of LSC. Inclusion of other LSC markers is warranted to confirm the absence of LSC. In adult AML patients, CD123 was found to be the best LSC marker, followed by CD7 (7^{th} place), CD56 (10^{th}) and CD2 (13^{th}). Our previous research supports that CD123 is most sensitive in identifying LSC[17,19,28]. Furthermore, as the order of markers is the same in our study, we suggest similar LSC surface properties in paediatric and adult AML.

In this study, 11% of all patients were classified as high risk for developing relapse based on LSC load with high specificity determined by ROC analysis (96%, not shown), while in adult AML, over 50% of patients could be classified as LSC^{high}[14]. Of all nine paediatric patients with high LSC load, seven patients eventually relapsed (78%), having increased LSC load at time of relapse with the same aberrant marker expression (data not shown). Similar results have been reported in a study with adult AML[29].

Of all LSC^{low} patients, 39% developed a relapse. Hence, we suggest that the current multiparameter flow cytometry approach must be further improved applying more LSC antibodies simultaneously[19,30]. At the time of writing, two LSC^{high} patients did not relapse. In addition to technical improvements of LSC measurements (i.e. addition of markers, higher cell numbers), LSC measurements during therapy (for instance at MRD time points[22]) could aid in accurately assessing patient prognosis as outcome also greatly depends on many different factors during therapy.

Next to increasing sensitivity of LSC identification and thereby identifying patients at higher risk to relapse, defining a group of patients with very low risk to relapse can possibly spare these patients from undergoing more intensive treatment. Low numbers of $CD34^{+}$ cells is a prognostically favourable variable as it is in adult AML. These results need to be confirmed in a different cohort taking into account suggested improvements.

Despite the fact that the number of patients included in this study is relatively small compared to studies in adult AML, our results do confirm that the frequency of LSC at diagnosis in paediatric AML patients can distinguish patients more likely to fail current

treatment regimens. Next to this, as in adults, CD34-negative patients were found to have better RFS. Unfortunately, drawback of the limited patient number is the unfeasibility in the present to further establish the association between LSC burden and well-known clinical and biological features.

As survival after relapse is significantly poorer, one of the most hopeful approaches to increase the OS rate is the prevention of relapse. The incorporation of $CD34^{+}/CD38^{-}$ LSC frequency and conventional MRD measurement are warranted for future risk group stratification. In addition, identifying patients at risk for relapse using LSC load, may direct the use of more personalized targeted medicine or experimental therapies targeting the LSCs[31].

ACKNOWLEDGEMENTS

Grant of the Clinical Research Fund (Ghent University Hospital) for B. Denys; Grant of the Foundation against Cancer, Belgium, for T. Lammens; grant of Cancer Plan, action 29 (KP_29_020) for J. Philippé. This study was partially financially supported by KIKA (grant 70, "Prognostic value of "Minimal Residual Disease" detection in children with acute myeloid leukemia".) (VHJvdV, VdH). The research for this manuscript was (in part) performed within the framework of the Erasmus Postgraduate School Molecular Medicine.

FURTER SUPPLEMENTARY TEXT/TABLES/FIGURES

Further supplemental data can be found online at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fbjh.14991&file=bjh14991-sup-0001-SupInfo.pdf

REFERENCES

1. de Rooij, J. et al. Pediatric AML: From Biology to Clinical Management. J. Clin. Med. 4, 127–149 (2015).

2. Zwaan, C. M. et al. Collaborative efforts driving progress in pediatric acute myeloid leukemia. J. Clin. Oncol. 33, 2949–2962 (2015).

3. Kaspers, G. J. L. et al. Improved outcome in pediatric relapsed acute myeloid leukemia: Results of a randomized trial on liposomal daunorubicin by the international BFM study group. J. Clin. Oncol. 31, 599–607 (2013).

4. Rubnitz, J. E., Gibson, B. & Smith, F. O. Acute myeloid leukemia. Hematol. Oncol. Clin. North Am. 24, 35–63 (2010).

5. Faulk, K., Gore, L. & Cooper, T. Overview of therapy and strategies for optimizing outcomes in de novo pediatric acute myeloid leukemia. Pediatric Drugs 16, 213–227 (2014).

6. Pui, C., Carroll, W., Meshinchi, S. & Arceci, R. J. Biology, Risk Stratification, and Therapy of Pediatric Acute Leukemias: An Update. J. Clin. Oncol. 29, 551–565 (2011).

7. van der Velden, V. H. et al. Clinical significance of flowcytometric minimal residual disease detection in pediatric acute myeloid leukemia patients treated according to the DCOG ANLL97/MRC AML12 protocol. Leukemia 24, 1599–1606 (2010).

8. Tierens, A. et al. Residual disease detected by flow cytometry is an independent predictor of survival in childhood acute myeloid leukemia; results of the NOPHO-AML 2004 study. Br. J. Haematol. 174, 600–609 (2016).

9. Noren, D. P. et al. A Crowdsourcing Approach to Developing and Assessing Prediction Algorithms for AML Prognosis. PLoS Comput. Biol. 12, 1–16 (2016).

10. Dick, J. E. Stem cell concepts renew cancer research. Stem Cells 112, 4793–4807 (2008).

11. Corces-Zimmerman, M. R. & Majeti, R. Pre-leukemic evolution of hematopoietic stem cells: the importance of early mutations in leukemogenesis. Leukemia 28, 2276–2282 (2014).

12. Passegué, E., Jamieson, C. H. M., Ailles, L. E. & Weissman, I. L. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? Proc. Natl. Acad. Sci. 100, 11842–11849 (2003).

13. Guzman, M. L. & Allan, J. N. Leukemia Stem Cells in Personalized Medicine. Stem Cells 32, 844–851 (2014).

14. Terwijn, M. et al. Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. PLoS One 9, e107587 (2014).

15. Costello, R. T. et al. Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. Cancer Res. 60, 4403–4411 (2000).

16. Witte, K.-E. et al. High Proportion of Leukemic Stem Cells at Diagnosis Is Correlated with Unfavorable Prognosis in Childhood Acute Myeloid Leukemia. Pediatr. Hematol. 28, 91–99 (2011).

17. van Rhenen, a et al. Aberrant marker expression patterns on the CD34+CD38− stem cell compartment in acute myeloid leukemia allows to distinguish the malignant from the normal stem cell compartment both at diagnosis and in remission. Leukemia 21, 1700–1707 (2007).

18. Horton, S. J. & Huntly, B. J. P. Recent advances in acute myeloid leukemia stem cell biology. Haematologica 97, (2012).

19. Zeijlemaker, W. et al. A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. Leukemia 30, 439–46 (2016).

20. de Bont, E. et al. Excellent outcome in pediatric AML with response guided chemotherapy without allogeneic HSCT in first complete remission: Results from protocol DB-AML01. in ASH Annual Meeting, Orlando (2015).

21. Kalina, T. et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia 26, 1986–2010 (2012).

22. Zeijlemaker, W. et al. Absence of leukemic CD34+ cells in acute myeloid leukemia is of high prognostic value: A longstanding controversy deciphered. Br. J. Haematol. 171, 227–238 (2015).

23. Gray, R. J. Competing Risk. Ann. Stat. 16, 1141–1154 (1988).

24. Dang, H., Chen, Y., Kamel-Reid, S., Brandwein, J. & Chang, H. CD34 expression predicts an adverse outcome in patients with NPM1-positive acute myeloid leukemia. Hum. Pathol. 44, 2038–2046 (2013).

25. Kanda, Y. et al. The clinical significance of CD34 expression in response to therapy of patients with acute myeloid leukemia. Cancer 88, 2529–2533 (2000).

26. de Jonge, H. J. M. et al. Prognostic impact of white blood cell count in intermediate risk acute myeloid leukemia: Relevance of mutated NPM1 and FLT3-ITD. Haematologica 96, 1311–1317 (2011).

27. Ran, D. et al. Heterogeneity of leukemia stem cell candidates at diagnosis of acute myeloid leukemia and their clinical significance. Exp. Hematol. 40, 155-165.e1 (2012).

28. Chávez-González, A. et al. Expression of CD90, CD96, CD117, and CD123 on Different Hematopoietic Cell Populations from Pediatric Patients with Acute Myeloid Leukemia. Arch. Med. Res. 45, 343–350 (2014).

29. Ho, T. et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. Blood 128, 1671–1679 (2016).

30. Voskova, D., Schnittger, S., Schoch, C., Haferlach, T. & Kern, W. Use of five-color staining improves the sensitivity of multiparameter flow cytomeric assessment of minimal residual disease in patients with acute myeloid leukemia. Leuk. Lymphoma 48, 80–8 (2007).

31. Wouters, R., Cucchi, D., Kaspers, G. J., Schuurhuis, G. J. & Cloos, J. Relevance of leukemic stem cells in acute myeloid leukemia: heterogeneity and influence on disease monitoring, prognosis and treatment design. Expert Rev. Hematol. 7, 791–805 (2014).

SUPPLEMENTARY DATA

Patients

From the >100 paediatric AML patients diagnosed between 2010-2014 and treated according to the Dutch-Belgian paediatric AML-01 (DB AML-01) protocol (EudraCT 2009-014462-26), 86 patients were analysed with an additional stem cell identification tube at diagnosis. Next to stem cell identification, standard morphological, immunological, cytogenetic and molecular criteria according to the WHO 2008 classification was included in immunophenotypic characterization at diagnosis. Exclusion criteria were AML with PML-RARA, secondary AML, untreated refractory anaemia with excess blasts (RAEB) and RAEB in transformation. For this study there were no selection criteria other than the availability of stem cell flow cytometry data.

The cohort statistics of the study population versus the total DB-AML01 study population are CR rate: 93% versus 90%; 4-year survival probability: 84% versus 78%. As the additional flow cytometry measurement acquires extra cells, we could reason that only samples with enriched cell number are included. However, a high WBC count at diagnosis is considered an prognostic factor for poor outcome whereas our cohort had a better outcome.

Table S1.

		N	Median (Range)
Age (years)		86	7 (0-16)
White blood cell count (x 10^9/l)		81	24 (2-449)
Blast in bone marrow at diagnosis (percentage)		81	75 (0-98)
CD34 expressing cells (percentage)			
White blood cells		86	6 (0-93)
Blast cells		86	18 (0-99)
Time to			
First complete remission (days)		80	32 (12-122)
First relapse (months)		80	20 (1-50)
		N	%
Gender (N=86)	Male	43	50
	Female	43	50
Mutation (N=86)	FLT3	16	19
	NPM1	7	8
Karyotype (N=84)	Normal	19	22
	Balanced translocations	30	35
	Unbalanced abnormalities	15	17
	Complex	20	23
Cytogenetic risk group (N=56)	Favorable	16	19
	Intermediate	38	44
	Adverse	2	2
French American British-type	M0	3	4
(N=86)	M1	13	15
	M2	14	16
	M4	18	21
	M5	19	22
	M6	1	1
	M7	5	6
	unknown	13	12

FLT3; fms like tyrosine kinase (internal tandem duplication or tyrosine kinase domain mutations), NPM1; Nucleophosmin 1

Table S2.

PacB	PacO	FITC	PE	PerCPCy5.5	PECy7	APC	APCH7
HLA-DR	CD45	CD123	CD56/CD2	CD34	CD117	CD7	CD38

Marker	Fluorochrome	Clone	Source	Catalogue number
HLA-DR	PacB	L243	Biolegend	307624
CD45	PacO	HI30	Invitrogen	MHCD4530
CD123	FITC	AC145	MiltenyiBiotec	130-090-897
CD56	PE	C5.9	Cytognos	CYT-56PE
CD2	PE	39C1.5	Beckman Coulter	A07744
CD34	PerCP-Cy5.5	8G12	BD Bioscience	347222
CD117	PECy7	104D2D1	Beckman Coulter	IM3698
CD7	APC	124-1D1	eBioscience	17-0079-42
CD38	APCH7	HB7	BD Bioscience	646786

Table S3.

Variables		Hazard ratio	Standard error	95% CI	P value
LSC load		2.49	0.43	1.08 - 5.73	0.03
Time to CR		0.00	0.00	0.99 - 1.01	0.31
Risk group	(ref) intermediate	--	--	--	--
	favorable	0.24	0.62	0.07 - 0.82	0.02
	adverse	0.60	1.03	0.08 - 4.53	0.62
Karyotype	(ref) normal	--	--	--	--
	balanced transl.	0.29	0.48	0.11 - 0.74	0.01
	unbalanced abn.	0.95	0.46	0.39 - 2.32	0.90
	complex	0.55	0.48	0.22 - 1.41	0.21
FAB classification	(ref) M0	--	--	--	--
	M1	1.04	1.08	0.12 - 8.63	0.97
	M2	1.25	1.07	0.15 - 10.22	0.84
	M4	0.77	1.08	0.09 - 6.42	0.81
	M5	0.90	1.08	0.11 - 7.53	0.93
	M6	3.49	1.42	0.22 - 56.37	0.38
	M7	0.42	1.42	0.03 - 6.77	0.54
Gender	female	0.77	0.34	0.39 - 1.49	0.43
FLT3	ITD/|TKD	2.00	0.38	0.96 - 4.19	0.07
NPM1	mutated	0.86	0.61	0.26 - 2.82	0.80
Core binding factor	mutated	0.23	0.74	0.06 - 1.00	0.05
WBC count (x 10^9/l)	> 50	2.14	0.35	1.08 - 4.24	0.03
blast BM (%)		1.01	0.01	0.99 - 1.02	0.36
blasts PB (%)		1.01	0.01	1.00 - 1.02	0.06

SE: standard error, FAB: French-American-British, FLT3: FMS-like tyrosine kinase 3, ITD: internal tandem duplication, TKD: tyrosine kinase domain, NPM1: Nucleophosmin, WBC: white blood cell, BM: bone marrow, PB: peripheral blood

Table S4.

	Variables	Hazard ratio	Standard error	95% CI	P value
Univariate	LSC load	2.49	0.43	1.08 - 5.73	0.03
	White Blood Cell Count	1.08	0.35	1.01 - 4.24	0.03
	Core Binding Factor	0.23	0.74	0.06 - 1.00	0.05
Multivariate	LSC load	4.09	0.67	1.10 - 15.20	0.04
	White Blood Cell Count	2.52	0.42	1.10 - 5.77	0.03
	Core Binding Factor	0.27	0.78	0.04 - 0.82	0.03

CHAPTER 5

APPLICABILITY AND REPRODUCIBILITY OF ACUTE MYELOID LEUKEMIA STEM CELL ASSESSMENT IN A MULTI-CENTER SETTING

Diana Hanekamp, Alexander N. Snel, Angèle Kelder, Willemijn J. Scholten, Naeem Khan, Marlen Metzner, Maria Irno-Consalvo, Mayumi Sugita, Anja de Jong, Sjoerd Oude Alink, Harrie Eidhof, Miriam Wilhelm, Michaela Feuring-Buske, Jennichjen Slomp, Vincent H.J. van der Velden, Edwin Sonneveld, Monica Guzman, Gail J. Roboz, Francesco Buccisano, Paresh Vyas, Sylvie Freeman, Costa Bachas, Gert J. Ossenkoppele, Gerrit J. Schuurhuis, Jacqueline Cloos

Br J Haematol. 2020 Apr 2;190(6):891-900

ABSTRACT

Leukemic stem cells (LSC) have been experimentally defined as the leukemia propagating population and are thought to be the cellular reservoir of relapse in acute myeloid leukemia (AML). Therefore, LSC measurements are warranted to facilitate accurate risk stratification. Previously, we published the composition of a one-tube flow cytometric assay characterized by the presence of 13 important membrane markers for LSC detection. Here we present the validation experiments of the assay in several large AML research centers both in Europe and the United States. Variability within instruments and sample processing showed high correlations between different instruments ($R_{pearson}$>0.91, p<0.001). Multi-center testing introduced variation in reported LSC percentages but was found below clinical relevant threshold. Clear gating protocols resulted in all laboratories being able to perform LSC assessment of the validation set. Participating centers were nearly unanimously able to distinguish LSC^{high} (>0.03% LSC) from LSC^{low} (<0.03% LSC) despite inter-laboratory variation in reported LSC percentages. This study proves that the LSC assay is highly reproducible. These results together with the high prognostic impact of LSC load at diagnosis in AML patients render the one-tube LSC assessment a good marker for future risk classification.

INTRODUCTION

Acute myeloid leukemia (AML) is a heterogeneous group of diseases, with the shared feature of proliferation of immature myeloid blasts in the bone marrow and blood. The classification of AML has changed dramatically over the last decades, and is mainly based on chromosomal abnormalities and gene mutations underlying each individuals' disease (Döhner *et al*, 2017). Despite this advancement, risk classification remains suboptimal as a proportion of patients will relapse regardless of absence of poor-risk factors at diagnosis. Improved detection of measurable residual disease (MRD) during therapy by immunophenotypic and molecular methods has shown low levels of persisting disease in patients in morphologic remission (Jongen-Lavrencic *et al*, 2018), essential for further therapy choices. Presumably, this MRD compartment encompasses leukemic stem cells (LSCs). LSCs are pivotal for underlying leukemia propagation, therapy resistance, and as cellular reservoir of relapse (Gupta *et al*, 2009; Gerber *et al*, 2012; van Rhenen *et al*, 2005). Recent studies have correlated high LSC frequencies at the time of diagnosis with presence of MRD and subsequent poor prognosis (Terwijn *et al*, 2014; Hanekamp *et al*, 2017). The implementation of LSC measurements in the clinic is therefore instrumental for risk stratification and facilitating the selection of appropriate treatment protocols (Hourigan *et al*, 2019).

Several studies identified LSCs by (cyto)genetic and functional characteristics (Ng *et al*, 2016; Gerber *et al*, 2012; Won *et al*, 2015). Apart from these assays, LSCs can be immunophenotypically identified based on the principle that LSC can aberrantly express antigens. These flow cytometric assays can easily be implemented in most AML diagnostic workups. Although different cellular compartments are shown to possibly contain leukemia initiating cells (Sarry *et al*, 2011; Ng *et al*, 2016), the CD34+CD38- compartment is the most established (Terwijn *et al*, 2014; van Rhenen *et al*, 2007; Costello *et al*, 2000). The use of an antibody panel with CD34 and CD38 has therefore been the basis of many studies, discriminating the CD34+CD38- cell fraction, which contains both LSCs and normal hematopoietic stem cells (HSCs), from other cells.

For optimal discrimination between LSCs and HSCs, multiple markers were identified, highlighting the heterogeneity of AML LSC (Hanekamp *et al*, 2017; Terwijn *et al*, 2014). We previously tested many of the proposed LSC markers in a large cohort of AML patients and selected those which showed the best distinction between HSC and LSC and, moreover, identified the highest LSC burden (Zeijlemaker *et al*, 2016). After omitting redundant markers, 13 markers (i.e. CLEC12A, TIM-3, CD7, CD11b, CD22, CD56, CD33, CD45RA, CD123, CD44 and backbone markers CD34, CD38, CD45) remained necessary for correct identification. This panel of markers was arranged in a single 8-color flow cytometry antibody panel, combining the first six markers together in one fluorescence channel, hereinafter referred to as the "Combi" channel, with the potential to be easily implemented in other laboratories (Zeijlemaker *et al*, 2016).

Medical laboratory assays are essential to support clinicians to provide optimal treatment choices for patients; hence their results and reports should be of the best achievable quality. Research in improving these assays is ongoing, but is also directed towards standardization (Cloos *et al*, 2018; Zeijlemaker *et al*, 2018) to facilitate multi-institutional collaborations.

Here we evaluated the technical and analytical feasibility of the 8-color LSC single tube assay, as well as standardization of the process. The study is conducted in several large research centers both in Europe and the United States, with extensive flow cytometry experience, but not with the assessed LSC-assay. We show that limited training result in highly concordant results, allowing other centers to independently validate the clinical utility of LSC testing in AML. These results, together with the high prognostic impact of LSC load at AML diagnosis, render the one-tube LSC assessment a good marker for future risk classification.

MATERIALS AND METHODS

Instruments, setups and samples

Used instruments are listed in Table S1. We previously described the setup of flow cytometers, as based on EuroFlow instructions (Cloos *et al*, 2018; Kalina *et al*, 2012). Sample information and details regarding harmonization of all machines are described in supplemental materials and methods.

Study setup

A schematic overview of the study setup is shown in Figure 1 and in detail described in supplementary text. In short, three cryopreserved diagnosis samples were used to evaluate inter-instrument variance and four cryopreserved diagnosis samples were used for inter-laboratory processing. Gating was trained on these latter four samples, and validated in 10 FCS files of representative diagnosis AML samples.

Statistics

All results showed complete gating strategy listing the number of events for all relevant populations (lymphocytes, (CD34+) blasts, CD34+CD38-, LSC and HSC). LSC percentages ≥0.03% was classified as LSC^{high} or LSC^{low} <0.03% (Terwijn *et al*, 2014; Zeijlemaker *et al*, 2018). Since percentages found in LSC^{low} patients are low, variances calculated as coefficient of variation are high. ±0.5log was considered as acceptable error with limited effect on prognostic value (Figure S1). Variation was calculated using the Excel function VAR.P. Pearson correlation coefficients were calculated when reported percentages were compared between laboratories or machines.

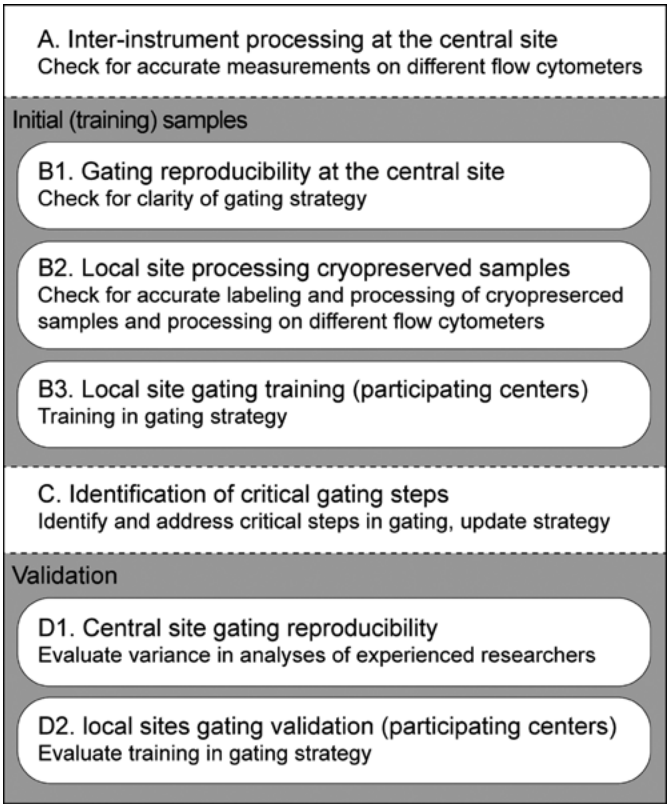


Figure 1. Schematic overview of study setup. The study can be divided in three parts: a) pre-analysis by the central site, b) analysis of four initial samples by six researchers of the central site and participating centers and c) the validation of both the coordinating center and the participating centers.

RESULTS

A. Central site inter-instrument processing

To evaluate the influence of different flow cytometers, the assay was first assessed on different platforms at the central site. Expression plots of the total blast population and CD34+ blast subpopulation of two representative samples measured on BD LSRFortessa and BC Gallios EX were compared to BD FACSCanto II (Figure 2). Lymphocyte and LSC percentages were analyzed in all samples and showed high correlation with results on BD FACSCanto II (Figure 2C).

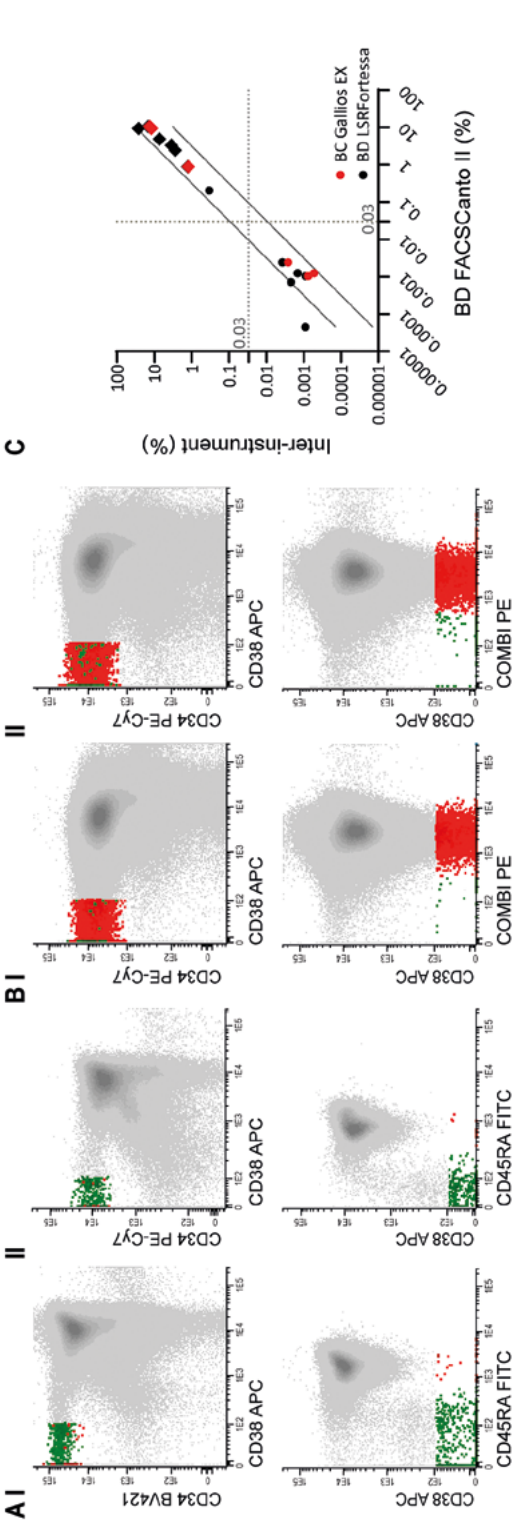


Figure 2. Central site inter-instrument processing. Multiple samples with similar processing were measured across multiple flow cytometric instruments (A) or platforms (B). A) Training sample 1 was measured on BD FACSCanto II (I) and BD LSRFortessa (II). FACS plots of complete blasts (top row) and CD34+ blasts (bottom row). CD33 and CD34 were exchanged in channel (see table S3) for standardization within the coordinating institute. B) A diagnostic AML sample was measured on BD FACSCanto II (I) and BC Gallios EX (II). FACS plots for complete blast population (top row) and CD34+ blast population (bottom row). C) LSC percentages (dots) and lymphocyte percentages (diamonds) analyzed in samples measured on BD LSRFortessa (black) and BC Gallios EX (red) compared to BD FACSCanto II. 0.5log error is depicted as diagonal lines. 0.03% clinical stem cell cutoff depicted as dotted lines. Pearson correlation coefficients between BC Gallios EX and BD FACSCanto II and BD LSRFortessa and BD FACSCanto II were r=1.00, p=<0.001 and r=0.91, p=<0.0001 respectively. Grey: (CD34+) population; Red: LSC; Green: HSC.

B. Training

B1. FCS files generated at central site with analysis by six researchers from central site

Gating reproducibility within the trained team at the central site was evaluated by having six researchers perform LSC assessments on FCS files generated at central site (n=4) (Figure 3 and Figure S2). Since LSCs are heterogeneous in marker expression, the gating of all individual markers was analyzed (Table S4). The largest variance was found for LSCs gated positive for Combi (mean $2.7*10^{-6}$, range $0.0\text{-}9.7*10^{-6}$). Without prior knowledge of inclusion of a CD34-negative patient (Zeijlemaker *et al*, 2015), all researchers from the central site were able to identify this CD34-negative sample (T3).

B2. FCS files generated at local sites with analysis by one researcher from central site

The same set of cryopreserved samples were sent to all participating laboratories accompanied by a protocol describing flow cytometer setup (supplemental text). Generated FCS files were uploaded by the local sites to a designated repository and checked by one researcher from the central site. Percentages of LSC were analyzed in all samples (Figure 3, Figure S3 and Table S5). While there was variation in the number of WBCs measured (T1 mean 1551352 events, range 460484-4538067; T2 mean 2137499 events, range 475915-3153669; T3 mean 1807224 events, range 548304-4218299; T4 mean 1318815 events, range 478873-3808045) measurements were overall highly comparable. Representative expression plots of the total blast population and expression plots of the CD34+ blast subpopulation of sample T2 show high resemblance between different local sites (Figure S2). Analysis performed by one researcher from the central site resulted in LSC percentages resembling LSC percentages found in analysis of files generated within the central site (Figure 3). Sample T1 from laboratory 4 contained 295370 WBC and could therefore not be analyzed.

B3. FCS files generated at local sites with analysis from local site

The participating centers were asked for analysis of the files using advised gating strategy (supplemental data 1). Analyses were uploaded to the repository and reviewed by the central site. When gating could evidently be optimized (i.e. WBC gate included debris) feedback was sent to the local site and analysis could be revised. An average of 1.9 (range 1-3) analysis rounds were needed to come to final gating results (two laboratories did not need feedback, four laboratories needed feedback once, while one laboratory needed feedback twice). Results reported by the local sites showed more variability (especially in sample T3 and T4) compared to analyses within the central site (Figure 3 and Figure S3 and corresponding Table S6). In all samples the largest variation was found in percentages $CD44^{positive}$ LSC (mean 0.0003621%, range 0.0000016%-0.0008108%). In conclusion, variance mostly results from data analysis not data collection.

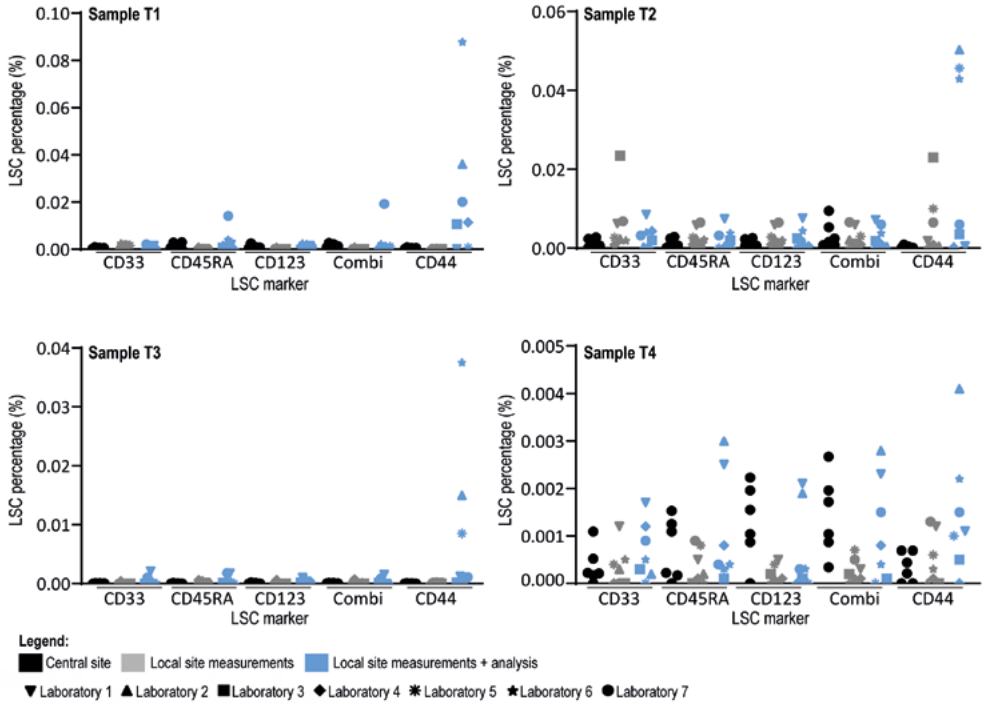


Figure 3. Results of central site and local sites on initial samples. Results of analysis of FCS files generated at central site by six researchers from central site (black), results of FCS files generated at local sites with analysis by one researcher from central site (grey) and results of FCS files generated at local sites with analysis from local site (blue) on samples T1-T4. All individual results are shown as percentage leukemic stem cells of the complete white blood cell compartment. Laboratories are specified using different symbols, showing that differences in LSC percentages is not consistently explained by one laboratory. Axes run from 0.000-0.030% (clinical relevant cut-off) for sample T1 and T2, and 0.000-0.003% for samples T3 and T4. CD44 is not depicted due to high variance (shown in Figure S3).

C. Critical steps in gating strategy

As a consequence of training the participating centers, a number of critical steps in the gating strategy were elucidated.

Exact gating of WBC compartment. Since the burden of LSC is presented as percentage of complete WBC load, correct gating of this population is critical. However, gating for WBCs was not described. Furthermore, as no live-dead marker is included in the stem cell tube, gating for viable WBC versus debris is solely based on scatter properties. Therefore, it is recommended to start with gating the lymphocytes in a CD45/SSC plot. Since lymphocytes are low in side and forward scatter properties, they can act as starting point for the WBC gate (Figure S4A).

Discriminating CD34+ blasts from CD34- blasts. Leukemic blasts can differ in CD34 expression, and gating of CD34positive blasts can therefore be challenged by CD34dim

blasts, the CD34/CD38 plot can be illustrative (Figure S4B). After gating LSC and HSC, the gating strategy supports so called back-gating. Both LSC and HSC often present as cluster in SSC/FSC, CD45/SSC and CD34/SSC plots. Scattered LSC (or HSC) events low in the CD34+ gate therefore direct to narrowing of the CD34gate (Figure S4C).

All tips for optimal gating were reported to all laboratories before the validation step was initiated.

D. Validation

D1. Additional FCS files generated at central site with analysis by researchers from central site

To evaluate the reproducibility of LSC analysis in clinical practice, ten additional representative diagnostic FSC files were selected for analysis. Of these samples, four were LSC^{high} (V5, V7, V8 and V10), five were LSC^{low} (V1, V2, V3, V4, V6, V9), of which one was CD34-negative (V1). Two samples were around the cutoff (V5 LSC^{high}, V6 LSC^{low}). Results were reported similar to routine practice: which includes gating of all LSC markers (i.e. CD33, CD44, CD123, CD45RA and Combi) separately and selecting the best marker to ultimately report as LSC load (Figure S5A, detailed in Table 1). As LSC are frequently covered by more than one LSC marker, the selection of a different marker did not always result in identification of a distinct different population (see supplemental gating strategy for examples). Discrimination between high (≥0.03%) and low (<0.03%) LSC load was unanimously among all researchers in 13/14 samples (93%). Sample V3 showed discordance between the researchers, as one researcher included $CD45^{high}$ cells as leukemic blasts and CD34+CD38- cells with higher scatter properties as LSC, in absence of LSC markers. The selected 'best' marker (Table S7) showed high resemblance (i.e. 5/6 or 6/6 researchers chose the same marker) in part of the samples (6/14; 43%), and lower resemblance (i.e. 3-4 of the 6 researchers chose the same marker) in the other part of the samples (8/14; 57%). CD45RA was selected as best LSC marker in 63%, followed by Combi (18.3%), CD123 (13.3%) and CD33 (5.0%). Marker CD44 was never selected.

Repeat analyses by individual operators is evaluated in a select set of samples and is shown in Figure S6 and corresponding table S8. The variance introduced by repeated analyses is minor, and had no effect on outcome (i.e. LSC^{low} remained LSC^{low}, LSC^{high} remained LSC^{high}).

D2. Additional FCS files generated at central site with analysis by local sites

LSC analyses of the local sites were evaluated in the same set of ten FSC files (Figure S5B). Participating centers were unanimously able to distinguish LSC^{high} samples V7, V8 and V10, but LSC percentages (Table 1) and selected markers differed (Table S7). Sample V5 was identified in 6/7 laboratories as LSC^{high}, but identified as LSC^{low} in one participating laboratory with 0.029% (nevertheless very close to the cut off of 0.03%). All results in LSC^{high} patients were within ±0.5log error (Figure 4), which was identified as acceptable

Table 1. Percentages and events validation set

	Sample	T1	T2	T3	T4	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Intra-laboratory															
Researcher 1	%	0.001	0.002	0.000	0.001	0.000	0.000	0.002	0.001	0.051	0.018	0.121	2.359	0.001	0.222
	Events	24	91	0	42	0	11	19	15	612	364	2509	35834	8	3669
Researcher 2	%	0.001	0.001	0.000	0.003	0.000	0.000	0.047	0.000	0.051	0.018	0.125	2.369	0.001	0.205
	Events	32	32	0	92	0	7	457	7	627	352	2528	36939	9	3453
Researcher 3	%	0.002	0.001	0.000	0.001	0.000	0.000	0.002	0.000	0.052	0.017	0.125	2.388	0.000	0.217
	Events	48	28	1	42	0	8	18	3	634	334	2513	36086	1	3492
Researcher 4	%	0.000	0.001	0.000	0.001	0.000	0.000	0.003	0.000	0.037	0.014	0.112	2.051	0.000	0.177
	Events	16	40	0	42	3	7	31	6	460	280	2275	31804	105	6339
Researcher 5	%	0.003	0.003	0.000	0.002	0.000	0.002	0.009	0.000	0.051	0.019	0.101	2.438	0.001	0.230
	Events	113	181	1	75	1	82	85	7	620	397	2127	37642	6	3938
Researcher 6	%	0.001	0.002	0.000	0.002	0.000	0.000	0.001	0.000	0.048	0.020	0.117	2.284	0.000	0.204
	Events	48	86	0	92	0	10	12	3	580	397	2451	34002	4	3451
Inter-laboratory															
Laboratory 1	%					0.000	0.001	0.007	0.000	0.051	0.018	0.141	2.248	0.001	0.266
	Events					3	17	64	5	627	351	2829	33456	14	4388
Laboratory 2	%					0.000	0.000	0.001	0	0.047	0.011	0.086	1.933	0.000	0.098
	Events					2	5	11	1	573	223	1803	30155	2	1668
Laboratory 3	%					0.000	0.000	0.001	0.000	0.029	0.011	0.070	1.580	0.001	0.110
	Events					2	4	7	3	446	223	1534	28601	6	2422
Laboratory 4	%					0.000	0.001	0.028	0.000	0.036	0.014	0.085	2.017	0	0.185
	Events					8	16	285	4	435	277	1859	30494	0	2995
Laboratory 5	%					0	0.000	0.002	0	0.036	0.014	0.106	2.329	0.000	0.244
	Events					0	6	21	0	444	250	2302	32983	4	3884
Laboratory 6	%					0	0.000	0.003	0.000	0.049	0.013	0.088	3.016	0.001	0.070
	Events					0	9	30	5	519	241	1698	32568	8	1082

Table 1. (continued)

	Sample	T1	T2	T3	T4	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Laboratory 7	%					0.000	0.001	0.008	0.000	0.044	0.014	0.077	2.364	0.002	0.299
	Events					2	22	83	8	574	268	610	37017	19	5180
Intra-laboratory	Mean	0.001	0.002	0.000	0.002	0.000	0.001	0.011	0.000	0.048	0.018	0.117	2.315	0.000	0.209
	Minimum	0.000	0.001	0.000	0.001	0.000	0.000	0.001	0.000	0.037	0.014	0.101	2.051	0.000	0.177
	Maximum	0.003	0.003	0.000	0.003	0.000	0.002	0.047	0.001	0.052	0.02	0.125	2.438	0.001	0.230
	Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000
Inter-laboratory	Mean					0.000	0.000	0.007	0.000	0.042	0.013	0.093	2.212	0.001	0.182
	Minimum					0.000	0.000	0.001	0.000	0.029	0.011	0.070	1.580	0.000	0.070
	Maximum					0.000	0.001	0.028	0.000	0.051	0.018	0.141	3.016	0.002	0.299
	Variance					0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.171	0.000	0.007

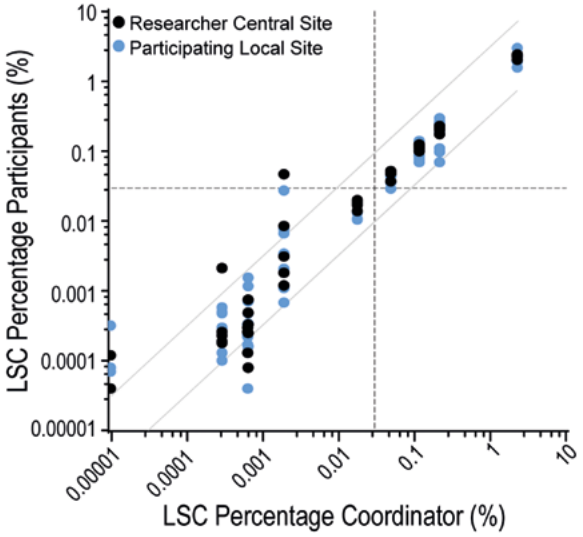


Figure 4. Results of central site and local sites on validation FCS files. Ten representative diagnostic AML samples were selected and corresponding flow cytometry files were sent to six researchers from the central site and seven participating centers for analysis. Results were reported back as leukemic stem cell percentage analyzed by the most reliable stem cell marker (or markers). Results reported by one researcher from the central site compared to LSC percentages reported by all other participants (researchers from the central site in black, participating laboratories in blue). Previously determined (and validated) cut-off of 0.03% is shown as dotted line. Result above clinical validated cutoff fall within ±0.5 log error, shown as grey diagonal lines. Lower percentages fall outside ±0.5 log error but are clinically irrelevant.

error (see Materials and methods). There was a high correlation in detected LSC burden (mean r=0.999, range 0.998-1.000, $p<0.001$) between participating centers and the central site. In analyses of the local sites, CD45RA was selected as best marker in most samples (38.6%), followed by Combi (27.1%). CD44 was selected as best marker in 17.1%, while never selected by researchers from the central site. While none of the researchers from the central site selected the same marker for all samples, three participating institutes did.

E. Clinical implications

Implementation of the assay has important clinical implications as it allows the identification of patients that have distinct different prognosis, and could therefore be added to future risk classification. Correct characterization of patients with a significant poor outcome (LSC^{high}) and patients with a significant better outcome (i.e. CD34negative (Zeijlemaker *et al*, 2015)) was evaluated. The 14 samples analyzed within this study included four patients with high LSC load, and two CD34negative patients. Three patients were correctly identified as LSC^{high} in all analyses. The remaining sample was correctly categorized in 12/13 analyses. The two CD34negative patients were recognized as LSC^{low} in all analyses, but only researchers from the coordinating institute noted that these patients were CD34negative.

It is important to note that the participating centers were not specifically asked for this conclusion. Figure 4C shows higher variation in samples below the clinical 0.03% cut-off, as is to be expected. However, the clinical value of the exact frequency of LSC at time of diagnosis is not thoroughly established and possibly obscured due to higher intra- and inter-laboratory variability. In summary, the data reveal that the current dichotomization between LSC^{high} and LSC^{low} is more robust, and should therefore be adopted.

DISCUSSION

Identification of patients with high LSC load at time of diagnosis, allows identification of patients with poor disease outcome very early in the disease course (Zeijlemaker *et al*, 2018; Terwijn *et al*, 2014). Similarly, identification of patients that lack aberrant leukemic CD34+ (stem) cells allows identification of patients with a distinct better prognosis before the response of therapy can be perceived (Zeijlemaker *et al*, 2015). While the contribution of CD34+CD38- LSC to poor disease outcome is demonstrated in several studies (Plesa *et al*, 2017; Jentzsch *et al*, 2017; Hwang *et al*, 2012), LSC measurements are not clinically implemented because of the seemingly complex process that requires specific experience and standardization in laboratories involved. This study among 8 institutes shows that the one-tube LSC assay is highly reproducible between several large flow cytometry AML centers in both Europe and the United States after a relatively simple training.

Sophisticated 8-10 color flow cytometry is at the basis of the diagnosis, characterization and monitoring of hematological malignancies. Correct implementation of the technique and standardization in its applications is of high importance and several guidelines to achieve this have recently been published (Lacombe *et al*, 2016; Solly *et al*, 2019). In this study we demonstrate that harmonization between flow cytometers is required for comparable results. Here, the use of BD's FC beads or setup according to Euroflow protocol was adequate to result in comparable measurements as percentages of lymphocytes, blasts, CD34+ blasts, CD34+CD38dim and CD34+CD38- fractions were decidedly comparable among all institutes (data not shown).

A defined gating strategy is essential for laboratories aspiring to incorporate any flow cytometric assessment. To highlight the effectiveness of our gating strategy our training emphasized on samples low in LSC frequency (three LSC^{low}, one CD34negative), since analyses of low-frequent cell populations is sensitive to errors. Nonetheless, limited feedback from experienced researchers was sufficient to train new researchers to gate according to protocol and achieve a high degree of comparison with results found by experienced researchers. During the training phase, a number of critical steps in the gating strategy were elucidated which are accentuated in this article.

The development of a standardized antibody panel for LSC detection (Zeijlemaker *et al*, 2016) helps to strive towards standardization. This tube simplifies LSC assessment in routine AML flow cytometry work-up, as well as limiting costs and number of cells needed. While the LSC tube consists of the best (most discriminative, high negative predictive

value and most sensitive) markers for identification of LSC, some were preferred over others. Markers CD45RA and Combi often show distinct separation between HSC and LSC as two separate 'tails' within the CD34+CD38- fraction. CD44 is used as marker for LSC (Zeijlemaker *et al*, 2016) based on overexpression of CD44 compared to expression on normal hematopoietic (stem) cells (Cao *et al*, 2016), but correct identification of normal-high expression versus overexpression is difficult when only one 'tail' is present. For specific purposes, exclusion of CD44 could be suggested to allow incorporation of additional antibodies.

The training in analysis was confirmed by ten representative diagnostic AML samples, ranging from high levels of LSC to absence of LSC, mimicking the clinical setting. As expected, a high degree of correlation is found in samples with higher frequencies of LSCs in contrast to samples with low frequencies. Previously, a cut-off of 0.03% was identified as clinically prognostic relevant (Terwijn *et al*, 2014; Zeijlemaker *et al*, 2018). In our validation, five samples with percentages proximal to this cut-off were correctly classified in 4/5 cases, with the remaining sample being misclassified as LSC^{low} by one institute (with 0.029% just below cut-off). Analysis of the exact percentage of LSC below the 0.03% cut-off is prone to higher variation, but critical analysis of LSC^{low} patients is crucial for the identification of CD34negative patients, associated with an overall good prognosis (Zeijlemaker *et al*, 2015, 2018). Further research should be undertaken to evaluate whether CD34negative patients are correctly identified and discriminated from LSC^{low} patients.

Whereas addition of LSC measurements at diagnosis might lead to further improvement of risk group stratification, post-induction MRD measurements are valuable for guiding post-remission strategies (Buccisano *et al*, 2019). MRD in AML is a rapidly evolving area with fast developments in designs and approaches. While the introduction of next-generation sequencing MRD detection certainly holds promise for the future, the combination with flow cytometry showed that both techniques contributed independently to the prognostic value of the patient cohort (Jongen-Lavrencic *et al*, 2018). Combining flow cytometry MRD measurements with post-induction LSC measurements, even improved the prognostic classification further (Zeijlemaker *et al*, 2018). Since LSC in MRD situation are rare events, correct gating of LSCs, but also measurement of sufficient WBCs is critical. The possibility and practicability of LSC MRD measurements are not yet described in this manuscript and should be explored as proceeding of this multicenter international group.

Current limitations of this study include the selection of a restricted set of patient samples which conceivably does not cover the complete cellular heterogeneity seen within the AML population. It could therefore be argued that the implementations of the assay in centers need to be evaluated in prospective multicenter studies. Furthermore, all samples measured were cryopreserved mononuclear cells. Ideally, fresh samples would be first measured at the coordinating institute, deemed suited for the training and then immediately sent to and measured at the participating centers. This was considered impractical due to introducing more variability from poorer viability. As the effect of

different sample processing could therefore not extensively be analyzed, the use of a standardized protocol is therefore warranted (Cloos *et al*, 2018; Schuurhuis *et al*, 2018).

In summary, we show that the one-tube LSC assay is highly reproducible for many different FC experienced laboratories after a relatively simple training. Since the tube is useful for finding almost all CD34+CD38- stem cells and requires limited sample, it can be implemented in clinical studies. The high concordance between different laboratories is particularly valuable for use in multicenter studies. These results together with the high prognostic impact of LSC load at diagnosis in AML patients render the one-tube LSC assessment a good marker for future risk classification.

ACKNOWLEDGEMENTS

The authors thank all MRD team members of the Amsterdam UMC for participation in the intra-laboratory analyses. We thank Erik Huys and Frank Preijers from the department of Laboratory Medicine-Laboratory for Hematology, Radboud UMC for their measurements on BC Navios and discussions on harmonization of flow cytometry platforms. We thank Hetty J. Bontkes, Martine Reijm and Jolien C. Hollander from the department of Clinical Chemistry, Medical Immunology (Amsterdam UMC, location VUmc) for help and access to BC Gallios. Antibodies used for these validation studies were provided by BD.

AUTHORSHIP CONTRIBUTIONS

DH, GJS and JC initiated and designed the study. DH, AK, WJS, NK, MM, MIC, MS, AdJ, SOA, and MW analyzed the flow cytometry data. DH coordinated the study, collected all data, conducted and analyzed all integrated analyses. DH performed the statistical analysis. DH wrote the manuscript which was further revised by CB and JC and reviewed by all coauthors.

CONFLICT OF INTEREST DISCLOSURE

Financial support for part of this research has been received by BD biosciences. A lyophilized version of the LSC tube is currently in production and will become available, with royalty payments for intellectual property rights to the VUmc.

SUPPLEMENTARY TEXT/TABLES/FIGURES

Supplemental data can be found online at https://onlinelibrary.wiley.com/doi/10.1111/bjh.16594.

REFERENCES

Buccisano, F., Maurillo, L., Schuurhuis, G.J., Del Principe, M.I., Di Veroli, A., Gurnari, C. & Venditti, A. (2019) The emerging role of measurable residual disease detection in AML in morphologic remission. *Seminars in Hematology*, **56,** 125–130 Available at: https://www.sciencedirect.com/science/article/abs/pii/S003719631830101X?via%3Dihub [Accessed August 23, 2019].

Cao, H., Heazlewood, S.Y., Williams, B., Cardozo, D., Nigro, J., Oteiza, A. & Nilsson, S.K. (2016) The role of CD44 in fetal and adult hematopoietic stem cell regulation. *Haematologica*, **101,** 26–37 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26546504 [Accessed April 10, 2019].

Cloos, J., Harris, J.R., Janssen, J.J.W.M., Kelder, A., Huang, F., Sijm, G., Vonk, M., Snel, A.N., Scheick, J.R., Scholten, W.J., Carbaat-Ham, J., Veldhuizen, D., Hanekamp, D., Oussoren-Brockhoff, Y.J.M., Kaspers, G.J.L., Schuurhuis, G.J., Sasser, A.K. & Ossenkoppele, G. (2018) Comprehensive Protocol to Sample and Process Bone Marrow for Measuring Measurable Residual Disease and Leukemic Stem Cells in Acute Myeloid Leukemia. *Journal of visualized experiments : JoVE*, **March 5,** Available at: http://www.ncbi.nlm.nih.gov/pubmed/29553571 [Accessed June 14, 2018].

Costello, R.T., Mallet, F., Gaugler, B., Sainty, D., Arnoulet, C., Gastaut, J. a. & Olive, D. (2000) Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. *Cancer Research*, **60,** 4403–4411.

Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F.R., Büchner, T., Dombret, H., Ebert, B.L., Fenaux, P., Larson, R.A., Levine, R.L., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G.J., Sanz, M., Sierra, J., Tallman, M.S., Tien, H.-F., Wei, A.H., et al (2017) Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*, **129,** 424–447 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27895058 [Accessed January 31, 2018].

Gerber, J.M., Smith, B.D., Ngwang, B., Zhang, H., Vala, M.S., Morsberger, L., Galkin, S., Collector, M.I., Perkins, B., Levis, M.J., Griffin, C. a, Sharkis, S.J., Borowitz, M.J., Karp, J.E. & Jones, R.J. (2012) A clinically relevant population of leukemic CD34(+)CD38(-) cells in acute myeloid leukemia. *Blood*, **119,** 3571–7 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3325044&tool=pmcentrez&rendertype=abstract.

Gupta, P.B., Chaffer, C.L. & Weinberg, R.A. (2009) Cancer stem cells: mirage or reality? *Nature Medicine*, **15,** 1010–1012 Available at: papers3://publication/doi/10.1038/nm0909-1010.

Hanekamp, D., Cloos, J. & Schuurhuis, G.J. (2017) Leukemic stem cells: identification and clinical application. *International Journal of Hematology*, **105,** 549–557 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28357569 [Accessed August 4, 2017].

Hourigan, C.S., Gale, R.P. & Walter, R.B. (2019) Refining AML outcome prediction. *Leukemia*, **33,** 283–284.

Hwang, K., Park, C.-J., Jang, S., Chi, H.-S., Kim, D.-Y., Lee, J.-H., Lee, J.H., Lee, K.H., Im, H.-J. & Seo, J.-J. (2012) Flow cytometric quantification and immunophenotyping of leukemic stem cells in acute myeloid leukemia. *Annals of Hematology*, **91,** 1541–1546 Available at: http://link.springer.com/10.1007/s00277-012-1501-7 [Accessed July 4, 2019].

Jentzsch, M., Bill, M., Nicolet, D., Leiblein, S., Schubert, K., Pless, M., Bergmann, U., Wildenberger, K., Schuhmann, L., Cross, M., Pönisch, W., Franke, G.-N., Vucinic, V., Lange, T., Behre, G., Mrózek, K., Bloomfield, C.D., Niederwieser, D. & Schwind, S. (2017) Prognostic impact of the CD34+/CD38− cell burden in patients with acute myeloid leukemia receiving allogeneic stem cell transplantation. *American Journal of Hematology*, **92,** 388–396 Available at: http://doi.wiley.com/10.1002/ajh.24663 [Accessed July 4, 2019].

Jongen-Lavrencic, M., Grob, T., Hanekamp, D., Kavelaars, F.G., al Hinai, A., Zeilemaker, A., Erpelinck-Verschueren, C.A.J., Gradowska, P.L., Meijer, R., Cloos, J., Biemond, B.J., Graux, C., van Marwijk Kooy, M., Manz, M.G., Pabst, T., Passweg, J.R., Havelange, V., Ossenkoppele, G.J., Sanders, M.A., Schuurhuis, G.J., et al (2018) Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *New England Journal of Medicine*, **378,** 1189–1199 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29601269 [Accessed April 18, 2018].

Kalina, T., Flores-Montero, J., van der Velden, V.H.J., Martin-Ayuso, M., Böttcher, S., Ritgen, M., Almeida, J., Lhermitte, L., Asnafi, V., Mendonça, A., de Tute, R., Cullen, M., Sedek, L., Vidriales, M.B., Pérez, J.J., te Marvelde, J.G., Mejstrikova, E., Hrusak, O., Szczepański, T., van Dongen, J.J.M., et al (2012) EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia*, **26,** 1986–2010 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3437409&tool=pmcentrez&rendertype=abstract.

Lacombe, F., Bernal, E., Bloxham, D., Couzens, S., Porta, M.G.D., Johansson, U., Kern, W., Macey, M., Matthes, T., Morilla, R., Paiva, A., Palacio, C., Preijers, F., Ratei, R., Siitonen, S., Allou, K., Porwit, A. & Béné, M.C. (2016) Harmonemia: a universal strategy for flow cytometry immunophenotyping—A European LeukemiaNet WP10 study. *Leukemia*, **30,** 1769–1772 Available at: http://www.nature.com/articles/leu201644 [Accessed November 29, 2018].

Ng, S.W., Mitchell, A., Kennedy, J.A., Chen, W.C., McLeod, J., Ibrahimova, N., Arruda, A., Popescu, A., Gupta, V., Schimmer, A.D., Schuh, A.C., Yee, K.W., Bullinger, L., Herold, T., Görlich, D., Büchner, T., Hiddemann, W., Berdel, W.E., Wörmann, B., Cheok, M., et al (2016) A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*, **540,** 433–437 Available at: http://www.nature.com/doifinder/10.1038/nature20598.

Plesa, A., Dumontet, C., Mattei, E., Tagoug, I., Hayette, S., Sujobert, P., Tigaud, I., Pages, M.P., Chelghoum, Y., Baracco, F., Labussierre, H., Ducastelle, S., Paubelle, E., Nicolini, F.E., Elhamri, M., Campos, L., Plesa, C., Morisset, S., Salles, G., Bertrand, Y., et al (2017) High frequency of CD34+CD38-/low immature leukemia cells is correlated with unfavorable prognosis in acute myeloid leukemia. *World journal of stem cells*, **9,** 227–234 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29321824 [Accessed July 4, 2019].

van Rhenen, A., van Dongen, G., Kelder, A., Rombouts, E.J., Feller, N., Moshaver, B., Walsum, M.S. -v., Zweegman, S., Ossenkoppele, G.J. & Jan Schuurhuis, G. (2007) The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. *Blood*, **110,** 2659–2666 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2007-03-083048.

van Rhenen, A., Feller, N., Kelder, A., Westra, A., Rombouts, E., Zweegman, S., van der Pol, M., Waisfisz, Q., Ossenkoppele, G.J. & Schuurhuis, G.J. (2005) High Stem Cell Frequency in Acute Myeloid Leukemia at Diagnosis Predicts High Minimal Residual Disease and Poor Survival. *Clinical Cancer Research*, **11,** 6520–6527 Available at: http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-05-0468.

Sarry, J., Murphy, K., Perry, R., Sanchez, P. V, Secreto, A., Keefer, C., Swider, C.R., Strzelecki, A., Cavelier, C., Récher, C., Mas, V.M., Delabesse, E. & Carroll, M. (2011) Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J. Clin. Invest.*, **121,** 384–395.

Schuurhuis, G.J., Heuser, M., Freeman, S., Béné, M.-C., Buccisano, F., Cloos, J., Grimwade, D., Haferlach, T., Hills, R.K., Hourigan, C.S., Jorgensen, J.L., Kern, W., Lacombe, F., Maurillo, L., Preudhomme, C., van der Reijden, B.A., Thiede, C., Venditti, A., Vyas, P., Wood, B.L., et al (2018) Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood*, **131,** 1275–1291 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29330221 [Accessed January 31, 2018].

Solly, F., Angelot-Delettre, F., Ticchioni, M., Geneviève, F., Rambaud, H., Baseggio, L., Plesa, A., Debliquis, A., Garnache-Ottou, F., Roggy, A., Campos, L., Aanei, C., Rosenthal-Allieri, A., Georget, M., Lachot, S., Jacob, M., Robillard, N., Wuilleme, S., Andre-Kerneis, E., Cornet, E., et al (2019) Standardization of Flow Cytometric Immunophenotyping for Hematological Malignancies: The FranceFlow Group Experience. *Cytometry Part A*, cyto.a.23844 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/cyto.a.23844 [Accessed August 23, 2019].

Terwijn, M., Zeijlemaker, W., Kelder, A., Rutten, A.P., Snel, A.N., Scholten, W.J., Pabst, T., Verhoef, G., Löwenberg, B., Zweegman, S., Ossenkoppele, G.J. & Schuurhuis, G.J. (2014) Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. *PLoS ONE*, **9,** e107587 Available at: http://dx.plos.org/10.1371/journal.pone.0107587.

Won, E., Kim, H., Park, R., Choi, S.-Y., Shin, J.H., Suh, S.-P., Ryang, D.-W., Szardenings, M. & Shin, M.-G. (2015) Direct confirmation of quiescence of CD34+CD38- leukemia stem cell populations using single cell culture, their molecular signature and clinicopathological implications. *BMC Cancer*, **15,** 217 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4391681&tool=pmcentrez&rendertype=abstract.

Zeijlemaker, W., Grob, T., Meijer, R., Hanekamp, D., Kelder, A., Ham, J., Oussoren-Brockhoff, Y.J.M., Snel, A.N., Veldhuizen, D., Scholten, W.J., Maertens, J., Breems, D.A., Pabst, T., Manz, M.G., van der Velden, V.H.J., Slomp, J.J.J.J., Preijers, F.W.M.B., Cloos, J., van de Loosdrecht, A.A., Löwenberg, B., et al (2018) CD34+CD38- leukemic stem cell frequency to predict outcome in acute myeloid leukemia. *Leukemia*, **33,** 1102–1112 Available at: http://www.ncbi.nlm.nih.gov/pubmed/30542144 [Accessed March 5, 2019].

Zeijlemaker, W., Kelder, A., Scholten, W.J., Snel, A.N., Veldhuizen, D., Cloos, J., Ossenkoppele, G.J. & Schuurhuis, G.J. (2016) A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. *Leukemia*, **30,** 439–46 Available at: http://dx.doi.org/10.1038/leu.2015.252.

Zeijlemaker, W., Kelder, A., Wouters, R., Valk, P.J.M., Witte, B.I., Cloos, J., Ossenkoppele, G.J. & Schuurhuis, G.J. (2015) Absence of leukaemic CD34+ cells in acute myeloid leukaemia is of high prognostic value: A longstanding controversy deciphered. *British Journal of Haematology*, **171,** 227–238.

CHAPTER 6

AML/NORMAL PROGENITOR BALANCE INSTEAD OF TOTAL TUMOR LOAD (MRD) ACCOUNTS FOR PROGNOSTIC IMPACT OF FLOWCYTOMETRIC RESIDUAL DISEASE IN AML

Diana Hanekamp, Jesse M. Tettero,
Gert J. Ossenkoppele, Angèle Kelder,
Jacqueline Cloos, Gerrit Jan Schuurhuis

Cancers (Basel). 2021 May 26;13(11):2597

ABSTRACT

Measurable residual disease (MRD) in AML, assessed by multicolor flow cytometry, is an important prognostic factor. MRD is commonly calculated relative to white blood cells (WBC-MRD). MRD positivity is defined as MRD≥0.1% by European LeukemiaNet. The AML part of the primitive/progenitor compartment (CD34+,CD117+,CD133+) is part of MRD, and referred to as primitive marker MRD (PM-MRD), and together with the size of total progenitor compartment (PM as % of WBC) constitutes MRD: PM-MRDxPM%=WBC-MRD. We explored the relative contribution of both parameters to MRD.

In HOVON/SAKK study H102 (300 patients), analyses showed that the PM% parameter itself had no prognostic impact, implying that prognostic impact of MRD resulted from the PM-MRD parameter only. Based on two objectively assessed cut-off points (1.62% and 10%), PM-MRD offered an independent prognostic parameter and identified three patient groups with different prognosis. The use of PM% even resulted in MRD false-negativity/positivity in part of patients. In particular, in the European LeukemiaNet based MRD negative (MRD<0.1%) group a subgroup with poor prognosis (n=31) was identified based on PM-MRD≥10%. This MRD false negativity urges to consider PM-MRD for MRD analysis. Prognostic impact of MRD may thus mainly originate from AML progenitor load (PM-MRD) instead of total leukemic load (WBC-MRD).

INTRODUCTION

In the treatment of acute myeloid leukemia (AML), assessment of measurable residual disease (MRD) is of high importance. Nowadays, MRD is assessed with molecular methods (PCR and next generation sequencing), and for the majority of cases, multicolor flow cytometry (MFC). MRD thus assessed was found highly prognostic for disease outcome in many studies[1–6]. MFC-MRD is classically defined as number of cells with a leukemic associated immunophenotype (LAIP), usually expressed as percentage of the complete white blood cell (WBC) population. European Leukemia Net (ELN) recommendations define MRD positivity as ≥0.1% $LAIP^+$ cells/WBC[7]. While MFC offers the major method to assess MRD, variation among laboratories is high[8]. Therefore, there is urge for harmonization and, wherever possible standardization, of assays. Whereas standardization for sample acquisition, shipment and storage can be achieved by strict rules[9–12], the standardization in analysis and report of MRD remains a challenge[7].

Major criticism on MFC-MRD is that it is relatively subjective, since it requires extensive knowledge of normal bone marrow (NBM) differentiation patterns[7]. Furthermore, one should realize that a LAIP may not cover the whole leukemic population, and therefore may lead to under-estimation of the leukemic load at follow up. Moreover, LAIPs may change during/after therapy[13–15]. In addition, background of LAIP antigens on normal cells may lead to low LAIP specificity and thereby low sensitivity in the detection of specific, i.e. AML defining, LAIPs[16–18].This is especially important at lower MRD levels, when background levels may approach patient MRD levels.

Recently, we postulated an alternative MRD approach that, with relatively simple mathematics, potentially enabled simplification and better standardization of MFC-MRD[16], because it allows to quantitatively and more objectively assess MRD. Also, since it is known that CD34-positive, CD117-positive and CD133-positive leukemia compartments contain normal progenitors and leukemia initiating and -propagating cells[19–21], we hypothesized that such biologically important cells may be equally informative for relapse initiation as compared to total leukemic load that includes non-dividing more mature cells. This would circumvent the use of total progenitor frequencies and thereby the use of WBC count. Since this WBC parameter is subject to large variations, amongst others depending on sample transport/storage, such would be very important for standardization/harmonization of MRD.

In the present study, we tested MRD calculated on primitive marker (PM-MRD) and compared it to the standardly used classical MRD that is calculated on the total WBC compartment. We further investigated the quantitative effects of variables that contribute to total MRD, and showed that prognostic impact of MRD mainly originates from prognostic impact of PM-MRD. We did this by re-analyzing the recent Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON) and Swiss Group for Clinical Cancer Research (SAKK) clinical trial 102[22,23], which encompassed 300 patients.

PATIENTS, MATERIALS AND METHODS

Patients and normal controls. For inclusion criteria and treatment protocols of patients we refer to two recently published papers[22,23]. Background levels for MRD and PM-MRD were analyzed in ten normal bone marrows (NBM), obtained after informed consent from (age-matched) healthy donors (median age 71 years, range 58-77) undergoing cardiac surgery. To evaluate the effect of therapy on background expression levels, additionally regenerative bone marrow (RBM) was analyzed in eight patients. Details are in Suppl. text 5.

IMMUNOPHENOTYPING

Immunophenotyping was performed as previously described[24]. Flow cytometry was performed on FACS CANTO (BD Biosciences, San Jose, CA, USA) with either 6- or 8-color antibody panels (see supplemental table 2 in Zeijlemaker, et al.[23]). Analyses of flow cytometric measurements were performed with Infinicyt™ software (Cytognos, Salamanca, Spain). LAIPs used in this study and frequencies of use are shown in Suppl. Table 1.

MRD assessment in the present study was based on LAIP approach. Gating was performed following strict criteria concerning forward/side scatter properties of the LAIP, expression pattern of CD45, primitive marker (PM) expression (i.e. CD34, CD117 and/or CD133), and expression of a myeloid marker (i.e. CD13, CD33 and/or HLADR), to exclude a-specificity and non-relevant cell types. In accordance with the biological role of cells of the PM cells, CD34 was used in most LAIPs in CD34 positive AML, CD117 and CD133 in LAIPs in CD34low/neg AML cases. NBM and RBM were gated according to the strict standard MRD gating procedure[10]. Details regarding LAIP background in NBM and RBM and application in MRD calculations are in Supplementary text 5.

A QUANTITATIVE APPROACH FOR MRD CALCULATION

The quantitative model for MRD is shown in Suppl. Figure 1A and, in more detail, in Supplementary Figure 1 of Schuurhuis et al.[16], and hypothetically contains all factors contributing to final MRD: specificity of the LAIP (termed f), size of the total PM compartment in MRD sample (termed g), percentage of the PM compartment that is covered with LAIP at time of diagnosis (termed h), and LAIP coverage of the PM compartment at time of MRD sampling (d+b).

Classical MRD (MRDclass) is defined as used in previous studies[24,25], which uses aberrancies on primitive marker (PM, CD34/CD117/CD133) cells with inclusion of the total (AML+normal) PM compartment (as % of WBC, referred to as PM%), and in which we corrected MRD for h in part of the cases. Calculated MRD (MRDcalc) uses only aberrancies on PM cells and PM%. When subsequently including f, h or f+h, MRD is termed MRDcalc(with f), MRDcal(with h), and MRDcalc(with f and h), respectively.

The established model for the AML part of the PM compartment (referred to as PM-MRD) is shown in Suppl. Figure 1B and, in more detail, in Supplementary Figure 1B of

Schuurhuis et al.[16], and hypothetically contains all factors contributing to final PM-MRD: i.e. specificity of the LAIP (termed f), percentage of the PM compartment that is covered with LAIP at time of diagnosis (termed h) and LAIP PM coverage at time of MRD sampling (d+b, now termed p). Since PM-MRD is a new parameter, there is no "PM-MRDclass", and we therefore simply use the term PM-MRD instead. Inclusion of f, h and f+h are referred to as PM-MRD(with f), PM-MRD(with h), and PM-MRD(with f+h). To assess LAIP specificity (f), we analyzed the same ten NBM and eight RBM samples as for MRD.

It is important to note that the relationship between MRD and PM-MRD in the present study is: MRD=PM-MRD x PM percentage (all at the MRD/PM-MRD assessment time point). No more mature populations beyond the progenitor stage are considered.

STATISTICAL ANALYSIS

In survival Kaplan-Meier analyses, event free survival (EFS) was used, defined as time between sampling after complete remission (CR) and date of relapse/progressive disease or death. Patients with no event were censored at date of last follow-up. Kaplan-Meier analyses were performed using the survival R package. Statistical analyses were performed using SPSS version 22.0 software. Outcome between groups was compared using log-rank test. The MaxStat package of R identified optimal cut-off point(s) for PM-MRD. The prognostic value of both MRD (MRDclass. and MRDcalc) and PM-MRD (cut-off 1.62% and 10.0%) for EFS was investigated in multivariable models including variables that were significant in univariate analyses.

RESULTS

SPECIFICITY: EXPRESSION OF LAIPS ON NON-LEUKEMIC WBC

Knowledge of background LAIP expression patterns on normal cells is required to avoid false positive results. Background may differ between LAIPs[17], and was therefore established for each different LAIP. The LAIPs used are shown in Figure 1A and Suppl. Table 1. Range of background levels in NBM controls is 0.000-0.036%; median: 0.003% (Figure 1A). All are far below 0.1%, which confirms the relative solidity of this ELN consensus threshold to distinguish $MRD^{low/neg}$ from $MRD^{high/pos}$,[7] and thereby suggests that, for a large part of the LAIPs, reliable MRD assessment is possible below the 0.1% threshold level.

MRD DEFINITIONS AND CALCULATIONS

In order to gain insight in the contribution of LAIP specificity (f) and LAIP coverage (h) to MRD calculation (for details: Supplementary text (paragraphs 1-3), we first evaluated the accuracy of the quantitative model. The H102 data were re-analyzed in the most basic version (MRDcalc, with no correction for f and h) and compared this to the MRD as we used in previous studies[24] (termed MRDclass), using the ELN consensus of 0.1 (% of WBC) cut-off(7). As expected, MRDcalc (Figure 2B) performs largely similar to the already

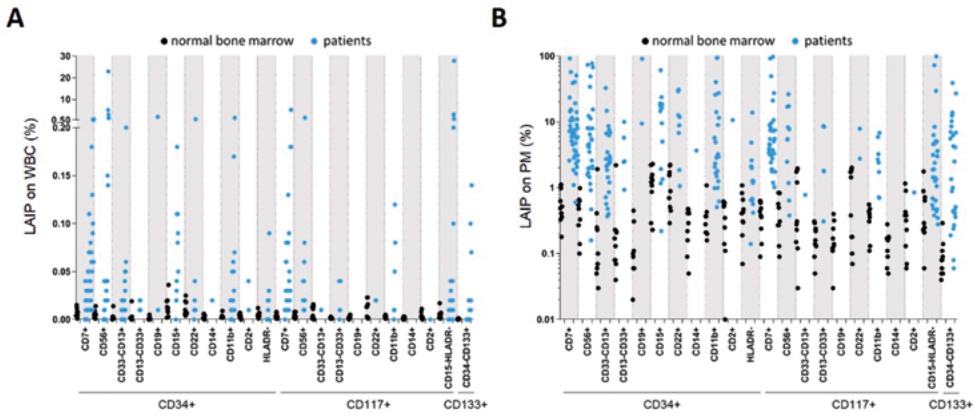


Figure 1. Background expression of LAIPs in normal bone marrow and patient samples. Background levels for (A) MRD and (B) PM-MRD were analyzed in ten normal bone marrow samples (in black); In blue are patients included in the HOVON SAKK 102 clinical trial. Immunophenotyping was performed as previously described[23,24]. Background and LAIPS were gated following strict criteria concerning forward/side scatter properties, expression pattern of CD45, primitive marker expression, and expression of a myeloid marker to exclude a-specificity[10,24]. (A) Background expression was median 0.003 (range 0%-0.036%). (B) Background levels of LAIPs within the primitive compartment range 0.01-2.30 (median 0.29%) in NBM. For reasons of clarity, the presentation of LAIPs has been simplified by not including the myeloid marker used (CD13 or CD33).

prospectively validated MRDclass[24] (Figure 2A). Also, both MRDclass and MRDcalc allowed to define multiple threshold levels that resulted in multiple patient groups with different EFS (Figure 2D and 2E, respectively).

Next, we evaluated the MRDcalc model with incorporation of specificity f (further defined as median expression value for each individual LAIP in NBM, see also Supplementary text 1). This revealed largely similar prognostic value compared to MRDcalc (Figure 2C; compare with Figure 2B). Also for MRDcalc(with f) multiple cut-offs delivered further prognostic information (Figure 2F). Lastly, implementation of the coverage factor at diagnosis (variable h) in the MRDcalc model, resulted in clearly inferior models (Supplementary text 2 and Suppl. Fig. 2 A-D), as did the combination of f and h (Suppl. Fig. 2 D-F). Importantly, MRDcalc and MRDcalc(with f) performed quite similar predominantly at relatively high MRD levels.

SPECIFICITY: EXPRESSION OF LAIPS ON NON-LEUKEMIC PROGENITORS

It has been previously suggested that MRD expressed as percentage of the progenitor compartment (CD34+and/orCD117+and/orCD133+; termed PM-MRD) may be more easy to standardize[16,25]. Therefore, we employed this dataset and mathematical model to investigate this further. First, we looked at the specificity by determining the background levels of LAIPs within these primitive compartments in NBM (Figure 1B: median 0.29%;

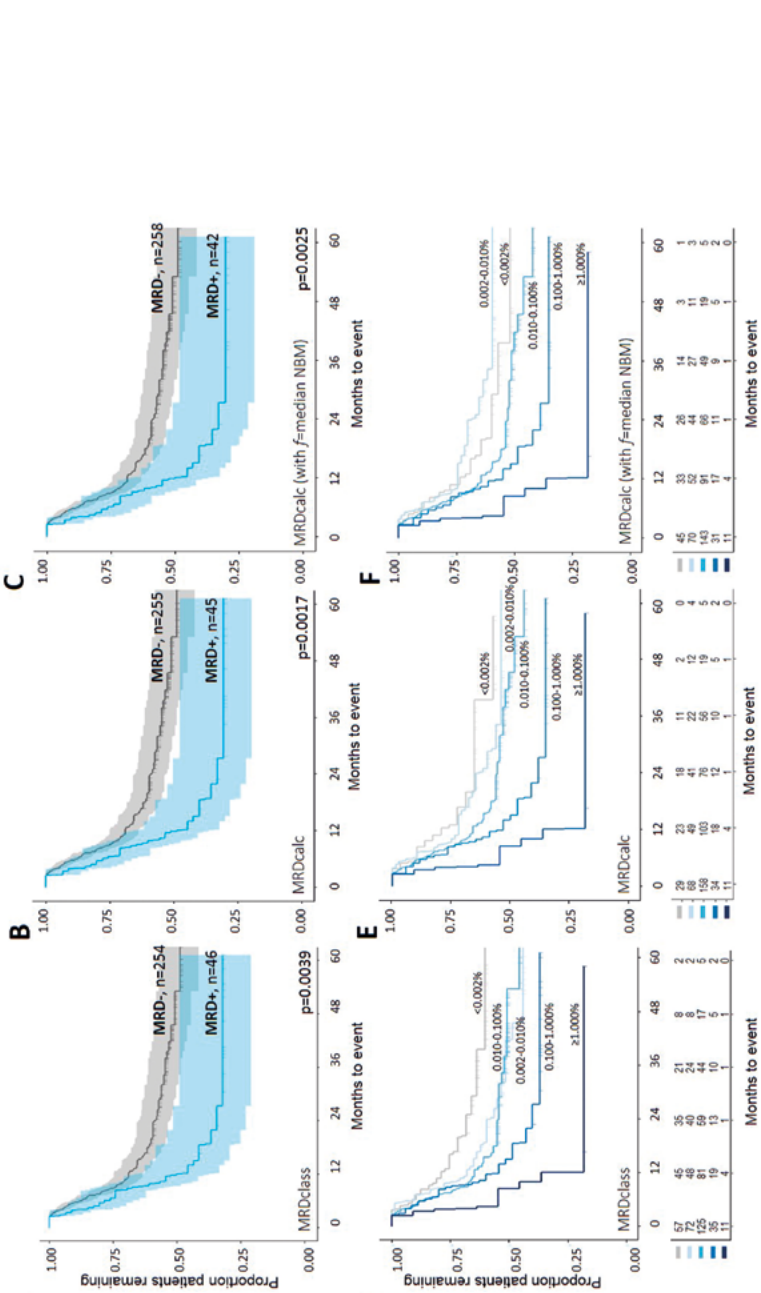


Figure 2. Calculated and classical MRD detection. Kaplan-Meier analyses comparing EFS of (WBC) MRD with or without application of the quantitative model. (A) Kaplan–Meier curve comparing EFS of classically used WBC MRD according to 0.1% cut-off (defining two patient groups: (MRD+ or MRD-). (B) Event free survival of AML patients based on calculated MRD (MRDcalc+ or MRDcalc-; without inclusion of variables f and h, and according to 0.1% cut-off). Patient groups are slightly different compared to MRDclass (resulting from round of and from subjective use of correction factor(23))). (C) Event free survival of AML patients based on calculated MRD (MRDcalc+ or MRDcalc-; with inclusion of variable f, and according to 0.1% cut-off). (D-F) Additional cut-offs in either (D) MRDclass, (E) MRDcalc without f, and (F) MRDcalc with f, show prognostic value in sub-groups defined within both the originally MRD+ group (≥0.1%; 2 sub-groups identified), and in the originally MRD-negative group (<0.1%; 3 sub-groups identified).

range 0.01-2.30). In contrast to MRDcalc PM-MRD can reach high levels to maximally 100% of LAIP positive cells in the PM compartment). Differences in background for different LAIPs are observed (for example CD34-/CD133+ LAIP had a higher than ten-fold lower background compared to the CD34+CD15+ LAIP), supporting the idea that these differences should be taken into account[16], especially in the lower ranges of PM-MRD (Figure 1B; more detailed explanation in Supplementary text 1). Since the primitive compartment is part of MRDcalc (MRD=PM-MRD x PM percentage), also for MRDcalc the backgrounds on the primitive cells do affect MRDcalc in its lower ranges (say 0.00%-0.03%, Figure 1A).

PM-MRD CALCULATION AND PROGNOSTIC IMPACT

We next explored the clinical relevance of considering the prognostic value of PM-MRD. The main parameters here are similar to MRDcalc, but, in contrast to the latter, the PM% is completely excluded (Patients, materials and methods and Suppl. Fig. 1B). Since in the previous paragraph we showed that f was important for the PM-compartment in MRD (Figure 1B), we included f in the analysis: PM-MRD(with f). Using MaxStat statistics two optimal cut-off levels could be identified (1.62% and 10%; Figure 3A) that each allowed identification of two distinct patient populations with different EFS (Figures 3B and C). Together the two cut-offs thus allowed discrimination of three patient groups with EFS 64.4% (n=101), 47.8% (n=134) and 33.8% (n=65) (Figure 3D). For MRD, no objective cut-off points could be identified using MaxStat statistics. To better compare with PM-MRD, we used, apart from the ELN consensus cut-off of 0.1% (results in Figure 2), an extra lower cut-off of 0.01%, subjectively assessed to be able to identify a good prognostic patient group with about similar number of patients as the best-performing PM-MRD group (<1.62%). Together the two MRD cut-offs identified patients with EFS of 59.1% (n=115), 49.0% (n=143) and 31.0% (n=42), for the three MRD groups (Figure 3E). Similar to MRDcalc (Figure 2F), PM-MRD allows to discriminate other patient groups with additional cut-offs (Figure 3F). However, below 1% PM-MRD, which corresponds to about 0.01% MRD, no further consequent differences in prognostic impact were observed (not shown). This makes sense since the LAIP backgrounds are in general in between 1% and 0.1% (Figure 1B). Suppl. Table 2 shows that PM-MRD is an independent prognostic factor. Similar to MRDcalc, the inclusion of variable h (± f) did not contribute to a better performing model (see Supplementary text 3 and Suppl. Fig. 3). Detailed comparison of PM-MRD and MRD is given in the next paragraph. As argued in the previous paragraph, correction of background by definition will have the most effect at low PM-MRD values, where PM-MRD approaches f (for details see Supplementary Text 1).

Previously we have shown that the 0.1% MRD cut-off defines patient groups with different outcome in different cytogenetic risk groups[24]. When applying this to PM-MRD, it turns out that the two cut-offs define three patient groups with different survival especially in the intermediate and poor cytogenetic risk groups (Figure 3 G-I; see also Supplementary text 4). For comparison, MRD (0.1% cut-off) is shown in Suppl. Fig. 4.

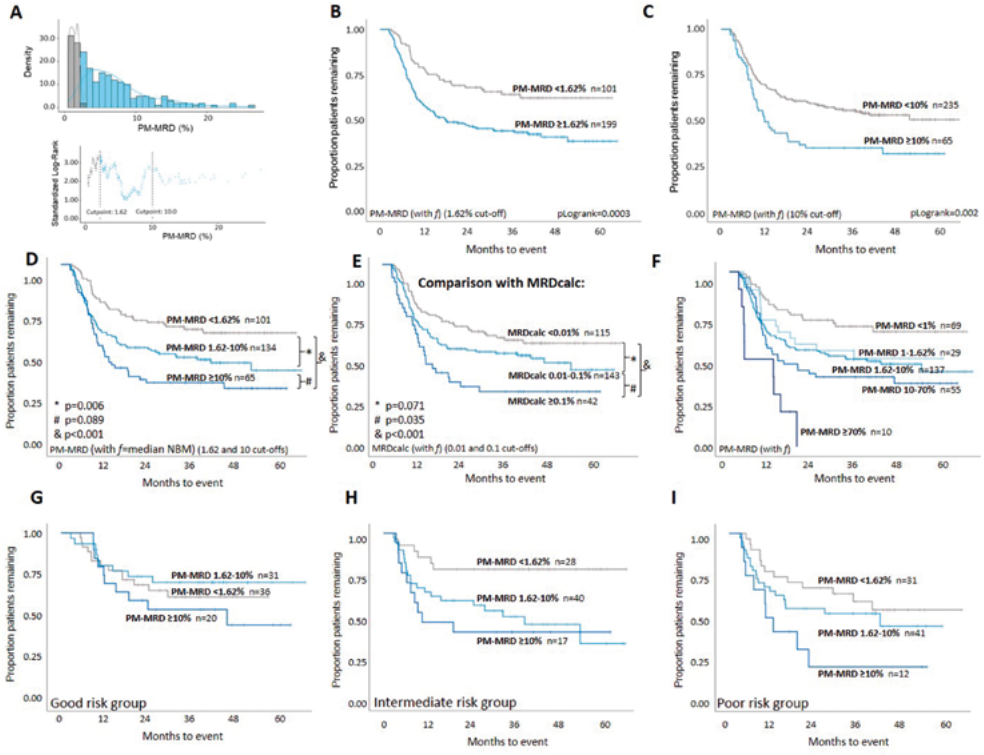


Figure 3. Prognostic impact of different PM-MRD cut-offs and corrected for background f. (A) Using standard Log-Rank analysis the optimal cut-offs of 1.62% and 10.0% were determined. (B) Using the cut-off of 1.62% allows discrimination of two patients groups with different EFS: 178 patients (59.3%) were PM-MRD(with-f)^{positive} with an EFS of 43.2%, versus 122 patients (40.7%) PM-MRD(with-f)^{negative} with an EFS of 64.4% (pLogrank= 0.0003). (C) For the 10.0% cut-off, these groups were 65 patients (21.7%) PM-MRD(with-f)^{positive} with an EFS of 33.8%, versus 235 patients (78.3%) PM-MRD(with-f)^{negative} with an EFS of 54.9% (pLogrank= 0.0016). (D) When both are combined, three patient groups with different EFS are shown: EFS 64.4% (n=101), 47.8% (n=134) and 33.8% (n=65). (E) For comparison, for MRD no such clear objective cut-offs could be defined. (F) Multiple cut-offs shows that PM-MRD refines prognostic impact beyond the application of the two objectively defined cut-offs. (G) Cut-off in Figure 3D applied for the cytogenetically favorable prognosis group (H) in the intermediate and (I) poor cytogenetic risk group.

It may be argued that more mature AML may not be suitable for the progenitor approach. However, our cohort included 46 mature AML cell patients (15.3%), classified as French–American–British (FAB) classifications M5 (n=38), M6 (n=6) or M7 (n=2), which all had primitive compartments that could be evaluated for LAIP expression. MRDcalc(with f) and PM-MRD(with f) (cut-offs at 0.1% and 10.0%, respectively) allowed to identify patients groups with distinct different EFS (pLogrank=0.009 and 0.001, respectively) (Suppl. Fig. 7).

COMPARISON OF PROGNOSTIC IMPACT OF PM-MRD VERSUS MRD

We have shown that both PM-MRD and MRD show prognostic value, however, with clear differences (Figures 2 and 3), which are reflected in the moderate correlation between PM-MRD and MRD ($r=0.737$, $p<0.001$). This implies that the PM%, which is the only difference between PM-MRD and MRD (MRD=PM-MRD x PM%), contributes to the discrepancies seen in prognostic value. However, in contrast to PM-MRD and MRD, there was no correlation between PM% and prognostic value (EFS, $r=0.039$; $p=0.512$). To illustrate this, in Figure 4A and B the whole relevant patient group (with PM% 0%-3.2%) was sub-divided in six PM% sub-groups. This did not result in consistent differences between the groups in EFS, in both Kaplan-Meier analysis (Figure 4A) and EFS at 36 months (Figure 4B).

This implies that the PM% does not contribute to the prognostic impact of MRD, which in turn would suggest that the prognostic impact of MRD would originate entirely from the PM-MRD component in it. To better understand this unexpected phenomenon, we explored the relationship between both PM-MRD and MRD on the one hand, and PM% on the other hand. Whereas PM-MRD as such inversely correlates with EFS in the whole unbiased patient group (resulting in the prognostic impact as seen in Figure 3), Figure 4C shows that PM-MRD was largely independent of PM%, which, in terms of prognostic impact, is in line with the lack of correlation between PM% and EFS (as seen in Figure 4A,B). In sharp contrast, MRD significantly increases with increasing PM-percentage (Figure 4D). These observations thereby reveal a strong discrepancy between MRD and EFS, but not between PM-MRD and EFS. This in turn suggest that the inclusion of the PM% in the calculations (leading to MRD) seems superfluous, and, in addition, may also cause MRD false negativity in part of the cases with very low PM%, or too high MRD values in part of the cases with high PM% (see Suppl. Text 6 for more details).

Supplementary Figures 9C-E confirm this finding: these show the distribution of individual patient values for EFS, PM-MRD, and MRD over the different PM% groups.

The quantitative contribution of this PM% parameter to the MRD parameter, as it is presently used in the clinic, should not be under-estimated: the difference between the patient with the lowest PM% (<0.01%) and highest PM% (≥3.2%) is a factor >320, compared to a factor 100 for PM-MRD. The main conclusion is therefore that it may not predominantly be the total leukemic load (MRD) that defines patient's outcome in remission AML, but the contribution of the AML part of the total progenitor compartment (PM-MRD), or in other words, the balance between AML and normal progenitors.

The clinically most important consequences for the present application of MRD concerns the prognostic meaning of the ELN consensus MRD cut-off of 0.1%: Figure 4E shows that, in the current study, the PM-MRD≥10%/MRD≥0.1% compartment contains 25 patients, who indeed are characterized by poor prognosis (see Kaplan-Meier plots at the right in the Figure, with further explanations in the legends). However, there is a PM-MRD≥10%/MRD<0.1% compartment with poor prognostic based on PM-MRD, but good prognostic based on MRD. This compartment contains 31 patients, now showing that

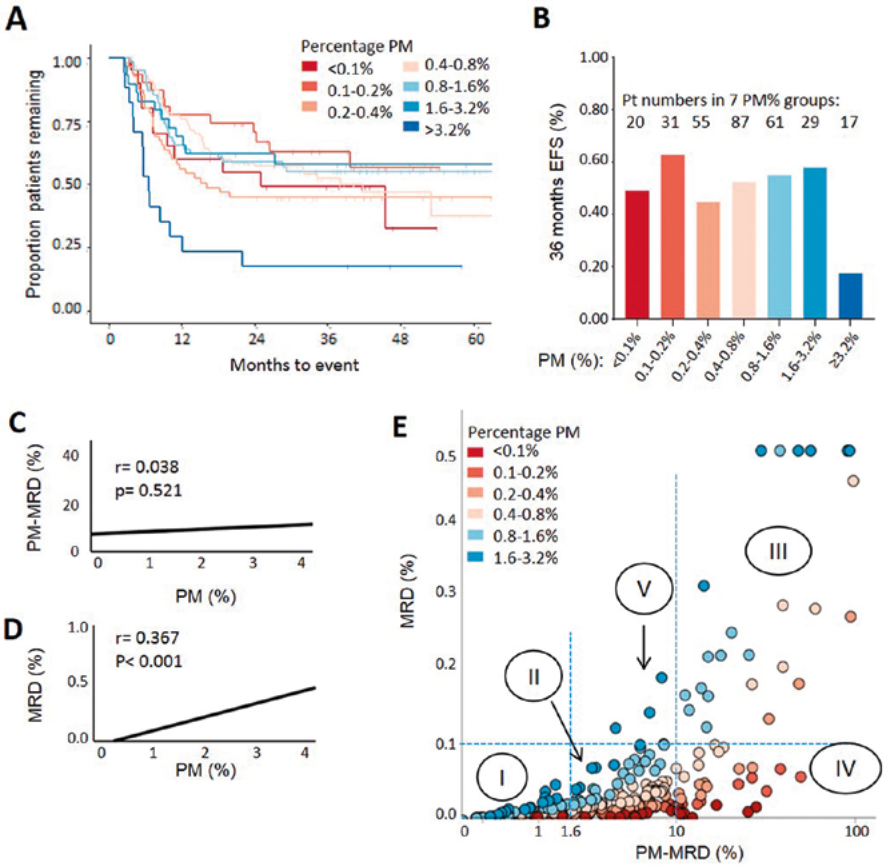


Figure 4. Role of the percentages of the whole (AML+normal) PM compartment ("PM%") in the prognostic value of PM-MRD and MRD. The PM% was subdivided in the following groups, <0.1%, 0.1%-0.2%, 0.2%-0.4%, 0.4%-0.8%, 0.8%-1.6%, 1.6%-3.2% and ≥3.2%. (A) Kaplan-Meier analysis shows that there was no consistent difference in EFS between the patients in the different PM% groups. Except for the group with PM% ≥3.2%, wherein 4/17 samples were from patients who had officially relapsed at the time point of sampling. (B) EFS at 36 months in the different PM% groups (data extracted from all individual patient data used in Figure 4A). (C) Correlation between PM-MRD and PM% (in the range 0%-3.2%) and (D) between MRD and PM%. (E).

over 50% of the patients with poor prognosis based on high PM-MRD percentage, are not accounted for by MRD≥0.1 %. The reason for this is that (very) low PM percentages (especially in the groups <0.1%-0.4%, see brownish symbols in Figure 4E) may result in low MRD values (PM-MRD x PM% <0.1), despite high, prognostically unfavorable, PM-MRD values (≥10), and thereby represent MRD false-negativity. A tentative way how PM-MRD may be reported to clinicians is outlined in Suppl. Text 7 and related Suppl. Tables.

DISCUSSION

This paper consists of three parts in which we have used the data from the clinical HOVON/SAKK 102 study[22] for which we already assessed prognostic impact for MRD[23] and which encompassed 300 evaluable patients. First we demonstrate the application of a simple mathematical model to objectively calculate MRD, i.e. to enumerate all parameters contributing to MRD, thereby enabling to report MFC-MRD more objectively. The most important parameters turned out to be i) the fraction of the (CD34+and/orCD117+and/orCD133+) progenitor compartment that, under MRD conditions, is covered with LAIP, ii) the background of a LAIP defined in normal bone marrow (termed f), and iii) the total (AML+normal) progenitor compartment as a percentage of WBC (PM%).

The second part introduces PM-MRD which is defined as the AML part of the total progenitor compartment and which is related to MRD by the formula MRD = PM-MRD x PM%. When using the mathematical model, it is revealed that PM-MRD offers a strong independent prognostic factor, defining three patient groups with different survival, and which in the lower regions, but especially at the higher ranges, may perform better than MRD. Of note, for MRD no such clear objective cut-offs could be identified: the cut-off of 0.1% is the result of a consensus within European LeukemiaNet and not objectively assessed.

In the third part we show that the prognostic impact of MRD is mainly, if not completely, due to the prognostic impact of PM-MRD component in it; the PM% part of MRD does not add to the prognostic impact of MRD, but instead in part of cases produces artificial and highly variable increases of PM-MRD, which may lead to either false-positivity or MRD false-negativity. This can be most accurately assessed for MRD false-negativity: when using the MRD consensus cut-off of 0.1%, the poorest prognostic PM-MRD group (≥10%) not only harbours almost all high MRD cases (≥0.1%), but in additions contains even more MRD low (<0.1%), but nevertheless poor prognosis, patients.

At this stage it must be emphasized that both the progenitor make up and the PM% made large contributions to the total tumour load (MRD): PM-MRD ranged from about 1% to 100% (factor 100), while the difference between the highest and lowest PM% values were factor >320 (range <0.01% - 3.2%). As a result in the formula MRD=PM-MRD x PM% the differences in MRD between lowest and highest value may be a factor of at least 10^4. On the other hand, for PM-MRD the difference between the highest (100%) and lowest (1%) is only a factor 100 at maximum, apparently without loss of prognostic information.

We conclude that PM-MRD seems to be the main, if not only, prognostically important part of MRD, which observation, especially seen the questionable role of PM%, should be taken into account when using classical MRD. Total tumour load in solid tumours, both before and especially after therapy, has a direct relationship with patient prognosis[26], and, seen the clinical impact of MRD in defining risk and/or adapt therapy, also for hematological tumours including AML. Despite the fact that primitive AML cells (progenitors and/or stem cells), when expressed as total load (per WBC), have been shown to have prognostic

impact[23,25,27–29], as far as we know the present paper for the first time in AML shows that it is not mainly the total tumour burden (MRD), but the balance between tumour and normal progenitor cells that is important for clinical outcome. It is tempting to speculate that at a certain threshold the fast growth of acute leukemic progenitors[19–21], may overrule initial differences in total tumor load (i.e. dividing plus non-dividing AML cells as percentage of WBC). The data retrospectively confirm one of our earlier HOVON/SAKK studies in which we examined in a small patient group (n=77) the added prognostic impact of PM-MRD (then called aPC fraction: aberrant primitive compartment) on the MRD negative patient group (MRD<0.1%)[25]. When re-considering the data, poor prognosis for the whole patient group (MRD+ plus MRD-) can almost fully accounted for by aPC ≥10% (roughly containing 65% MRD positive and 35% MRD negative cases).

Apart from this new insight in how to define residual disease, a major advantage of the PM-MRD approach is that it circumvents the use of the total progenitor load (PM%, i.e. CD34and/orCD117and/orCD133 as percentage of WBC), and thereby thus also of the WBC load. The uncertainty of WBC count, either resulting from the gating procedure of all viable WBC as such, or as a result of selective degradation of WBC sub-populations (e.g. granulocytes) upon sample storage or transport[9,16], will affect correct MRD quantitation, and thus makes PM-MRD potentially a more reliable measure for residual disease quantitation. As such, the wish for standardization of MFC MRD[16] may become in reach.

PM-MRD may also contribute to a more objective identification of MRD, independent of the large personal experience required for the classical MRD studies in which identification of aberrancies in differentiation patterns in AML BM, compared to normal BM, fulfilled a central role. Identification of progenitor compartments in follow-up BM is of wide-spread use in diagnostics and treatment, e.g. CD34 expression in stem cell transplantation[30,31]. If a consensus list of potential LAIPs would become available, this would be useful for new participants in the field. The putative contents of such a list is described in Suppl. Text 1, last paragraph. Progenitor identification even better allows the different-from-normal approach for all types of aberrant cells, which has been urged by European LeukemiaNet[7]. PM-MRD likely could be of major importance especially in multi-centre clinical studies aiming at refinement of risk assessment, to guide follow up therapy, or for use as a short term endpoint for survival.

In our results, we showed the redundancy of the previously used correction factor (termed h) to account for the percentage of LAIP covering blast or PM cells at time of diagnosis and translated to the MRD situation[3]. With the increasing knowledge on post-diagnostic immunophenotypic changes[13], uncertainly about the maintenance at follow up of diagnosis make up has become too large.

In our analyses we were not able to assess prognostic impact of PM-MRD below 1% because of LAIP background values (factor "f") becoming of similar magnitude as the PM-MRD values. In parallel the MRD level that roughly coincides with that 1% PM-MRD value, was 0.01%, below which also no clear prognostic impact was seen (Figure 2F). A more sophisticated background correction would be not to subtract the whole LAIP+

background, but the local background instead, i.e. only in the area in which (PM)-MRD is located and which may depend on the type of LAIP, on the patient and on the time point of MRD assessment. Such approaches are underway.

It remains to be seen how PM-MRD relates to Leukemia Stem Cell (LSC) MRD as reported before by us[23]. In that paper we show that classical MRD and LSC-MRD are supplementary.

In conclusion, we have introduced a mathematical model to calculate residual disease in an objective, quantitative, and reproducible way. The model also reveals a new way to calculate residual disease, i.e. PM-MRD, describing the make-up of the progenitor compartment in terms of normal and AML progenitors. PM-MRD not only reveals the presence of multiple patient groups with different prognosis, but also showed to represent the major prognostic part of MRD. The approach also reduces false-negativity as seen in classical MRD. We suggest current MRD institutes to verify these findings in their own databases (a handy Excel file to quickly calculate MRD and PM-MRD has been supplied as a supplementary file), which usually are already available herein (at least for CD34 and CD117) as part of the regular MRD measurements.

ACKNOWLEDGEMENT

This work was supported by Netherlands Cancer Foundation/Alpe d'HuZes, project: ALPE-2013-6371. We thank all participating centers of the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK), and the Dutch AML-MRD working party (Vincent H.J. van der Velden, Rik A. Brooimans, Jeroen G. te Marvelde (Erasmus University Medical Center), Jennichjen Slomp, Antoinette Heijs (Medisch Spectrum Twente, Medlon), and Frank Preijers and Erik Huys (Radboud University Medical Center) for flow cytometric analysis and efforts for a standardized protocol within the Netherlands.

REFERENCES

1. Venditti a, Buccisano F, Del PG, Maurillo L, Tamburini a, Cox C, et al. Level of minimal residual disease after consolidation therapy predicts outcome in acute myeloid leukemia. Blood. 2000;96(0006-4971 (Print)):3948–52.

2. San Miguel JF, Vidriales MB, López-Berges C, Díaz-Mediavilla J, Gutiérrez N, Cañizo C, et al. Early immunophenotypical evaluation of minimal residual disease in acute myeloid leukemia identifies different patient risk groups and may contribute to postinduction treatment stratification. Blood 2001;98(6):1746–51.

3. Feller N, van der Pol M a, van Stijn A, Weijers GWD, Westra AH, Evertse BW, et al. MRD parameters using immunophenotypic detection methods are highly reliable in predicting survival in acute myeloid leukaemia. Leukemia 2004;18(8):1380–90.

4. Kern W, Voskova D, Schoch C, Hiddemann W, Schnittger S, Haferlach T. Determination of relapse risk based on assessment of minimal residual disease during complete remission by multiparameter flow cytometry in unselected patients with acute myeloid leukemia. Assessment. 2004;104(10):3078–85.

5. 5. Buccisano, F., Hourigan, C. S. & Walter, R. B. The Prognostic Significance of Measurable ("Minimal") Residual Disease in Acute Myeloid Leukemia. Curr. Hematol. Malig. Rep. 12, 547–556 (2017)

6. Hourigan, C. S., Gale, R. P. & Walter, R. B. Refining AML outcome prediction. Leukemia 33, 283–284 (2019).

7. Schuurhuis GJ, Heuser M, Freeman S, Béné M-C, Buccisano F, Cloos J, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood [2018;131(12):1275–91.

8. Keeney M, Halley JG, Rhoads DD, Ansari MQ, Kussick SJ, Karlon WJ, et al. Marked Variability in Reported Minimal Residual Disease Lower Level of Detection of 4 Hematolymphoid Neoplasms: A Survey of Participants in the College of American Pathologists Flow Cytometry Proficiency Testing Program. Arch Pathol Lab Med [015;139(10):1276–80.

9. Johansson U, Bloxham D, Couzens S, Jesson J, Morilla R, Erber W, et al. Guidelines on the use of multicolour flow cytometry in the diagnosis of haematological neoplasms. Br J Haematol 2014;165(4):455–88.

10. Cloos J, Harris JR, Janssen JJWM, Kelder A, Huang F, Sijm G, et al. Comprehensive Protocol to Sample and Process Bone Marrow for Measuring Measurable Residual Disease and Leukemic Stem Cells in Acute Myeloid Leukemia. J Vis Exp 2018;March 5(133).

11. Kalina T, Flores-Montero J, van der Velden VHJ, Martin-Ayuso M, Böttcher S, Ritgen M, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia 2012;26(9):1986–2010.

12. Lacombe F, Bernal E, Bloxham D, Couzens S, Porta MGD, Johansson U, et al. Harmonemia: a universal strategy for flow cytometry immunophenotyping—A European LeukemiaNet WP10 study. Leukemia 2016;30(8):1769–72.

13. Zeijlemaker W, Gratama JW, Schuurhuis GJ. Tumor heterogeneity makes AML a "moving target" for detection of residual disease. Cytom Part B Clin Cytom 2014;86(1):3–14.

14. Cui W, Zhang D, Cunningham MT, Tilzer L. Leukemia-associated aberrant immunophenotype in patients with acute myeloid leukemia: Changes at refractory disease or first relapse and clinicopathological findings. Int J Lab Hematol 2014;36(6):636–49.

15. Zhou Y, Moon A, Hoyle E, Fromm JR, Chen X, Soma L, et al. Pattern associated leukemia immunophenotypes and measurable disease detection in acute myeloid leukemia or myelodysplastic syndrome with mutated NPM1. Cytom Part B Clin Cytom 2019;96(1):67–72.

16. Schuurhuis GJ, Ossenkoppele GJ, Kelder AA, Cloos J. Measurable residual disease in acute myeloid leukemia using flowcytometry: approaches for harmonization/ standardization. Expert Rev Hematol 2018;11(12):17474086.2018.1549479.

17. Sui JN, Chen QS, Zhang YX, Sheng Y, Wu J, Li JM, et al. Identifying leukemia-associated immunophenotype-based individualized minimal residual disease in acute myeloid leukemia and its prognostic significance. Am J Hematol 2019;94(5):528–38.

18. Camburn AE, Petrasich M, Ruskova A, Chan G. Myeloblasts in normal bone marrows expressing leukaemia-associated immunophenotypes. Pathology 2019;51(5):502–6.

19. Quek L, Otto GW, Garnett C, Lhermitte L, Karamitros D, Stoilova B, et al. Genetically distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. J Exp Med 2016;213(8):1513–35.

20. Beghini A, Corlazzoli F, Del Giacco L, Re M, Lazzaroni F, Brioschi M, et al. Regeneration-associated WNT signaling is activated in long-term reconstituting AC133bright acute myeloid leukemia cells. Neoplasia 2012;14(12):1236–48.

21. Blair A, Sutherland HJ. Primitive acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo lack surface expression of c-kit (CD117). Exp Hematol 2000;28(6):660–71.

22. Löwenberg B, Pabst T, Maertens J, van Norden Y, Biemond BJ, Schouten HC, et al. Therapeutic value of clofarabine in younger and middle-aged (18-65 years) adults with newly diagnosed AML. Blood 2017;129(12):1636–45.

23. Zeijlemaker W, Grob T, Meijer R, Hanekamp D, Kelder A, Ham J, et al. CD34+CD38- leukemic stem cell frequency to predict outcome in acute myeloid leukemia. Leukemia [2018;33:1102–12.

24. Terwijn M, van Putten WLJ, Kelder A, van der Velden VHJ, Brooimans R a., Pabst T, et al. High Prognostic Impact of Flow Cytometric Minimal Residual Disease Detection in Acute Myeloid Leukemia: Data From the HOVON/SAKK AML 42A Study. J Clin Oncol 2013;31(31):3889–97.

25. Terwijn M, Kelder A, Snel AN, Rutten AP, Scholten WJ, Oussoren YJM, et al. Minimal residual disease detection defined as the malignant fraction of the total primitive stem cell compartment offers additional prognostic information in acute myeloid leukaemia. Int J Lab Hematol 2012;34(4):432–41.

26. Stein WD, Huang H, Menefee M, Edgerly M, Kotz H, Dwyer A, et al. Other Paradigms: Growth Rate Constants and Tumor Burden Determined Using Computed Tomography Data Correlate Strongly With the Overall Survival of Patients With Renal Cell Carcinoma. Cancer J 2009;15(5):441–7.

27. Terwijn M, Zeijlemaker W, Kelder A, Rutten AP, Snel AN, Scholten WJ, et al. Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. PLoS One 2014;9(9):e107587.

28. Zeijlemaker W, Kelder A, Cloos J, Schuurhuis GJ. Immunophenotypic Detection of Measurable Residual (Stem Cell) Disease Using LAIP Approach in Acute Myeloid Leukemia. Curr Protoc Cytom 2019;91(1).

29. Bradbury C, Houlton AE, Akiki S, Gregg R, Rindl M, Khan J, et al. Prognostic value of monitoring a candidate immunophenotypic leukaemic stem/progenitor cell population in patients allografted for acute myeloid leukaemia. Vol. 29, Leukemia. Nature Publishing Group; 2015. p. 988–91.

30. Murugesan M, Nair C, Nayanar S, Pentapati K. Flow cytometric enumeration of CD34+ hematopoietic stem cells: A comparison between single- versus dual-platform methodology using the International Society of Hematotherapy and Graft Engineering protocol. Asian J Transfus Sci. 2019;13(1):43–6.

31. Vaezi M, Shakouri M, Souri M, Setaredan MA, Mossavi SA, Mohammadi S, et al. The effects of lower CD34 yields after lowe dose G-CSF induction on long-term autologous stem cell transplantation outcome: A single center study. Transfus Apher Sci. 2018 Apr 1;57(2):265–71.

32. Zhou Y, Othus M, Araki D, Wood BL, Radich JP, Halpern AB, et al. Pre- and post-transplant quantification of measurable ('minimal') residual disease via multiparameter flow cytometry in adult acute myeloid leukemia. Leukemia 2016;30(7):1456–64.

33. Walter RB, Buckley SA, Pagel JM, Wood BL, Storer BE, Sandmaier BM, et al. Significance of minimal residual disease before myeloablative allogeneic hematopoietic cell transplantation for AML in first and second complete remission. Blood 2013;122(10):1813–21.

34. Lacombe, F., Dupont, B., Lechevalier, N., Vial, J. P. & Béné, M. C. Innovation in Flow Cytometry Analysis. HemaSphere 1 (2019) 3(2):e173. doi:10.1097/HS9.0000000000000173

SUPPLEMENTARY DATA

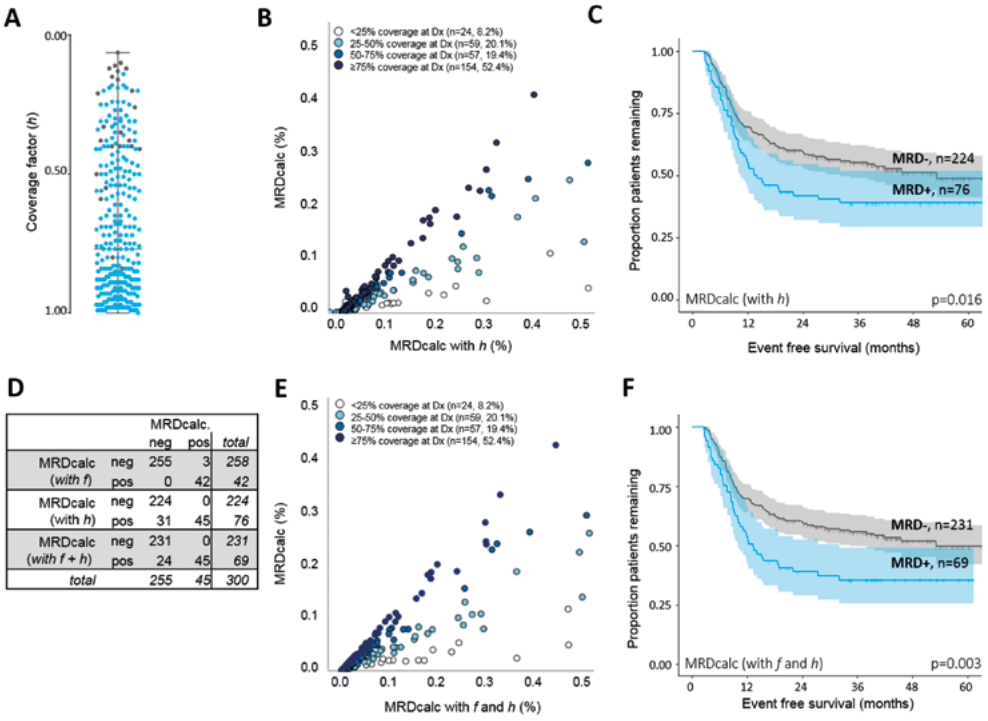
SUPPLEMENTAL TEXT 5. NORMAL BONE MARROW VERSUS POST-CHEMOTHERAPY BONE MARROW

In most studies, BM collected from healthy donors serves as reference for normal expression patterns. To assess remaining effects of such regeneration on LAIP expression in normal progenitors, we evaluated the expression of all LAIPs, used at the relevant MRD sampling points, but now from AML patients who had no LAIP at diagnosis and with no evidence for newly-formed of LAIPs on AML cells during/after therapy (regenerating bone marrow, RBM). These RBM patients were treated according to HOVON-SAKK clinical trial 132, which was METC approved (central study approval number 2013-539; VUmc local approval number 2014-228). RBM samples were collected after cycle II, identical to patients in the HOVON-SAKK clinical trials 102 and 132.

Suppl. Figure 5 and corresponding Suppl. Table 1 show the expression of LAIPs in RBM, with NBM shown too, and demonstrate that multiple LAIPs that compared to NBM are differently expressed (i.e. higher, with the exception of CD34+HLADR-) during regeneration. Since background levels in RBM are thus in general higher compared to NBM, and since this background is subtracted within the model, more patients are classified $MRD^{negative}$ compared to the model where NBM background levels are used. In WBC-MRD (Suppl. Figure 6), when using RBM, 38 patients are identified as $MRD^{positive}$ compared to 42 patients in MRDcalc. (Figure 2C).

In our study BM had regenerated before sampling and normal BM is likely the best control here. In other studies, when assessing MRD after the first induction cycle of chemotherapy, RBM is used[17,32,33]. Since therapies and thereby their effect on the BM are different for non-AML patients, we do not favor the use of background levels obtained in non-AML patients(6). However, there is no full guarantee that the selection of patients without LAIPs present at diagnosis and who received similar therapy as LAIP+ patients, do not have acquired LAIP expression on leukemic cells. This might have resulted from selection during therapy of very low frequent leukemic LAIP+ cells present at diagnosis. Therefore the choice for median frequencies in RBM to subtract from MRD data points seems better than subtraction of the mean, which may contain some relatively high values from cases with acquired LAIP expression on leukemic cells. Altogether, the choice for either NBM or RBM depends on status of BM regeneration, and is also important in light of future computational automated programs[34].

SUPPLEMENTAL FIGURES



		MRDcalc. neg	pos	total
MRDcalc (with *f*)	neg	255	3	*258*
	pos	0	42	*42*
MRDcalc (with *h*)	neg	224	0	*224*
	pos	31	45	*76*
MRDcalc (with *f* + *h*)	neg	231	0	*231*
	pos	24	45	*69*
total		*255*	*45*	*300*

Figure S2. Constructing the MRD model. (A) Distribution of coverage factors (h) within the study. Coverage factor is calculated as percentage PM covered with LAIP at time of diagnosis. In blue the coverage factor of patients for whom the implementation of the coverage factor did not alter the MRD status (positive or negative). In grey coverage factors of patients (n=24) who were MRD- in the MRD model without variables, but MRD+ when both variables are used (Suppl. Fig. 2C). (B) Scatter plot showing that patients with very low LAIP coverage (< 25%) at diagnosis (shown in white), show least correlation between MRD results in the model without variables (MRDcalc) versus the model when only the coverage factor is implemented (MRDcalc with h). (C) EFS of patients classified MRDpositive or MRDnegative according to 0.1% cut-off, but MRD now calculated according to the formula with incorporation of h. Note that the curves are closer to each other compared to a model with f (Figure 2C). (D) Table showing that 31 patients of the 255 patients who were MRD negative in MRDcalc (no f and h corrections applied) now are MRD positive (second row). The discrepant results originate from different rounding up around 0.1% cut-off for the two approaches. MRDcalc (without f and h incorporated is almost similar to MRDcalc with f included (first row: only 3/258 patients become MRDpositive). (E) Similar to B but now both f and h included. (F) EFS curves of patients classified MRDpositive or MRDnegative according to 0.1% cut-off, but MRD now calculated according to the formula with both f and h incorporated. Table in Suppl. Fig. 2D shows that 24 patients who were MRD negative in MRDcalc (no f and h corrections applied) now are MRD positive (third row).

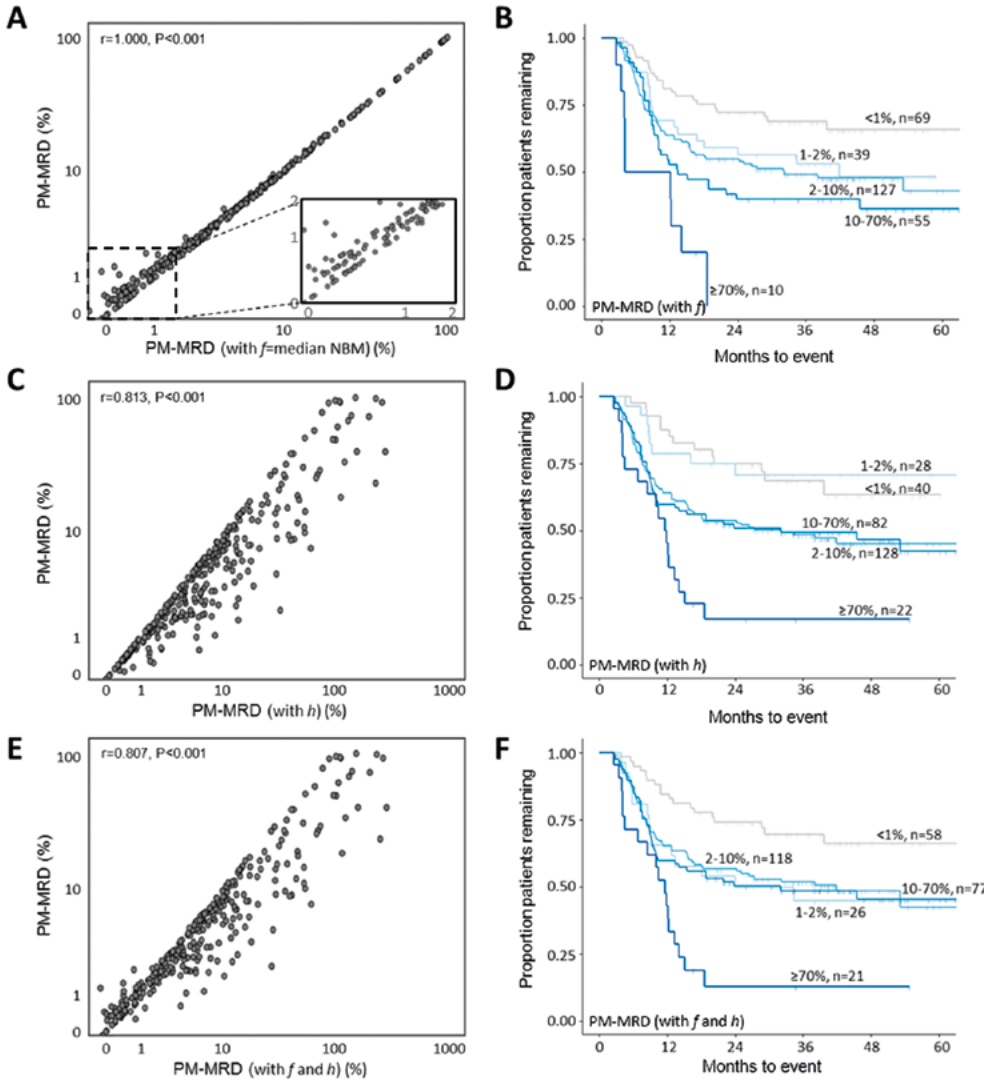


Figure S3. PM-MRD with implementation of one or both variables using multiple cut-offs. The figure summarizes the effect on PM-MRD of specificity f in (A-B)(cut-offs slightly modified compared to earlier shown in Figure 3F), of the coverage factor h in (C-D), and f+h in (E-F). Scatter plots on the left (A,C,E) show the correlation between the model without variables (PM-MRD) and all tested variations. Kaplan-Meier analyses (B,D,F) show EFS for all tested models when using multiple cut-offs. Note that applying h, with or without f (C and E, resp) in general results in higher PM-MRD compared to PM-MRD (without f and h) and also in PM-MRD with f only (A). However, at the lowest PM-MRD levels, incorporation of f results in lower PM-MRD levels (A, see also inset in A), even when f and h are combined (E). This shows that at low MRD levels may have strong impact on PM-MRD.

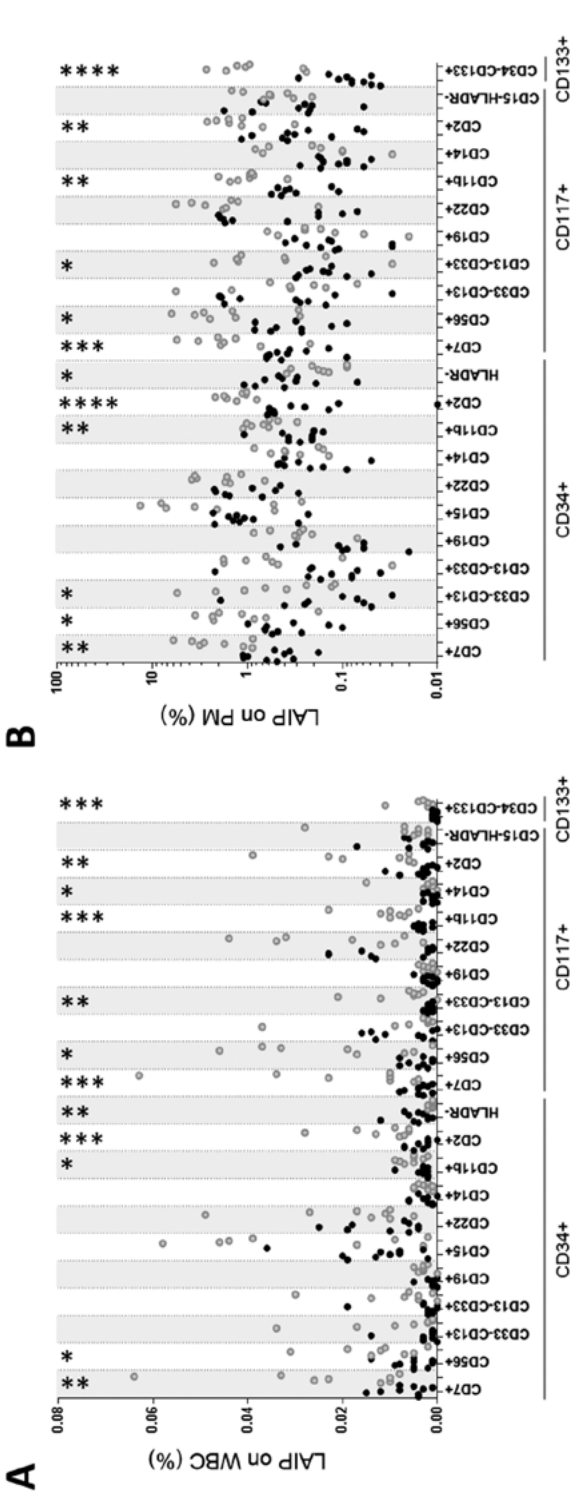


Figure S5. Background LAIP expression on normal BM versus post-chemotherapeutic BM. We evaluated the background expression of all LAIPs using regenerating bone marrow (shown in grey) as (A) percentage within all white blood cells or (B) percentage within the progenitor compartment (PM). At the top , the significance of difference between normal BM and regenerating BM is indicated. * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.

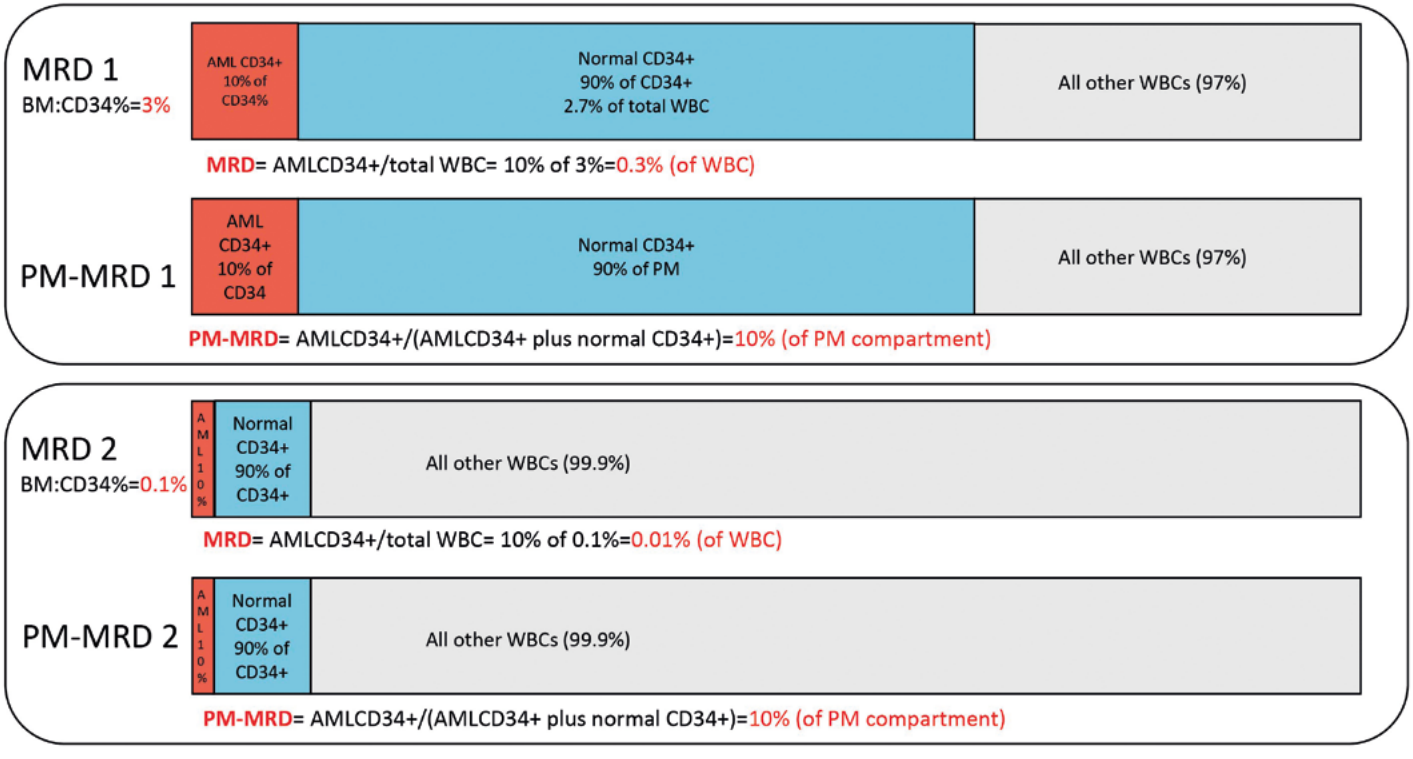


Figure S8. Disturbed relationship between MRD and PM-MRD. PM-MRD and MRD positivity in a schematic example of 2 clinical samples differing in CD34 percentage. In example 1, the AML CD34+ cells make up 10% of the total CD34+ compartment (in brown). Here PM-MRD thus is 10%. Since the BM CD34+ percentage in this example is 3% (of WBC; in brown+blue), the AML part of the total WBC population, representing MRD, is 10% of 3%=0.3% In example 2, PM-MRD is again 10% (in brown), but since the BM CD34+ percentage (in brown+blue) in this example is 0.1% (of WBC), the AML part of the total WBC population (MRD) is 10% of 0.1%=0.01%. These examples at the more extreme conditions of BM CD34% show that samples 1 and 2 have similar PM-MRD (10% of CD34+ compartment) suggesting similar poor prognosis for both (see Figure 3D), while, on the contrary, this sample 1 has a high MRD (0.3%) suggesting very poor prognosis (see Figure 2F), while sample 2 has very low MRD (0.01%) suggesting good prognosis (see Figure 2F)

SUPPLEMENTAL TABLES

Supplemental table 1. Primitive markers and LAIPs in patient population

PM	LAIP	MRD in patients Median (range)	PM-MRD in patients Median (range)	Cases (number)
CD34	CD7+	0.03 (0.00-0.99)	7.14 (0.60-91.71)	47
	CD56+	0.03 (0.00-23.0)	4.50 (0.16-77.6)	30
	CD33-C13+	0.01 (0.00-0.20)	2.60 (0.37-32.75)	26
	CD13-CD33+	0.02 (0.01-0.02)	2.53 (0.93-10.02)	5
	CD13-CD117+	0.19 (0.14-0.24)	17.03 (6.91-27.14)	2
	CD19+	1.00 (0.01-1.98)	50.12 (9.45-90.79)	2
	CD15+	0.09 (0.00-0.28)	15.49 (0.22-61.05)	14
	CD22+	0.01 (0.00-0.92)	10.42 (1.06-31.10)	8
	CD14+	0.02 (0.02-0.02)	3.67 (3.67-3.67)	1
	CD11b+	0.02 (0.00-1.54)	2.96 (0.51-94.42)	25
	CD2+	0.03 (0.01-0.04)	5.36 (0.00-10.71)	2
	HLADR-	0.01 (0.00-0.09)	1.23 (0.14-13.94)	14
CD117	CD7+	0.03 (0.00-5.28)	4.49 (1.02-97.62)	33
	CD56+	0.03 (0.01-1.56)	5.49 (0.38-26.29)	12
	CD33-CD13+	0.01 (0.01-0.01)	0.78 (0.78-0.78)	1
	CD13-CD33+	0.01 (0.00-0.04)	1.80 (0.31-8.62)	6
	CD19+			0
	CD22+	0.02 (0.02-0.02)	5.33 (2.77-7.89)	2
	CD14+			0
	CD11b+	0.01 (0.00-0.12)	2.70 (0.70-6.84)	9
	CD2+	0.00 (0.00-0.00)	0.85 (0.85-0.85)	1
	CD15-HLADR-	0.01 (0.00-27.9)	1.91 (0.28-99.06)	31
CD133	CD34-	0.01 (0.00-0.24)	1.93 (0.00-39.02)	28

FURTHER SUPPLEMENTAL TEXT/FIGURES/TABLES

Further supplemental data can be retrieved via email.

CHAPTER 7

MOLECULAR MINIMAL RESIDUAL DISEASE IN ACUTE MYELOID LEUKEMIA

Diana Hanekamp, Mojca Jongen-Lavrencic, Tim Grob, François G. Kavelaars, Adil al Hinai, Annelieke Zeilemaker, Claudia A.J. Erpelinck-Verschueren, Patrycja L. Gradowska, Rosa Meijer, Jacqueline Cloos, Bart J. Biemond, Carlos Graux, Marinus van Marwijk Kooy, Markus G. Manz, Thomas Pabst, Jakob R. Passweg, Violaine Havelange, Gert J. Ossenkoppele, Mathijs A. Sanders, Gerrit-Jan Schuurhuis, Bob Löwenberg, Peter J.M. Valk

N Engl J Med. 2018 Mar 29;378(13):1189-1199

ABSTRACT

Patients with acute myeloid leukemia (AML) often reach complete remission, but relapse rates remain high. Next-generation sequencing enables the detection of molecular minimal residual disease in virtually every patient, but its clinical value for the prediction of relapse has yet to be established.

METHODS

We conducted a study involving patients 18 to 65 years of age who had newly diagnosed AML. Targeted next-generation sequencing was carried out at diagnosis and after induction therapy (during complete remission). End points were 4-year rates of relapse, relapse-free survival, and overall survival.

RESULTS

At least one mutation was detected in 430 out of 482 patients (89.2%). Mutations persisted in 51.4% of those patients during complete remission and were present at various allele frequencies (range, 0.02 to 47%). The detection of persistent DTA mutations (i.e., mutations in DNMT3A, TET2, and ASXL1), which are often present in persons with age-related clonal hematopoiesis, was not correlated with an increased relapse rate. After the exclusion of persistent DTA mutations, the detection of molecular minimal residual disease was associated with a significantly higher relapse rate than no detection (55.4% vs. 31.9%; hazard ratio, 2.14; P<0.001), as well as with lower rates of relapse-free survival (36.6% vs. 58.1%; hazard ratio for relapse or death, 1.92; P<0.001) and overall survival (41.9% vs. 66.1%; hazard ratio for death, 2.06; P<0.001). Multivariate analysis confirmed that the persistence of non-DTA mutations during complete remission conferred significant independent prognostic value with respect to the rates of relapse (hazard ratio, 1.89; P<0.001), relapse-free survival (hazard ratio for relapse or death, 1.64; P = 0.001), and overall survival (hazard ratio for death, 1.64; P = 0.003). A comparison of sequencing with f low cytometry for the detection of residual disease showed that sequencing had significant additive prognostic value.

CONCLUSIONS

Among patients with AML, the detection of molecular minimal residual disease during complete remission had significant independent prognostic value with respect to re-lapse and survival rates, but the detection of persistent mutations that are associated with clonal hematopoiesis did not have such prognostic value within a 4-year time frame. (Funded by the Queen Wilhelmina Fund Foundation of the Dutch Cancer Society and others.)

INTRODUCTION

Acute myeloid leukemia (AML) is a heterogeneous group of clonal hematopoietic stem cell disorders with variable response to therapy[1-3]. Although the majority of patients with newly diagnosed AML have morphologic complete remission after they are treated with intensive induction chemotherapy, relapse rates remain high[2]. Decisions about the choice of post remission therapy in patients with AML currently depend on the identification of a selected set of genetic markers at diagnosis and the detection of residual disease with multiparameter flow cytometry[2,4]. Quantitative molecular evaluation during complete remission could further improve prognostication of outcomes in patients with AML.

The potential of the detection of molecular minimal residual disease after treatment to predict disease relapse in patients with AML has been explored, but assessment of molecular minimal residual disease is not widely established in clinical practice. Previous studies have dealt with only a few leukemia-specific genetic aberrations[5-11]. Next-generation sequencing enables comprehensive, simultaneous detection of somatic mutations that are often patient-specific, both at diagnosis and during treatment[5,12]. Initial studies showed the complex dynamics of residual mutations after induction therapy and the possible association between the persistence of certain somatic mutations and risk of relapse[12,13].

In determining whether molecular monitoring may be applicable in patients with AML, the phenomenon of age-related clonal hematopoiesis (also known as clonal hematopoiesis of indeterminate potential),[14-17] a condition characterized by the recurrence of gene mutations (allele frequency, >2%) in healthy persons with no evidence of hematologic disease, has added an extra layer of complexity. Persons with age-related clonal hematopoiesis have a slightly increased risk of developing hematologic cancers over time[14,15,18]. Mutations in the epigenetic regulators DNMT3A, TET2, and ASXL1 (i.e., DTA mutations) are most common in persons with age-related clonal hematopoiesis[14-19]. Residual leukemia-specific mutations that are present in the bone marrow during complete remission may represent either residual leukemic cells or age-related clonal hematopoiesis[14,15,17]. Whether posttreatment persistence of genetic mutations associated with age-related clonal hematopoiesis in the bone marrow from patients with AML has an effect on the disease course remains unclear.

We evaluated a large cohort of patients with AML to investigate whether targeted molecular monitoring with next-generation sequencing could add clinical value for predicting the recurrence of leukemia.

METHODS

STUDY DESIGN

The study was designed by the first two and the last two authors, who wrote the manuscript with input from the other authors. The authors vouch for the completeness and accuracy of

the data and analysis. No one who is not an author contributed to the manuscript. There was no commercial support for the study.

PATIENTS AND CELL SAMPLES

Between 2001 and 2013, we obtained samples of bone marrow or peripheral blood from 482 patients, between the ages of 18 and 65, who had a confirmed diagnosis of previously untreated AML (428 patients) or had refractory anemia with excess of blasts, with a score on the Revised International Prognostic Scoring System of more than 4.5, indicating a high or very high risk of relapse (54 patients). To be included in the study, patients had to be in either complete remission or complete remission with incomplete hematologic recovery (defined according to the European Leukemia Net recommendation; hereafter collectively referred to as complete remission), with less than 5% blast cells in the bone marrow[2,4], after receiving two cycles of induction chemotherapy (Fig. S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Among patients in whom at least one mutation was detected at diagnosis, samples were obtained during a defined period of remission, between 21 days and 4 months after the start of the second treatment cycle.

Patients were treated according to the clinical protocol of either the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON)20 or the Swiss Group for Clinical Cancer Research (SAKK). The treatment protocols and patient eligibility criteria have been described previously[21,22]. All the patients provided written informed consent. Details about the patients and cell samples are provided in the Supplementary Appendix.

TARGETED NEXT-GENERATION SEQUENCING AND MULTIPARAMETER FLOW CYTOMETRY

To detect the mutations in 54 genes that are often present in patients with hematologic cancers, we used targeted next-generation sequencing with the Illumina TruSight Myeloid Sequencing Panel (Illumina), following the manufacturer's protocol. Detection of residual disease with multiparameter flow cytometry was performed as described previously[23]. Details about these detection methods and data interpretation are provided in the Supplementary Appendix.

STATISTICAL ANALYSIS

The 430 patients in whom at least one mutation was detected at diagnosis were randomly assigned to either a training cohort (283 patients) or a validation cohort (147 patients); the two cohorts had similar clinical, cytogenetic, and molecular characteristics (Table 1, and Fig. S1 and Table S1 in the Supplementary Appendix). The primary end point was the 4-year cumulative incidence of relapse (defined according to the European Leukemia Net recommendation4), and the secondary end points were the 4-year rates of overall survival and relapse-free survival. Within each cohort, the difference in the incidence of relapse between patients in whom residual disease was detected and those in whom

residual disease was not detected was evaluated with the use of the method of Gray and the Fine and Gray model for competing risks. The log-rank test and the Cox proportional-hazards model were used for survival analyses. A two-sided P value of 0.05 or less was considered to indicate statistical significance. Details about the statistical analyses are provided in the Supplementary Appendix.

RESULTS

DETECTION OF MUTATIONS AT DIAGNOSIS

We performed targeted next-generation sequencing to detect gene mutations at diagnosis in samples obtained from 482 patients with AML (Fig. S1 in the Supplementary Appendix). We detected an average of 2.9 mutations per patient; at least 1 single mutation, which could potentially serve as a marker of residual disease, was present in 430 (89.2%) of the patients. Mutations in NPM1, DNMT3A, FLT3, and NRAS were among the most common detectable mutations at diagnosis (Table 1 and Figure 1A, and Table S1 in the Supplementary Appendix).

DETECTION OF MUTATIONS DURING COMPLETE REMISSION

We then performed targeted next-generation sequencing to detect persistent mutations after induction therapy in samples of bone marrow obtained from 430 patients who were in complete remission. Persistent mutations were detected in 51.4% of the patients (Figure 1A, and Fig. S2A in the Supplementary Appendix). The rate at which mutations persisted was highly variable across genes. DTA mutations were most common, persisting at rates of 78.7% for DNMT3A, 54.2% for TET2, and 51.6% for ASXL1 (Figure 1A). In contrast, the majority of mutations in genes related to the RAS pathway were cleared after induction therapy, with mutations in NRAS, PTPN11, KIT, and KRAS persisting at rates of 4.2%, 7.0%, 13.5%, and 12.5%, respectively.

Of note, the allele frequencies of the mutations that persisted during complete remission ranged from 0.02 to 47% (Figure 1B). This finding suggests that residual mutation-bearing cells could constitute a minor population of the cells or perhaps even a majority of the cells. An allele frequency of 50% is consistent with the presence of a heterozygous mutation in all cells. Thus, although the patients were in morphologic complete remission, which would typically imply that heterozygous mutations are present at allele frequencies lower than 2.5% (the equivalent of <5% blast cells in the bone marrow), the samples that were obtained during remission often contained mutations with much higher allele frequencies (Figure 1B).

Mutations that persisted after induction therapy at allele frequencies higher than 2.5% were often DTA mutations (Figure 1, and Fig. S2 and S3 in the Supplementary Appendix). In contrast, mutations in IDH1, IDH2, STAG2, TP53, and other genes only occasionally persisted after induction therapy at allele frequencies higher than 2.5%, and thus the allele

frequencies of these mutations were typically consistent with the state of morphologic complete remission (<5% blast cells in the bone marrow).

Table 1. Clinical, Cytogenetic, and Molecular Characteristics of the 430 Patients°

Characteristic	value
Age at diagnosis - yr	
Median	51
Range	18-66
Sex - no. (%)	
M	216 (50)
F	214 (50)
White-cell count per microliter at diagnosis - no. (%)	
≤ 100,000	387 (90)
>100,000	43 (10)
2017 European Leukemia Network risk classification at diagnosis - no. (%)	
Favorable	204 (48)
Intermediate	113 (26)
Adverse	113 (26)
No. of chemotherapy cycles to attain complete remission - no. (%)	
1	360 (84)
2	70 (16)
Consolidation therapy - no. (%)	
None	46 (11)
Chemotherapy	117 (27)
Autologous hematopoietic stem-cell transplantation	78 (18)
Allogeneic hematopoietic stem-cell transplantation	189 (44)
Cytogenetic analysis at diagnosis - no. (%)†	
t(8;21)	27 (6)
inv(16)	24 (6)
Complex karyotype	38 (9)
Monosomal karyotype	30 (7)
Mutation at diagnosis - no. (%)	
ASXL1	31 (7)
*CEBPA*double mutation	19 (4)
DNMT3A	141 (33)
FLT3	
Tyrosine kinase domain	53 (12)
Internal tandem duplication, low ratio	40 (9)
Internal tandem duplication, high ratio	51 (12)
NPM1	168 (39)
RUNX1	50 (12)
TET2	48 (11)

° The percentages may not sum up to 100 because of rounding.

† Karyotyping failed in 13 patients.

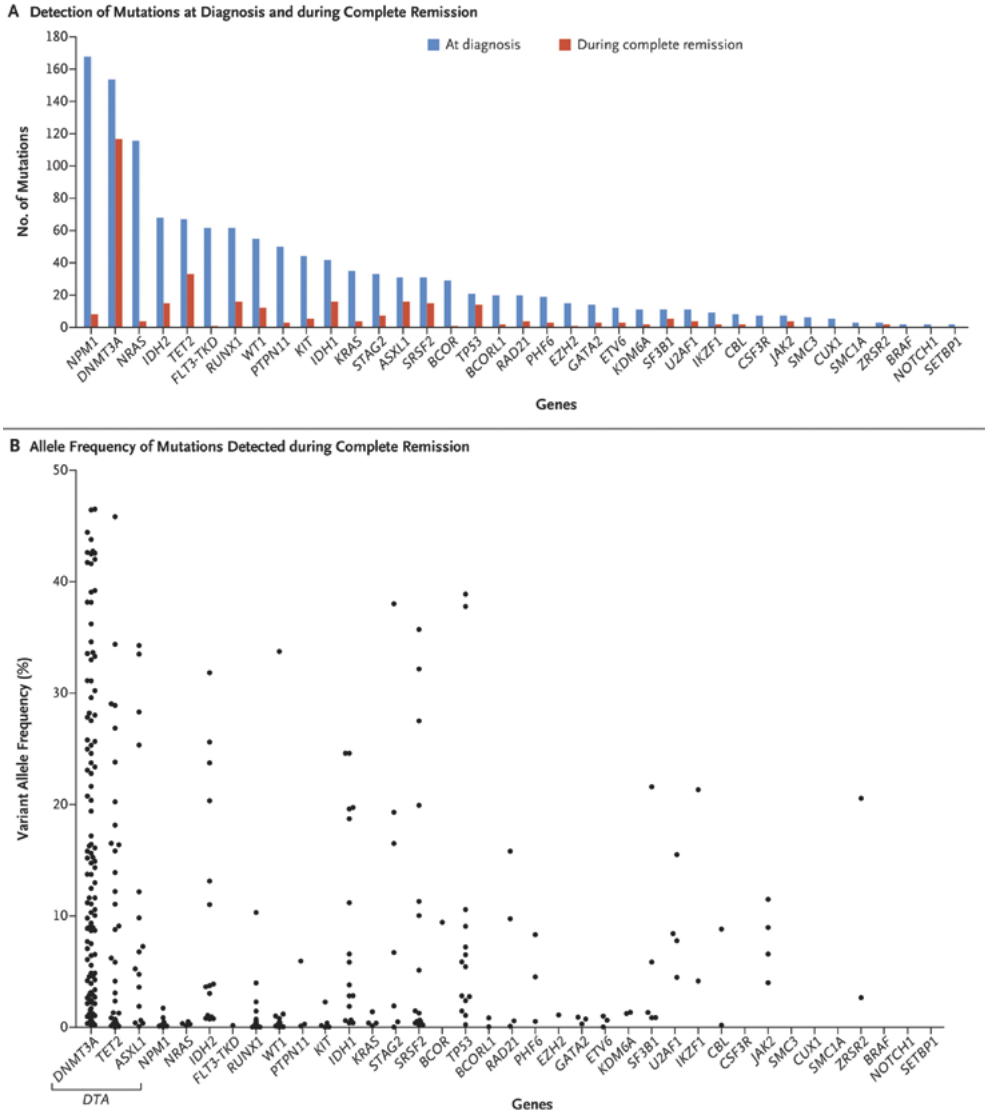


Figure 1. Detection of Mutations at Diagnosis and during Complete Remission and Allele Frequency of Mutations Detected during Complete Remission. Panel A shows the number of mutations in each leukemia-associated gene, both at diagnosis of acute myeloid leukemia and during complete remission, in 430 patients. Panel B shows the allele frequency of each mutation in each gene during complete remission in 430 patients. In male patients, the variant allele frequencies for PHF6, KDM6A, ZRSR2, BCOR, BCORL1, and STAG2 (on the X chromo-some) were divided by 2.

Because DTA mutations have been established as the most common gene mutations in persons with age-related clonal hematopoiesis[14-19], the persistent DTA mutations might have represented nonleukemic clones that repopulated the bone marrow after induction

therapy. Among patients who had both DTA mutations and non-DTA mutations at diagnosis, non-DTA mutations were generally cleared after induction chemotherapy, whereas DTA mutations often remained detectable during complete remission and were the only persistent mutations in 90 of 133 (67.7%) of those patients (Fig. S2 in the Supplementary Appendix). These observations are consistent with the notion that residual cells bearing DTA mutations after induction therapy represent nonleukemic clones rather than persistent malignant disease.

RELAPSE AND SURVIVAL END POINTS

In the training cohort (283 patients), we found that the detection of any persistent mutation during complete remission was associated with an increased risk of relapse (4-year relapse rate, 48.2% with detection vs. 32.4% with no detection; P=0.03) (Fig. S4A in the Supplementary Appendix). We then imposed various thresholds for allele frequency to determine whether the prognostic value of the persistent mutations would improve after the exclusion of mutations with a high allele frequency, which could indicate a state of clonal hematopoiesis. The correlation of persistent mutations with an increased relapse risk appeared to be independent of allele frequency. A correlation with relapse risk generally remained present when we excluded persistent mutations with allele frequencies at or above the following thresholds: 30% (P=0.09), 20% (P=0.11), 10% (P=0.01), 5% (P=0.04), 2.5% (P=0.007), and 1% (P=0.07) (Fig. S4 in the Supplementary Appendix). The exclusion of persistent mutations with certain allele frequencies had no clear effect on the relationship between persistent mutations and an increased relapse risk, thus precluding the identification of a threshold for allele frequency that could be used to distinguish populations at higher or lower risk for relapse. As we mentioned previously, the patients with persistent mutations at high allele frequencies were enriched for DTA mutations (Figure 1B).

We next determined whether persistent DTA mutations, which are associated with age-related clonal hematopoiesis, might be correlated with an increased relapse risk. We observed that the detection of persistent DTA mutations was not significantly associated with a higher 4-year relapse rate than no detection (P=0.29). The absence of a correlation was independent of allele frequency. No significant correlation of persistent DTA mutations with an increased relapse risk was apparent when we excluded persistent DTA mutations with allele frequencies at or above the following thresholds: 30% (P=0.91), 20% (P=0.66), 10% (P=0.89), 5% (P=0.82), 2.5% (P=0.53), and 1% (P=0.92) (Fig. S5 in the Supplementary Appendix). In contrast, among patients who had persistent DTA mutations during complete remission, coexisting persistent non-DTA mutations had high prognostic value with respect to relapse (4-year relapse rate, 66.7% with detection vs. 39.4% with no detection; P=0.002) (Figure 2A). Thus, in patients with persistent DTA mutations, the presence of residual disease that specifically included coexisting non-DTA mutations represented a predictor of impending relapse.

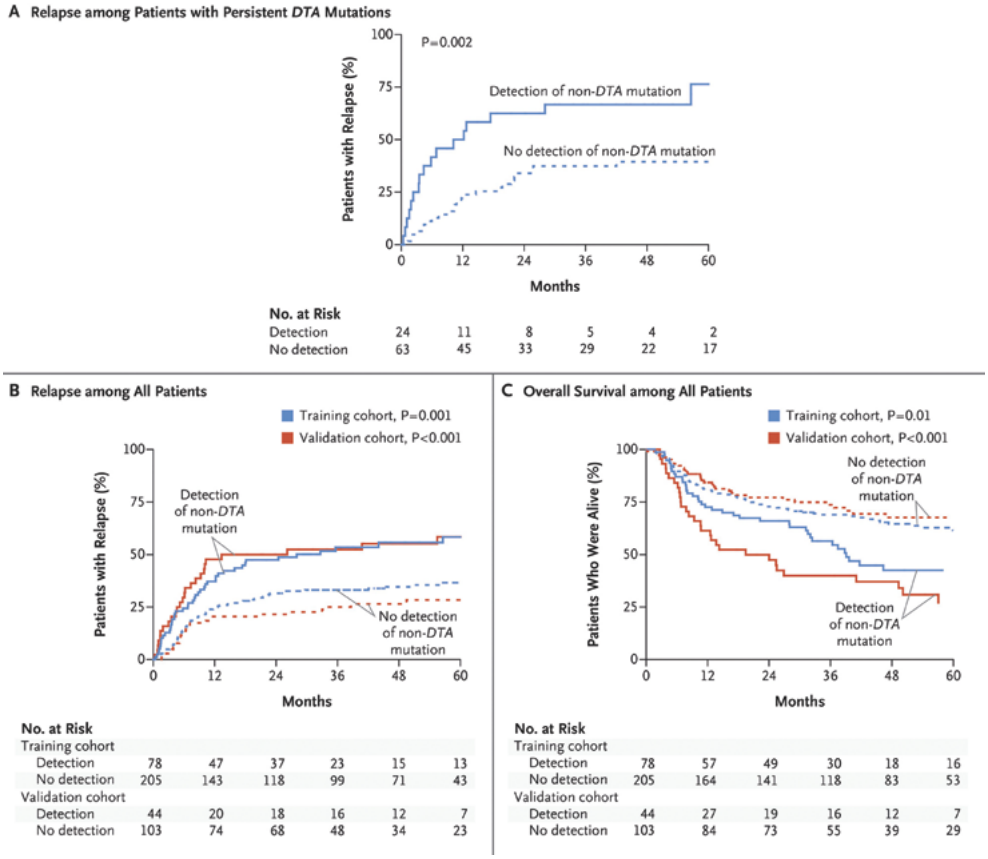


Figure 2. Rates of Relapse and Overall Survival. Shown is the cumulative incidence of relapse among patients in the training cohort with persistent DTA mutations, according to the detection of coexisting persistent non-DTA mutations during complete remission (Panel A), and among all patients in the training and validation cohorts, according to the detection of any persistent non-DTA mutations during complete remission (Panel B). Panel C shows the rate of overall survival among all patients in the training and validation cohorts, according to the detection of any persistent non-DTA mutations during complete remission. DTA mutations are mutations in DNMT3A, TET2, and ASXL1

We next assessed whether persistent non-DTA mutations might be correlated with an increased relapse risk. The detection of persistent non-DTA mutations at any allele frequency was strongly associated with an increased relapse risk (4-year relapse rate, 55.7% with detection vs. 34.6% with no detection; P=0.001) (Figure 2B), as well as with reduced relapse-free survival (4-year rate of relapse-free survival, 56.7% with detection vs. 36.6% with no detection; P=0.006) and reduced overall survival (4-year rate of overall survival, 65.3% with detection vs. 43.7% with no detection; P=0.01) (Figure 2C, and Fig. S6 in the Supplementary Appendix).

To assess the reproducibility of these results, we evaluated the effect of sequencing-based detection of persistent non-DTA mutations during complete remission on the rates

of relapse, relapse-free survival, and overall survival in the validation cohort (147 patients). The rates with detection versus no detection were as follows: 4-year relapse rate, 55.1% versus 26.5% (P<0.001); 4-year rate of relapse-free survival, 60.6% versus 35.6% (P<0.001); and 4-year rate of overall survival, 67.6% versus 37.1% (P<0.001) (Figure 2B and 2C, and Fig. S6 in the Supplementary Appendix). The results in the validation cohort confirmed the significant findings in the training cohort.

In the combined training and validation cohorts (a total of 430 patients), persistent non-DTA mutations were detected during complete remission in 28.4% of the patients. Detection of these mutations was associated with a significantly higher 4-year relapse rate than no detection (55.4% vs. 31.9%; hazard ratio, 2.14; 95% confidence interval [CI], 1.57 to 2.91; P<0.001), as well as with lower 4-year rates of relapse-free survival (36.6% vs. 58.1%; hazard ratio for relapse or death, 1.92; 95% CI, 1.46 to 2.54; P<0.001) and overall survival (41.9% vs. 66.1%; hazard ratio for death, 2.06; 95% CI, 1.52 to 2.79; P<0.001) (Fig. S6 in the Supplementary Appendix).

MULTIVARIATE AND SENSITIVITY ANALYSES

We performed multivariate analyses that accounted for the major established relevant prognostic factors, including age, white-cell count, 2017 European Leukemia Network risk classification, and the number of cycles of induction chemotherapy needed to attain complete remission. Sequencing-based detection of non-DTA mutations maintained significant independent prognostic value with respect to the rates of relapse (hazard ratio, 1.89; 95% CI, 1.34 to 2.65; P<0.001), relapse-free survival (hazard ratio for relapse or death, 1.64; 95% CI, 1.22 to 2.20; P=0.001), and overall survival (hazard ratio for death, 1.64; 95% CI, 1.18 to 2.27; P=0.003) (Table 2). No significant interactions were apparent between the detection of residual disease and the other prognostic factors in the multivariate model, type of consolidation therapy, or disease entity (AML vs. refractory anemia with excess of blasts) (data not shown).

In sensitivity analyses involving correction for variation in the time at which bone marrow specimens were obtained for sequencing analysis (within the remission period of 21 days to 4 months after the second treatment cycle), the prognostic value of sequencing-based detection of non-DTA mutations with respect to the rates of relapse, relapse-free survival, and overall survival remained unaffected (Table S2 in the Supplementary Appendix). In addition, an analysis that included post remission treatment with allogeneic stem-cell transplantation as a time-dependent variable conferred no effect on the prognostic value of the detection of residual disease (Table S3 in the Supplementary Appendix).

DETECTION OF RESIDUAL DISEASE WITH MULTIPARAMETER FLOW CYTOMETRY

Multiparameter flow cytometry is an increasingly used method for predicting relapse in patients with AML who are in complete remission[7,24]. We compared next-generation

Table 2. Multivariate Analysis of Prognostic Factors for Relapse, Relapse-free Survival, and Overall Survival.

	Relapse		Relapse-free Survival		Overall Survival	
Prognostic Factor	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
Molecular minimal residual disease: detection vs. no detection	1.89 (1.34-2.65)	<0.001	1.64 (1.22-2.20)	0.001	1.64 (1.18-2.27)	0.003
Age: per year	1.01 (0.99-1.03)	0.212	1.02 (1.00-1.03)	0.009	1.03 (1.01-1.04)	0.001
White-cell count per microliter at diagnosis: >100,000 vs. ≤100,000	2.16 (1.31-3.56)	0.003	2.03 (1.34-3.08)	0.001	2.02 (1.27-3.21)	0.003
2017 European Leukemia Network risk classification						
Intermediate vs. Favorable	1.67 (1.12-2.49)	0.012	2.01 (1.42-2.83)	<0.001	2.53 (1.72-3.72)	<0.001
Adverse vs. Favorable	1.83 (1.26-2.66)	0.002	2.21 (1.58-3.10)	<0.001	2.67 (1.83-3.92)	<0.001
Number of chemotherapy cycles to attain complete remission:						
2 cycles vs. 1 cycle	2.17 (1.50-3.15)	<0.001	2.43 (1.74-3.39)	<0.001	2.96 (2.09-4.21)	<0.001

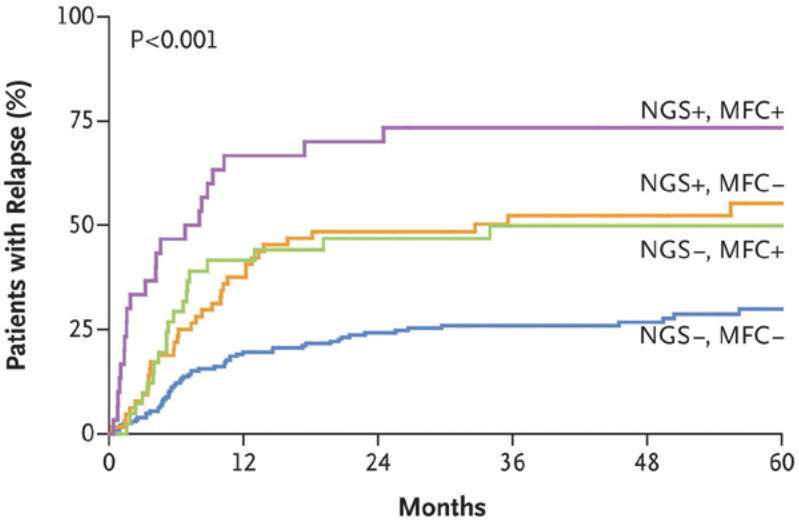
sequencing for the detection of persistent non-DTA mutations with flow cytometry for the detection of residual disease in a representative subgroup of 340 patients, from whom sufficient samples were obtained for both analyses. Concordant results (either detection or no detection on both assays) were found in 69.1% of the patients (30 patients with detection and 205 with no detection), whereas persistent non-DTA mutations were detected only on sequencing in 64 patients and only on flow cytometry in 41 patients. The 4-year relapse rate was 73.3% among patients in whom both assays were positive, 52.3% among those who had residual disease on sequencing but not on flow cytometry, 49.8% among those who had residual disease on flow cytometry but not on sequencing, and 26.7% among those in whom both assays were negative (Figure 3). In a multivariate analysis that combined the results of sequencing and flow cytometry, the combined use of the two assays for the detection of residual disease conferred independent prognostic value with respect to the rates of relapse ($P<0.001$), relapse-free ($P<0.001$), and overall survival ($P=0.003$) (Table S4 in the Supplementary Appendix).

DISCUSSION

In addition to the presence of genetic abnormalities at diagnosis, the continued presence of particular gene mutations during or after treatment carries prognostic information for certain genetically defined AML subtypes[5-11]. This applies, for example, to AML associated with a mutation in NPM1, for which the detection of a residual mutation in NPM1 transcripts during complete remission is indicative of an increased probability of relapse[8,9]. However, this example is only representative of a single-gene approach. We report the results of a systematic study that involved a large number of patients with AML, in which we used a genomewide approach to evaluate the persistence of multiple gene mutations for the effect on treatment outcomes. Patients were treated with intensive chemotherapy regimens and attained morphologic complete remission, with a median follow-up exceeding 3 years.

Of note, age-related clonal hematopoiesis[14-17], which is characterized by recurrent somatic mutations in leukemia-associated genes in persons with no apparent hematologic disease, adds a challenge in the detection of residual disease. Our study showed that the persistence of mutations that are most commonly associated with age-related clonal hematopoiesis (i.e., DTA mutations [mutations in DNMT3A, TET2, and ASXL1]) during complete remission did not contribute to a measurably increased risk of relapse within a follow-up period of 4 years in adults with AML who were younger than 65 years of age. This appeared to be true for mutations that were present at various allele frequencies, which suggests that the clone size in age-related clonal hematopoiesis yields no prognostic value with respect to the end points defined in this study.

The cells bearing DTA mutations appeared to persist and possess a selective clonal advantage over normal stem cells when they repopulated the bone marrow after induction therapy. This finding is consistent with the competitive clonal advantage of hematopoietic



No. at Risk						
NGS+, MFC+	30	8	7	5	4	4
NGS–, MFC+	41	22	18	14	11	7
NGS+, MFC–	64	39	30	22	15	11
NGS–, MFC–	205	153	130	101	69	42

Figure 3. Rate of Relapse According to Results of Next-Generation Sequencing and Multiparameter Flow Cytometry. Shown is the cumulative incidence of relapse, according to the presence of positive (+) or negative (–) results for the detection of persistent non-DTA mutations during complete remission on next-generation sequencing (NGS) and on multiparameter flow cytometry (MFC)

stem cells with deficiencies and mutations in DNMT3A and TET2, an advantage that has been reported previously[25-27]. The proliferative advantage of hematopoietic stem cells with DTA mutations and their capacity to withstand chemotherapy because of inherent resistance may explain why persistent premalignant DTA mutations were not correlated with an increased probability of relapse and thereby did not constitute a reliable molecular biomarker for the assessment of relapse risk.

It is possible that gene mutations other than DTA mutations also partially reflect clonal hematopoiesis. However, at this time, we cannot rigorously verify the possibility that gene mutations associated with age-related clonal hematopoiesis also reside as subfractions among the other gene abnormalities in leukemia cells. In addition, mutations in TP53, IDH1, and IDH2, along with genes related to the RAS pathway and spliceosome genes, have been shown to have distinct biologic features in the context of AML pathogenesis[28-31]. Therefore, in this study, we collectively considered non-DTA mutations to be abnormalities that are unrelated to clonal hematopoiesis.

Our study had a median follow-up of almost 40 months. Among patients with AML who have complete remission, most relapses generally occur within the first 4 years. We found that the continued persistence of DTA mutations was not associated with an increased

relapse risk, and thus these residual cells may not need to be eliminated to prevent relapse. However, the limited follow-up of 40 months does not rule out the possibility that persistent DTA mutations represent an increased risk of relapse at a later time point.

Although sequencing-based detection enables assessment for residual disease in virtually all patients with AML, it is imperfect in two ways. First, not all patients with residual mutation-bearing cells have a relapse. Second, some patients with no measurable residual disease have a relapse. It is conceivable that relapse estimation can be improved with the development of technological variations of sequencing-based approaches that have greater sensitivity or a broader scope (e.g., those with molecular barcoding, exome sequencing, or whole-genome sequencing) or with the identification of additional molecular and phenotypic markers so that quantitative minor clones or subclones associated with the leukemia are captured by the assay. In this respect, it is of particular interest that the use of multiparameter flow cytometry[7,24] — which identifies patients with AML who have an increased risk of relapse according to an entirely different approach that is based on a residual leukemia-associated immunophenotype[23,32] — can increase the yield of identification of residual leukemia during complete remission.

In this study, gene sequencing and multiparameter flow cytometry each had independent and additive prognostic value with respect to rates of relapse and survival in patients with AML. The detection of residual leukemia with both methods is associated with an excessively high probability of relapse (approximately 75%), and the absence of detection of residual disease with both methods is correlated with a relatively low probability of relapse (approximately 25%). Thus, the combined use of sequencing and flow cytometry during complete remission warrants further development and evaluation in clinical practice.

In conclusion, targeted sequencing-based detection of molecular minimal residual disease during complete remission was associated with an increased risk of relapse or death in patients with AML. However, over a 4-year follow-up period, the risk of relapse or death was not influenced by the persistence of genetic lesions that are associated with age-related clonal hematopoiesis.

ACKNOWLEDGEMENTS

We thank all the participating centers of the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK), where the clinical trials that formed the basis for this study were conducted; H. Berna Beverloo for performing cytogenetic analyses, Eric Bindels for performing next-generation sequencing, Remco Hoogenboezem for assisting with bioinformatics, and Egied Simons for assisting with preparation of the figures (Erasmus University Medical Center); and Vincent J. van der Velden (Erasmus University Medical Center), Jennichjen Slomp (Medisch Spectrum Twente, Medlon), and Frank Preijers (Radboud University Medical Center) for performing flow cytometry.

REFERENCES

1. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059-74.

2. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med 2015;373:1136-52.

3. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 2016;374:2209-21.

4. Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129:424-47.

5. Grimwade D, Freeman SD. Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"? Blood 2014;124:3345-55.

6. Kayser S, Walter RB, Stock W, Schlenk RF. Minimal residual disease in acute myeloid leukemia--current status and future perspectives. Curr Hematol Malig Rep 2015;10:132-44.

7. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. Leukemia 2017;31:1482-90.

8. Ivey A, Hills RK, Simpson MA, et al. Assessment of Minimal Residual Disease in Standard-Risk AML. N Engl J Med 2016;374:422-33.

9. Kronke J, Schlenk RF, Jensen KO, et al. Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group. J Clin Oncol 2011;29:2709-16.

10. Gaidzik VI, Weber D, Paschka P, et al. DNMT3A mutant transcript levels persist in remission and do not predict outcome in patients with acute myeloid leukemia. Leukemia 2017.

11. Bhatnagar B, Eisfeld AK, Nicolet D, et al. Persistence of DNMT3A R882 mutations during remission does not adversely affect outcomes of patients with acute myeloid leukaemia. Br J Haematol 2016;175:226-36.

12. Pastore F, Levine RL. Next-Generation Sequencing and Detection of Minimal Residual Disease in Acute Myeloid Leukemia: Ready for Clinical Practice? Jama 2015;314:778-80.

13. Klco JM, Miller CA, Griffith M, et al. Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. Jama 2015;314:811-22.

14. Genovese G, Kahler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 2014;371:2477-87.

15. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014;371:2488-98.

16. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood 2017;130:742-52.

17. Shlush LI. Age related clonal hematopoiesis (ARCH). Blood 2017.

18. Jan M, Ebert BL, Jaiswal S. Clonal hematopoiesis. Semin Hematol 2017;54:43-50.

19. Buscarlet M, Provost S, Zada YF, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood 2017;130:753-62.

20. Dutch–Belgian Cooperative Trial Group for Hematology–Oncology. Home page (http://www.hovon.nl)

21. Pabst T, Vellenga E, van Putten W, et al. Favorable effect of priming with granulocyte colony-stimulating factor in remission induction of acute myeloid leukemia restricted to dose escalation of cytarabine. Blood 2012;119:5367-73.

22. Lowenberg B, Pabst T, Maertens J, et al. Therapeutic value of clofarabine in younger and middle-aged (18-65 years) adults with newly diagnosed AML. Blood 2017;129:1636-45.

23. Terwijn M, van Putten WL, Kelder A, et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. J Clin Oncol 2013;31:3889-3897.

24. Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: consensus document from ELN MRD Working Party. Blood 2018 January 12 (Epub ahead of print).

25. Challen GA, Sun D, Jeong M, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet 2011;44:23-31.

26. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 2011;20:11-24.

27. Shlush LI, Zandi S, Mitchell A, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature 2014;506:328-33.

28. McKerrell T, Park N, Moreno T, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep 2015;10:1239-1245

29. Wong TNRG, Ramsingh G, Young AL, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature 2015;518:552-555.

30. Lindsley RCSW, Saber W, Mar BG, et al. Prognostic mutations in myelodysplastic syndrome after stem-cell transplantation. N Engl J Med 2017;376:536-547.

31. Desai P, Mencia-Trinchant N, Savenkov O, et al. Somatic mutations predict acute myeloid leukemia years before diagnosis. bioRxiv 2017 (https://www.biorxiv.org/content/early/2017/12/21/237941. opens in new tab).

32. Ravandi F, Jorgensen J, Borthakur G, et al. Persistence of minimal residual disease assessed by multiparameter flow cytometry is highly prognostic in younger patients with acute myeloid leukemia. Cancer 2017;123:426-35.

33. Valk PJ, Verhaak RG, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350:1617-28.

34. Taskesen E, Bullinger L, Corbacioglu A et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutation in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. Blood 2011;117:2469-75.

35. BBmap short-read aligner, and other bioinformatics tools. 2016. At http://sourceforge.net/projects/bbmap/.)

36. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213-9.

37. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009:25:2078-9.

38. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.

39. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568-76.

40. Indelocator. at http://www.broadinstitute.org/cancer/cga/indelocator.)

41. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association 1999;94:496- 509.

42. Mantel N. BD. Evaluation of response-time data involving transient states: an illustration using heart-transplant data. . J Am Stat Assoc 1974;69:81-6.

SUPPLEMENTARY DATA

PATIENTS AND CELL SAMPLES

Bone marrow aspirations or peripheral blood samples at diagnosis were taken after informed consent. Follow-up bone marrow samples of 430 out of 482 AML or RAEB patients in CR with mutations at diagnosis (Figure S1) were taken at least 21 days after the start of the second induction cycle. If additional samples were available, the most recent sample prior to start of consolidation therapy was selected. In case no consolidation therapy was given, the last sample that was available within a four-month interval from start of the second induction cycle was selected. The probabilities of relapse in AML patients with or without available samples did not differ (p=0.281). The median follow-up of the 430 AML cases was 39.7 months and the residual disease status was not available to the clinical investigator and did not influence the choice of consolidation therapy. Blasts and mononuclear cells at diagnosis were purified by Ficoll-Hypaque (Nygaard, Oslo, Norway) density gradient centrifugation and cryopreserved. Of all 430 AML cases reaching CR, white blood cells were isolated after induction treatment in 385 cases and mononuclear cells were subsequently purified in 45 cases. After thawing, cells were lysed in RLT solution with the addition of DTT (Qiagen, Venlo, The Netherlands).

DNA AND RNA ISOLATION

High quality DNA was extracted using the QIAsymphony (Qiagen, Venlo, The Netherlands). DNA concentration was measured by Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Wilmington, DE). RNA was isolated with RNA-Bee following the protocols of the manufacturer (Bio-Connect BV, Huissen, The Netherlands). CBFB-MYH11, RUNX1-RUNX1T1, FLT3 internal tandem duplication (ITD) and CEBPA mutations were determined as described previously[33,34]. TARGETED NGS The NGS libraries were paired-end sequenced (2x221bp) on an Illumina HiSeq 2500 System (Illumina, San Diego, CA) in Rapid Run mode. Since CBF fusion transcripts, CEBPA mutations and FLT3 ITDs cannot be reliably assessed with NGS on DNA, these molecular aberrations were excluded from the analyses.

NGS DATA ANALYSIS

The vast majority of amplicon target regions were completely paired-end sequenced. Overlap-based error-correction was utilized to attenuate any form of strand-specific error biases. Error-corrected paired-end reads aligned to the human genome version 19 (hg19) with BBMAP[35] followed by quality control to determine cases with insufficient number of reads for adequate variant calling. Single nucleotide variants (SNVs) and insertions-deletions (indels) at diagnosis were determined by MuTect[36], Samtools[37], GATK[38], Varscan[39], Indelocator[40] and Pindel[38]. Variant allele frequencies (VAF) of mutations detected at diagnosis were calculated as the ratio between the number of mutant and total reads. The persistence of mutations at follow-up, previously detected at diagnosis,

requires the detection of mutations at exceptionally low VAFs. A follow-up background error model was determined by calculating the VAF for each potential SNV within the set of target genes across all follow-up samples.

The detection of variants at low detection level is primarily reserved for highly discriminative insertion or deletion mutations sequenced at sufficient depth in both the follow-up samples of interest and the control set of remission samples. The strength of a site-specific error model is that it models the unique site-and-variant specific noise profile based on a large set of remission samples from patients who did not carry that specific mutation at diagnosis and thereby enables the assessment whether the variant remains persistent, defined as a statistical outlier, in the follow-up sample from the patient of interest. Since more complex insertion or deletion mutations are particularly distinct and the odds of detecting such variants as a consequence of sequencing or alignment errors is exceptionally low, such variants can be detected at higher sensitivity. The detection sensitivity of other mutations is variable and highly dependent on the average coverage for that specific locus for all samples, the observed error variance of the site-specific variant in the control set (a high variance results in decreased detection sensitivity) and the number of control sample available. The unique combination of patient-specific mutations, the application of a site-specific error model and strict detection criteria minimizes the odds that variants are erroneously called to persist. Quantile normalization of the calculated VAFs was performed per flow cell to mitigate the effect of qualitative differences amongst samples. All SNVs detected across the diagnostic samples were compiled and the background VAF distribution was determined for each individual SNV from follow-up samples lacking this SNV in the matched diagnostic sample. For the remaining follow-up samples the persistence of the SNV was considered confirmed when the VAF was an outlier compared to the background VAF distribution according to the Thompson-Tau test. A one-sided p-value <0.01 was considered statistically significant. Indels were processed and compared similarly, except for quantile normalization as there are infinitely many possible indel-configurations per locus.

STATISTICAL ANALYSES

The complete cohort of 430 AML patients was randomly split using Stata into a training (n=283) and validation cohort (n=147) (Table 1, Figure S1 and Table S1). Each patient received a pseudorandom number from a uniform distribution from 0 to 1. The random numbers generated were shuffled by sorting, allowing for random allocation of patients to the training or validation set. Differences in clinical, cytogenetic and molecular characteristics of the training and validation cohorts or NGS and flow cohorts were tested using the Fisher's exact test for categorical variables and Mann-Whitney U test for continuous variables. Clinical, cytogenetic and molecular characteristics of the training and validation cohorts were not significantly different (Table S1). The primary endpoint of the study was the cumulative incidence of relapse (CIR). Competing-risks regression

analysis was performed for relapse with adjustment for non-relapse mortality according to the method of Gray and the Fine & Gray model[41]. The secondary endpoints were relapse free (RFS) and overall survival (OS) which were analyzed using the log-rank test and the Cox proportional hazards model. Relapse and survival time was calculated from the sampling date until the date of the event of interest or censoring. RFS was defined from date of sampling to death, relapse or censoring, whichever came first. All statistical tests were two-sided and p-values <0.05 were considered statically. The proportional hazards assumption was tested by interaction with time and the interactions were evaluated in a standard way. The effect of allogeneic stem cell transplantation on CIR and OS was investigated in both multivariable models as a time-dependent covariate[42]. All p-values are two sided and p-values <0.05 were considered statically significant. Statistical analyses were performed with Stata Statistical Software, Release 14.1 (Stata, College Station, TX).

SUPPLEMENTAL FIGURES

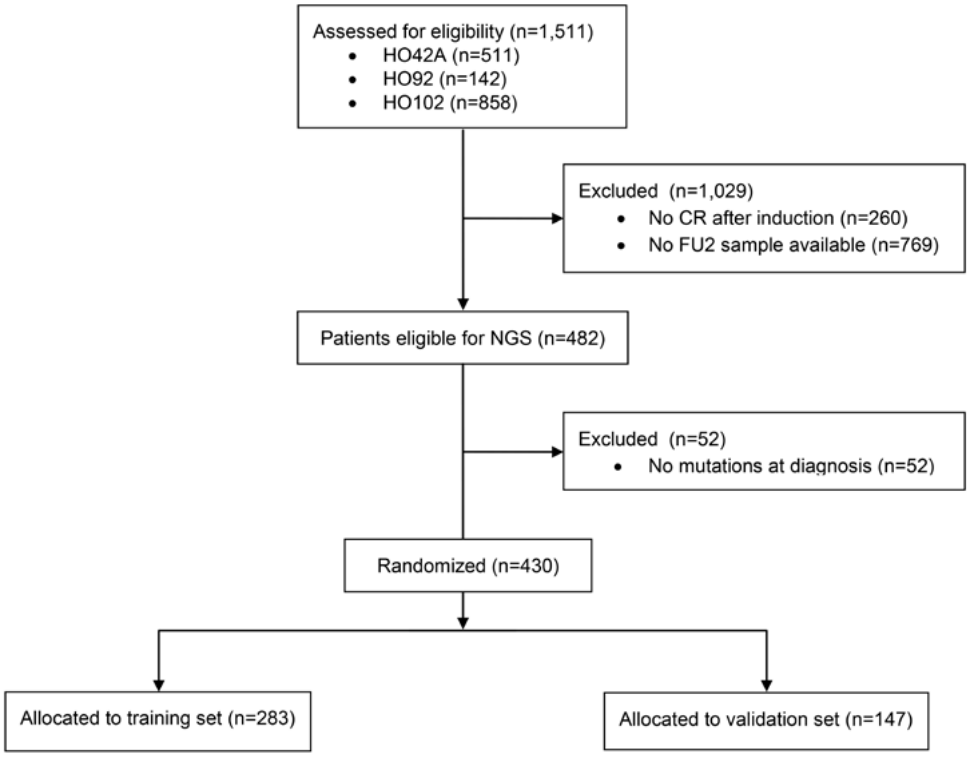


Figure S1. Consort diagram molecular residual disease study. Abbreviations: HO, HOVON-SAKK, Dutch-Belgian Hemato-Oncology Cooperative Group and the Swiss Group for Clinical Cancer Research; CR, Complete morphological Remission; FU2, Follow-up after induction cycle II; NGS, Next Generation Sequencing.

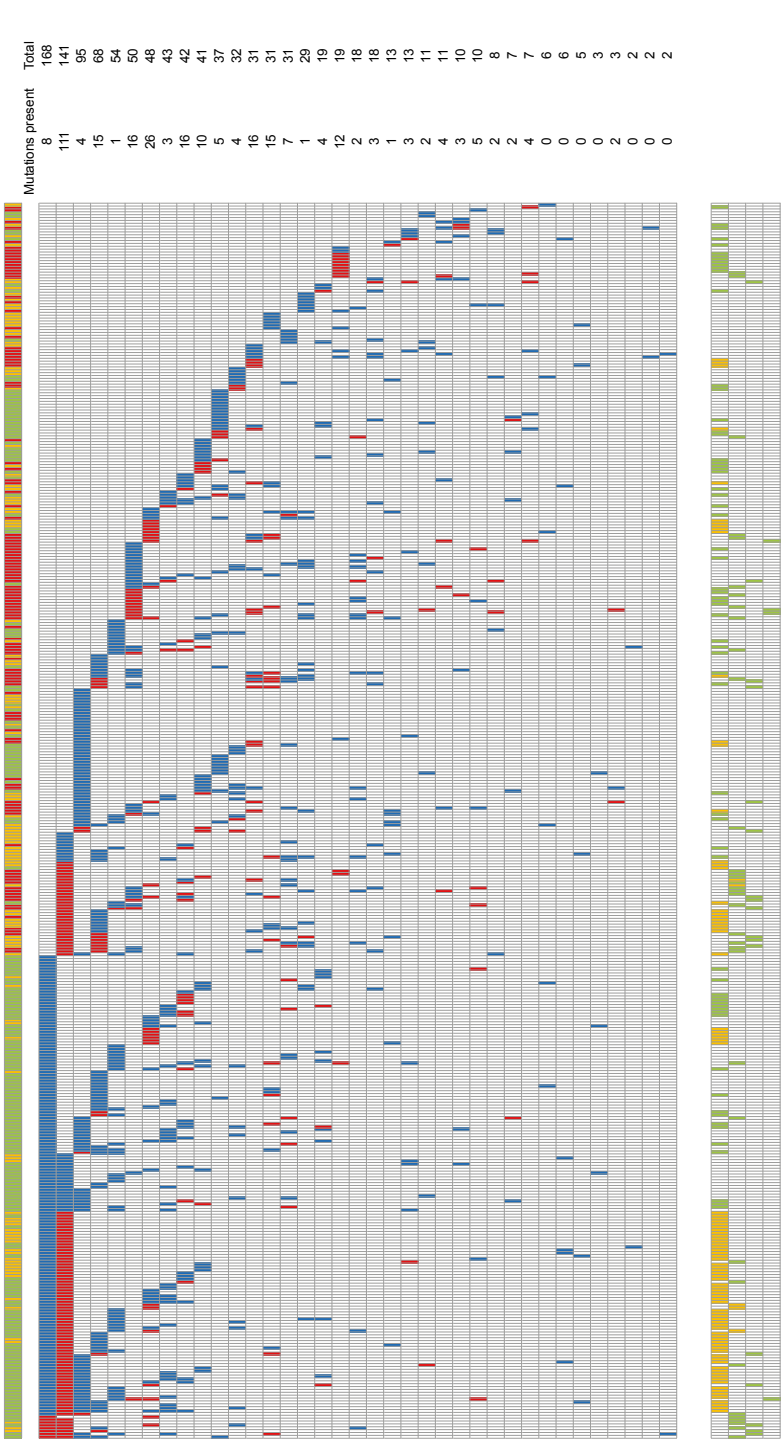


Figure S2A. Mutation status at diagnosis and after induction cycle II without VAF cut-off. Overview of the mutation status at diagnosis (present: blue) and after induction cycle II (present: red). Each column represents an individual patient. The upper panel indicates the ELN 2017 risk category (green: favorable risk, orange: intermediate risk, red: adverse risk). The total number of mutations present at diagnosis and the number remaining after induction cycle II are indicated on the right. The total number of residual mutations after induction treatment are summarised in the bottom figure (green: number of residual mutations present after induction treatment; orange: number of residual mutations exclusively in DNMT3A, TET2 and/or ASXL1). Abbreviations: ELN, European LeukemiaNet. VAF, Variant Allele Frequency

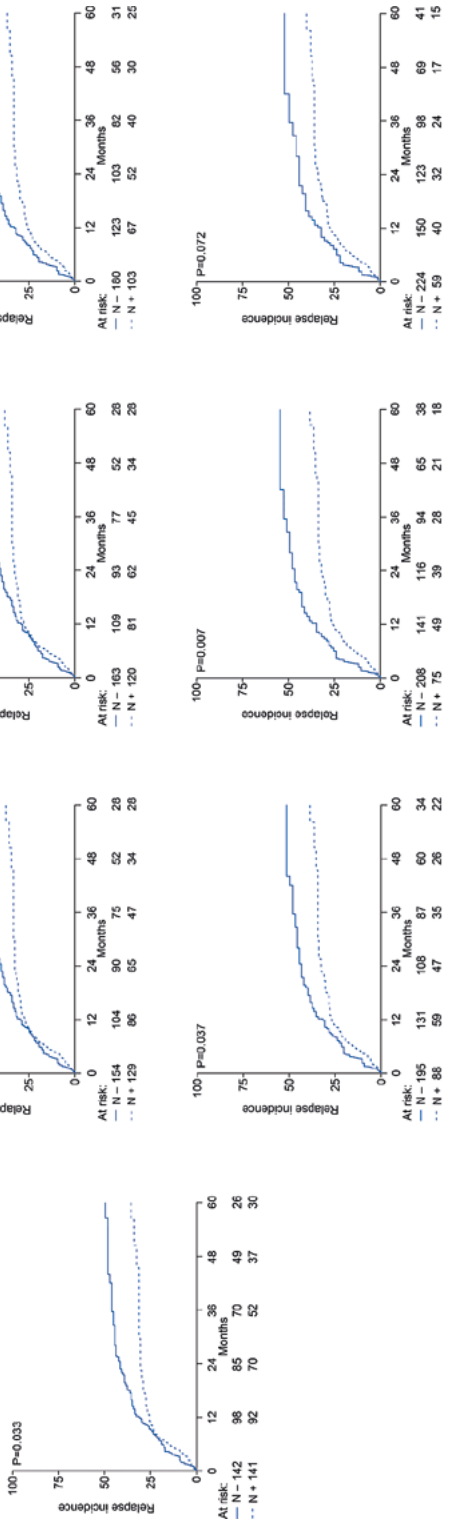


Figure S4. Cumulative incidence of relapse by any detectable mutation in CR with various VAF cut offs in AML patients allocated to the training set (no cut off (A), ≤30% (B), ≤20% (C), ≤10% (D) ≤5% (E), ≤2.5% (F) and ≤1% (G); solid line: mutations detectable in CR (N+); dashed line: mutations not detectable in CR (N-).

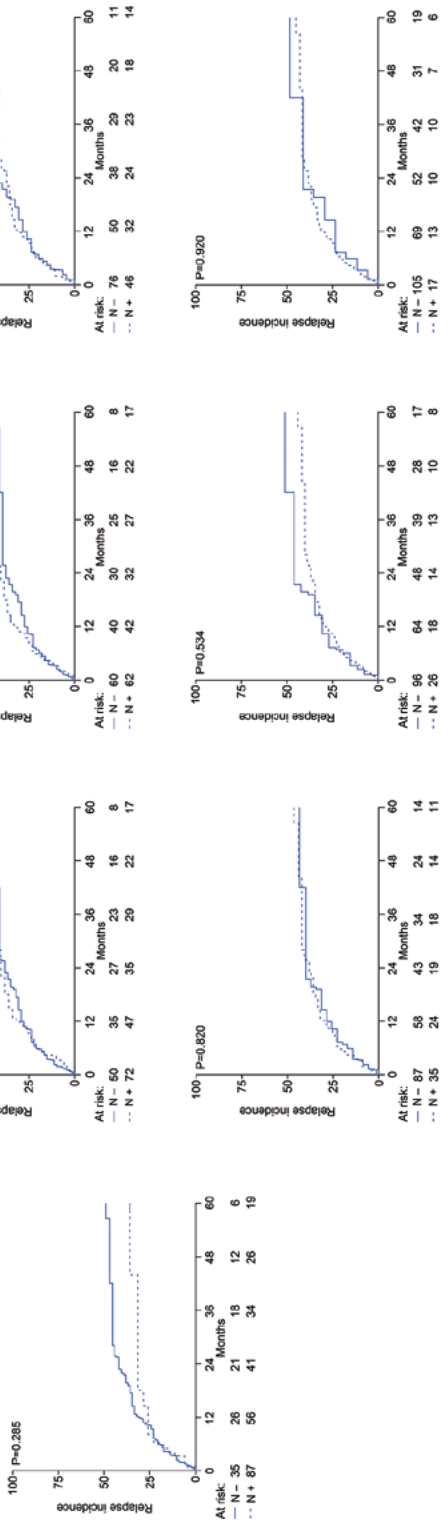


Figure S5. Cumulative incidence of relapse by detectable DNMT3A, TET2 and/or ASXL1 mutations in CR with VAF cut offs in DNMT3A, TET2 and/or ASXL1 mutant AML patients allocated to the training set. No cut off (A), ≤30% (B), ≤20% (C), ≤10% (D), ≤5% (E), ≤2.5% (F) and ≤1.0% (G); solid line: DTA mutations detectable (N+); dashed line: DTA mutations not detectable (N-).

SUPPLEMENTAL TABLES

Table S2. Multivariate Analysis with correction for sampling time

	Relapse		Relapse-free Survival		Overall Survival	
Prognostic Factor	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
Molecular minimal residual disease: detection vs. no detection	1.92 (1.37-2.70)	<0.001	1.66 (1.24-2.23)	0.001	1.66 (1.20-2.30)	0.002
Age: per year	1.01 (1.00-1.03)	0.114	1.02 (1.01-1.03)	0.004	1.03 (1.01-1.04)	0.001
White-cell count per microliter at diagnosis: >100,000 vs. ≤100,000	2.22 (1.37-3.59)	0.001	2.08 (1.37-3.16)	0.001	2.04 (1.28-3.24)	0.003
2017 European Leukemia Network risk classification						
Intermediate vs. Favorable	1.72 (1.15-2.55)	0.008	2.04 (1.45-2.87)	<0.001	2.58 (1.75-3.80)	<0.001
Adverse vs. Favorable	1.81 (1.25-2.63)	0.002	2.21 (1.57-3.09)	<0.001	2.67 (1.82-3.91)	<0.001
Number of chemotherapy cycles to attain complete remission:						
2 cycles vs. 1 cycle	2.10 (1.44-3.06)	<0.001	2.41 (1.73-3.36)	<0.001	2.97 (2.09-4.22)	<0.001
Sampling time in days	0.99 (0.98-1.00)	0.021	0.99 (0.99-1.00)	0.121	0.99 (0.99-1.00)	0.094

Table S3. Multivariate analysis with time-dependent correction for allogeneic HSCT

	Relapse		Relapse-free Survival		Overall Survival	
Prognostic Factor	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
Molecular minimal residual disease: detection vs. no detection	1.88 (1.34-2.64)	<0.001	1.64 (1.22-2.21)	0.001	1.64 (1.18-2.27)	0.003
Age: per year	1.00 (0.99-1.02)	0.556	1.02 (1.00-1.03)	0.024	1.03 (1.01-1.04)	0.001
White-cell count per microliter at diagnosis: >100,000 vs. ≤100,000	2.20 (1.34-3.63)	0.002	2.04 (1.35-3.10)	0.001	2.02 (1.27-3.20)	0.003
2017 European Leukemia Network risk classification						
Intermediate vs. Favorable	2.31 (1.52-3.51)	0.001	2.28 (1.58-3.28)	<0.001	2.56 (1.70-3.84)	<0.001
Adverse vs. Favorable	2.44 (1.64-3.62)	0.001	2.51 (1.75-3.59)	<0.001	2.71 (1.81-4.04)	<0.001
Number of chemotherapy cycles to attain complete remission:						
2 cycles vs. 1 cycle	2.07 (1.42-3.03)	<0.001	2.39 (1.72-3.33)	<0.001	2.95 (2.08-4.20)	<0.001
Post-remission therapy: allo HSCT vs. no allo HSCT)	0.45 (0.31-0.65)	<0.001	0.73 (0.54-0.99)	0.043	0.97 (0.70-1.34)	0.848

Table S4. Multivariate Analysis

	Relapse		Relapse-free Survival		Overall Survival	
Prognostic Factor	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
NGS-MRD vs. MFC-MRD						
NGS-MRD^{pos}/MFC-MRD^{neg} vs NGS-MRD^{neg}/MFC-MRD^{neg}	2.23 (1.43-3.49)	<0.001	1.79 (1.21-2.67)	0.004	1.82 (1.19-2.80)	0.006
NGS-MRD^{neg}/MFC-MRD^{pos} vs NGS-MRD^{neg}/MFC-MRD^{neg}	2.23 (1.35-3.68)	0.002	1.79 (1.11-2.88)	0.017	1.63 (0.96-2.77)	0.069
NGS-MRD^{pos}/MFC-MRD^{pos} vs NGS-MRD^{neg}/MFC-MRD^{neg}	3.92 (2.31-6.67)	<0.001	3.22 (1.99-5.22)	<0.001	2.38 (1.41-4.02)	0.001
Age: per year	1.01 (0.99-1.03)	0.173	1.02 (1.00-1.03)	0.013	1.03 (1.01-1.04)	0.002
White-cell count per microliter at diagnosis: >100,000 vs. ≤100,000	2.62 (1.66-4.13)	<0.001	2.23 (1.42-3.51)	0.001	1.94 (1.17-3.21)	0.001
2017 European Leukemia Network risk classification						
Intermediate vs. Favorable	1.61 (1.04-2.49)	0.034	1.91 (1.29-2.82)	0.001	2.42 (1.57-3.73)	<0.001
Adverse vs. Favorable	1.81 (1.19-2.76)	0.005	2.07 (1.41-3.03)	<0.001	2.52 (1.64-3.87)	<0.001
Number of chemotherapy cycles to attain complete remission:						
2 cycles vs. 1 cycle	2.29 (1.52-3.45)	<0.001	2.43 (1.66-3.55)	<0.001	3.19 (2.14-4.76)	<0.001

FURTHER SUPPLEMENTARY TEXT/TABLES/FIGURES

Further supplemental data can be found online at https://www.nejm.org/doi/suppl/10.1056/NEJMoa1716863/suppl_file/nejmoa1716863_appendix.pdf

CHAPTER 8

EARLY ASSESSMENT OF CLOFARABINE EFFECTIVENESS BASED ON MEASURABLE RESIDUAL DISEASE, INCLUDING AML STEM CELLS

Diana Hanekamp, Lok Lam Ngai,
Jeroen J.W.M. Janssen, Arjan van de Loosdrecht,
Gert J. Ossenkoppele, Jacqueline Cloos

Blood 2020 Nov 10;blood.2020007150.
doi: 10.1182/blood.2020007150

Acute myeloid leukemia (AML) is the most common form of acute leukemia in adults and is characterized by heterogeneity in cytogenetics and molecular aberrations[1]. While insights in the pathogenesis and clonal landscape of AML has increased, the backbone of induction therapy has been the same for many years. In 1973, the '7+3' cytarabine plus anthracycline chemotherapy was first described[2], achieving five-years overall survival (OS) in around 40%-50% in young patients[3], and 20-30% in elderly patients[4]. Several new (targeted) agents have been recently approved for AML[5], which look promising in a subgroup of patients[6], and some (untargeted) chemotherapeutic agents have been introduced. Clofarabine (2-chloro-2'-fluoro-deoxy-9-β-d-arabinofuranosyladenine; a second-generation purine nucleoside analog), has shown potential benefit in young[7] and older[8] AML patients. However, as clofarabine was associated with a risk of severe complications in some studies[9,10], while well-tolerated in others[11,12], further evaluation in dosing and scheduling was warranted. In a large phase III study of the Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON)–Swiss Group for Clinical Cancer Research (SAKK) Cooperative Groups, with 800+ patients enrolled, a significant favorable effect of clofarabine (added to idarubicin and Ara-C) was seen in the European leukemia net (ELN) intermediate-I prognostic risk subgroup[13] (event free survival (EFS) 26%±4 vs. 40%±5; Cox P=0.002; OS 29%±5 vs. 50%±6; Cox P<0.001)[10].

One of the secondary objectives of the study was the assessment of efficacy according to measurable residual disease (MRD). Multiparameter flow cytometry (MFC)-MRD identifies leukemic cells, which can be distinguished from normal cells based on the presence of leukemia associated immunophenotypes (LAIPs)[14]. In addition, specific antibody panels allowed MFC assessment of leukemic stem cells (LSCs)[15]. Samples for MFC-(LSC-) MRD detection were available for a subset of patients and the time point of MFC-MRD assessment after 2 cycles of treatment was used for further analysis; median 82 (43-252) days after start of therapy. See supplement for patient cohort information.

Here we further analyzed the results of this MFC-(LSC-)MRD detection in this trial showing clinical benefit of clofarabine in the ELN Intermediate-I risk group, demonstrating how MFC-(LSC-)MRD results after two cycles of chemotherapy would have predicted the beneficial effect of clofarabine at this early stage.

To qualify an assay for surrogate endpoint of survival benefit, the surrogate should be associated with the outcome and with the effect of the treatment and the outcome[16]. A positive trial showing clinical benefit for an investigational drug is therefore crucial to assess the validity of MFC-MRD to assess effectivity of a novel drug.

MFC-MRD analysis after cycle II was performed in 291 patients distributed over the four ELN2010 risk categories[13]. As clinical benefit was demonstrated in the Intermediate-I ELN risk group only, we focus this analysis on this subgroup in comparison to the remaining cohort. In the Intermediate-I patient group, median MFC-MRD level was 0.041% in the standard arm versus 0.018% in the clofarabine arm (n=49 vs. n=44 respectively; p=0.034) (Table 1). In the remaining cohort, median MFC-MRD level was 0.019% in the standard arm compared to 0.015% in the clofarabine arm (n=105 vs n=93 respectively; p=0.183)

Table 1. MFC-MRD and MFC-LSC-MRD results Intermediate-I patients

			No. of cases evaluated		
	Control	Clofarabine	Control	Clofarabine	p
Minimal residual disease					
MFC-MRD1 median% (range)	0.042 (0.00-3.51)	0.027 (0.00-3.02)	35	46	
MFC-MRD2 median% (range)	0.041 (0.00-11.75)	0.018 (0.00-0.61)	49	44	*
MFC-MRD status					
% MFC-MRD1 positive (n)	31.4 (11)	28.3 (13)	35	46	
% MFC-MRD2 positive (n)	28.6 (14)	20.4 (9)	49	44	
Leukemic stem cells					
MFC-LSC1 median% (range)	0.00 (0.00-0.09)	0.00 (0.00-0.03)	31	44	
MFC-LSC2 median% (range)	0.00 (0.00-2.09)	0.00 (0.00-0.00)	51	39	**
MFC-LSC-MRD status					
% MFC-LSC1 positive (n)	48.4 (15)	31.8 (14)	31	44	
% MFC-LSC2 positive (n)	41.4 (21)	20.5 (8)	51	39	

MFC; Multiparameter flow cytometry, MRD1/LSC1; after cycle I, MRD2/LSC2; after cycle II.
* P≤0.05, ** P≤0.01

(Table S2). Other time points e.g. after cycle I and consolidation show no significant differences (Table 1). When using the ELN recommended MFC-MRD cutoff (i.e. ≥0.1% LAIP cells/WBC)[17], the prevalence of MFC-MRD positive results seem to be lower in de clofarabine arm compared to the standard arm in the Intermediate-I risk group (20.4% versus 28.6%, respectively; not significant; Table 1) and in the remaining cohort (17.2% versus 23.8%, respectively; not significant; Table S2). Distinct differences in the prognostic value of MFC-MRD positivity between treatment arms were found in the Intermediate-I cohort. In the cumulative incidence of relapse analysis (CIR) (Figure 1A), clofarabine treated MFC-MRD^{negative} patients are less likely to relapse than MFC-MRD^{negative} patients treated without clofarabine (pLogrank=0.046). Similarly, although not significant, MFC-MRD^{positive} patients with clofarabine have lower CIR than MFC-MRD^{positive} patients without clofarabine (Figure 1A, p=0.09). These observations were not found in the remaining cohort (Supplement).

Similarly to MFC-MRD levels, lower levels of MFC-LSC-MRD were found in clofarabine treated patients versus patients without clofarabine in the Intermediate-I risk group (ranging from 0.00% to 2.5x10^{-4}%, n=39 vs. 0.00% to 2.09%, n=51 respectively; p=0.014), while no significant difference in MFC-LSC-MRD levels were found in the remaining cohort (clofarabine versus standard, ranging from 0.00% to 3.6x10^{-3}%, n=84 versus 0.00% to 8.6x10^{-3}%, n=120 respectively) (Table 1). For MFC-LSC-MRD detection, presence of any LSC (<0.00001%[15]) is associated with worse outcome. In the Intermediate-I group, more patients are classified MFC-LSC-MRD^{positive} after cycle 2 in the standard group compared

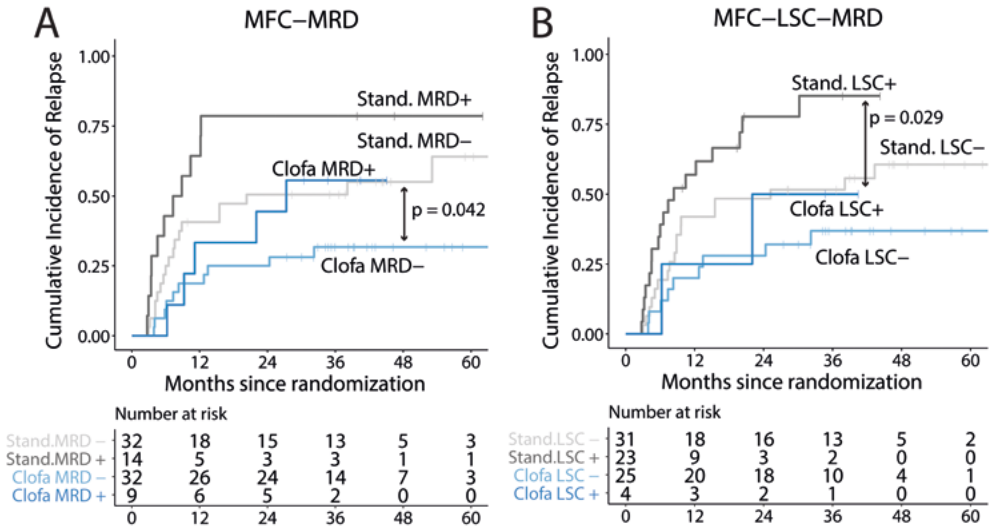


Figure 1. Cumulative incidence of relapse of MFC-MRD and MFC-LSC-MRD by treatment arm. ELN Intermediate-I risk patients are broken down by treatment arm (i.e. standard versus clofarabine 10mg, in grey or blue respectively), and presence or absence of (A) MFC-MRD or (B) MFC-LSC-MRD, using previously published cutoffs (i.e. ≥0.1% after chemotherapy cycle II is called MFC-MRD positive, ≥0.0000% after chemotherapy cycle II is called MFC-LSC-MRD positive[15]). (A) MFC-MRD negative patients treated with clofarabine have a lower incidence of relapse than MFC-MRD^{negative} patients in the standard arm. In parallel, MFC-MRD^{positive} patients with clofarabine have a lower incidence of relapse than MFC-MRD^{positive} patients without clofarabine. (B) Comparable to MFC-MRD, MFC-LSC-MRD show similar results; MFC-LSC-MRD negative patients treated with clofarabine have a distinct lower incidence of relapse than MFC-LSC-MRD^{negative} patients in the standard arm. In parallel, MFC-LSC-MRD^{positive} patients with clofarabine have a lower incidence of relapse than MFC-LSC-MRD^{positive} patients without clofarabine. pLogrank statistics shown when significant.

to the clofarabine treated group (41.2% versus 20.5%, respectively; p=0.043). In the remaining cohort, LSC presence was found in 37.5% in the standard arm versus 28.6% in the clofarabine treated group (p=0.185; Table S2). Although not significant, within the ELN Intermediate-I risk group, MFC-LSC-MRD^{positive} patients treated with clofarabine showed lower CIR than MFC-LSC-MRD^{positive} patients without clofarabine (Figure 1B, p=0.158). Furthermore, clofarabine treated MFC-LSC-MRD^{positive} patients perform similar to MFC-LSC-MRD^{negative} patients treated without clofarabine (Figure 1B, p=0.117), which was not found in the remaining cohort (Figure S1B).

The relatively small subgroup of patients benefiting from clofarabine made it difficult to perform further meaningful analyses such as influence of molecular aberrations (FLT3, NPM1), different time points and multivariate analyses. The results of these analyses are shown in the supplemental file (Figures S3-S5).

In summary, this study has revealed the important evidence that both MFC-MRD and MFC-LSC-MRD mirror the clinical effect of clofarabine by the differences in MFC-(LSC)-

MRD level between treatment arms. However, as treatment with clofarabine is associated with enhanced toxicities, the analyses may be hindered by assessment bias, as patients going off protocol within the first two cycles of chemotherapy due to toxicity, are not sampled for MRD detection. In addition, since among the different prognostic risk groups, treatment protocols differ, the effect of allogeneic or autologous stem cell transplantation could influence results seen in EFS and OS (see supplement).

Importantly, the proportion MFC-MRD positive patients is dependent on the chosen cut-off and could explain why MFC-MRD positivity between treatment groups did not reach significance. Although the current ELN recommended 0.1% cut-off is a robust cut-off in clinical trials, lower cut-offs can also render prognostic relevant MFC-MRD positivity[18,19]. In large multi-center data analysis the clinical relevance of multiple cut-offs is currently being investigated.

Despite these caveats, we show that the results of MFC-MRD and MFC-LSC-MRD after two cycles of chemotherapy reflect the beneficial clinical outcome of clofarabine within the ELN Intermediate-I risk group. In October 2018, the Food and Drug Administration issued a draft guidance on the use of MFC-MRD for accelerated approval[20]. They described that one of the requirements for surrogacy should comprise of clinical trials where treatment effects on the surrogate endpoint correspond to effects on the clinical outcome. We postulate that the difference in median MFC-(LSC)-MRD after treatments can be an interesting point to test this prospectively, as in the currently ongoing NCT03549351 trial[21].

In conclusion, we describe a positive clinical study in which MFC-(LSC-)MRD after two cycles of chemotherapy reflects the effectiveness of clofarabine and the improved clinical outcome in the Intermediate-I subgroup of AML patients. The results of this study point to a compelling need for further investigation of the use of MFC-MRD as instrument for surrogate short term endpoint of effectiveness of new therapies.

ACKNOWLEDGEMENTS

The authors thank all MRD team members of the Amsterdam UMC for sample workup and flow cytometry analysis of all samples.

AUTHORSHIP CONTRIBUTIONS

DH and JC initiated and designed the study. DH and LLN analyzed all data and performed statistical analysis. DH and LLN wrote the manuscript which was further revised by AvdL, GJO and JC and reviewed by all coauthors.

REFERENCES

1. Grimwade, D., Ivey, A. & Huntly, B. J. P. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood 127, 29–42 (2015).

2. Yates, J. W., Wallace, H. J., Ellison, R. R. & Holland, J. F. Cytosine arabinoside (NSC 63878) and daunorubicin (NSC 83142) therapy in acute nonlymphocytic leukemia. CANCER CHEMOTHER.REP. 57, 485–488 (1973).

3. Dombret, H. & Gardin, C. An update of current treatments for adult acute myeloid leukemia. Blood 127, 53–61 (2016).

4. Ossenkoppele, G. J. et al. Lenalidomide added to standard intensive treatment for older patients with AML and high-risk MDS. Leukemia (2020). doi:10.1038/s41375-020-0725-0

5. Paul, S., Rausch, C. R. & Jabbour, E. J. The face of remission induction. Br. J. Haematol. bjh.16353 (2019). doi:10.1111/bjh.16353

6. Stone, R. M. et al. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. N. Engl. J. Med. 377, 454–464 (2017).

7. Thomas, X. et al. Randomized Phase II Study of Clofarabine-Based Consolidation for Younger Adults With Acute Myeloid Leukemia in First Remission. J. Clin. Oncol. 35, 1223–1230 (2017).

8. Tiley, S. & Claxton, D. Clofarabine in the treatment of acute myeloid leukemia in older adults. Ther. Adv. Hematol. 4, 5–13 (2013).

9. Gill, H. et al. Clofarabine, cytarabine, and mitoxantrone in refractory/relapsed acute myeloid leukemia: High response rates and effective bridge to allogeneic hematopoietic stem cell transplantation. Cancer Med. (2020). doi:10.1002/cam4.2865

10. Löwenberg, B. et al. Therapeutic value of clofarabine in younger and middle-aged (18-65 years) adults with newly diagnosed AML. Blood 129, 1636–1645 (2017).

11. Stirrups, R. Clofarabine and cytarabine for acute myeloid leukaemia. Lancet. Oncol. 20, e402 (2019).

12. Crobu, V. et al. A real-world study on clofarabine and cytarabine combination in patients with relapsed/refractory acute myeloid leukemia. Mediterranean Journal of Hematology and Infectious Diseases 11, e2019032 (2019).

13. Dohner, H. et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood 115, 453–475 (2010).

14. Schuurhuis, G. J., Ossenkoppele, G. J., Kelder, A. A. & Cloos, J. Measurable residual disease in acute myeloid leukemia using flowcytometry: approaches for harmonization/ standardization. Expert Rev. Hematol. 11, 17474086.2018.1549479 (2018).

15. Zeijlemaker, W. et al. CD34+CD38- leukemic stem cell frequency to predict outcome in acute myeloid leukemia. Leukemia 33, 1102–1112 (2018).

16. Ciani, O. et al. Use of surrogate end points in healthcare policy: A proposal for adoption of a validation framework. Nat. Rev. Drug Discov. 15, 516 (2016).

17. Schuurhuis, G. J. et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood 131, 1275–1291 (2018).

18. Venditti, A. et al. GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. Blood 134, 935–945 (2019).

19. Shimomura, Y., Hara, M., Maruoka, H., Yabushita, T. & Ishikawa, T. Measurable residual disease evaluated by flow cytometry using leukemia associated immune phenotypes following allogeneic stem cell transplantation is associated with high relapse rates in patients with acute myeloid leukemia. Leukemia and Lymphoma 61, 745–748 (2020).

20. Hematologic Malignancies: Regulatory Considerations for Use of Minimal Residual Disease in Development of Drug and Biological Products for Treatment | FDA. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hematologic-malignancies-regulatory-considerations-use-minimal-residual-disease-development-drug-and. (Accessed: 23rd April 2020)

21. ClinicalTrials.gov. Evaluation of Measurable Residual Disease in Patients With Acute Myeloid Leukemia as Surrogate Endpoint for Survival. https://ClinicalTrials.gov/show/NCT03549351.

SUPPLEMENTARY DATA

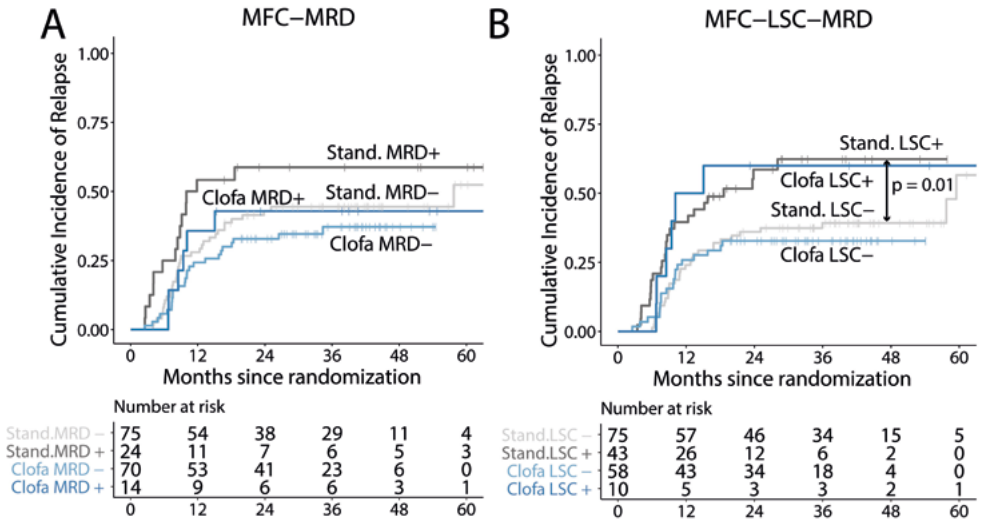


Figure S1. Cumulative incidence of relapse (CIR) of MFC-MRD and MFC-LSC-MRD within remaining ELN risk group patients without ELN Intermediate-I group. Remaining ELN risk group patients without ELN Intermediate-I group are stratified based on treatment arm (i.e. standard versus clofarabine 10mg, in grey or blue respectively), and presence or absence of (**A**) MFC-MRD or (**B**) MFC-LSC-MRD, using previously published cutoffs (i.e. ≥0.1% after chemotherapy cycle II is called MRD positive, ≥0.0000% after chemotherapy cycle II is called LSC positive[15]). (**A**) MFC-MRD negative patients treated with clofarabine have similar CIR compared to MFC-MRD negative patients treated without clofarabine (p=0.826). (**B**) With MFC-LSC-MRD, no differences in CIR are seen between treatment arms of MFC-LSC-MRD^{positive} patients and between treatment arms of MFC-LSC-MRD^{negative} patients. pLogrank statistics shown when significant.

Table S2. Complete cohort MFC-MRD and MFC-LSC-MRD

			No. of cases evaluated		
	Control	Clofarabine	Control	Clofarabine	p
Minimal residual disease					
MFC-MRD1 median% (range)	0.034 (0.00-15.68)	0.022 (0.00-52.46)	97	94	0.011
MRC-MRD2 median% (range)	0.023 (0.00-53.56)	0.015 (0.00-57.39)	105	93	0.183
MFC-MRD status					
% MFC-MRD1 positive (n)	35.1 (34)	22.3 (21)	97	94	0.052
% MFC-MRD2 positive (n)	23.8 (25)	17.2 (16)	105	93	0.252
Leukemic stem cells					
MFC-LSC1 median% (range)	0.00 (0.00-0.09)	0.00 (0.00-1.15)	106	94	0.360
MFC-LSC2 median% (range)	0.00 (0.00-2.09)	0.00 (0.00-0.00)	120	84	0.180
MFC-LSC-MRD status					
% MFC-LSC1 positive (n)	37.7 (40)	45.7 (43)	106	94	0.251
% MFC-LSC2 positive (n)	37.5 (45)	28.6 (24)	120	84	0.185

MFC; Multiparameter flow cytometry, MRD1/LSC1; after cycle I, MRD2/LSC2; after cycle II.

FURTHER SUPPLEMENTARY TEXT/TABLES/FIGURES

Further supplemental data can be found online at http://ars.els-cdn.com/content/image/1-s2.0-S0006497121006820-mmc1.pdf

CHAPTER 9

DISCUSSION AND FUTURE PERSPECTIVES

Acute myeloid leukemia (AML) is a heterogeneous group of clonal and oligoclonal stem cell disorders with variable response to therapy. Despite risk-directed chemotherapy, better supportive care and recent developments of novel therapies targeting specific genetic lesions overall survival (OS) remains to be improved.[1] Current treatment decisions in AML are strongly dependent on a selected number of clinically relevant cytogenetic and molecular genetic markers at diagnosis.[2] Although the majority of patients achieve complete hematological remission (CR) under current intensive induction chemotherapy, relapse rates remain relatively high.[3] A common characteristic of relapsed AML is increased chemoresistance.[4] Early detection or prevention of AML relapse is an urgent clinical need. In this thesis, we have investigated approaches to improve measurable residual disease (MRD) as prognostic factor for relapse development after intensive chemotherapy at several levels: 1) the role of leukemia stem cells (LSC) both at the level of clinical relevance and standardizing the assay, 2) potential improvement of MRD by refining multi-parameter flow cytometry MRD assessment and combining MFC with molecular techniques and 3) establishing an association between MRD level and treatment effectivity in order to investigate the use of MRD as surrogate endpoint.

ROLE OF LEUKEMIA STEM CELLS

The ultimate MRD assay would allow discrimination between cells without relapse initiating potential and those that can regrowth to a new leukemia. At the basis of this conceptual framework of leukemia initiating cells is the notion that LSC have the same potential for self-renewal, multidirectional differentiation, unlimited proliferation, resistance to cell death and multidrug resistance as normal hematopoietic stem cells (HSC; detailed in chapter 2). As such, for further refinement in AML relapse prediction, incorporation of LSC frequency at diagnosis and (perhaps more importantly) in MRD is warranted.

The characterization of LSC is based on the principle that healthy tissue-derived HSC do not express lineage-infidelity antigens, nor overexpress myeloid markers. The antigens aberrantly expressed by LSCs are miscellaneous and include, amongst others: CD2, CD7, CD11b, CD22, CD33, CD44, CD45RA, CD56, CD123, CD366 (TIM3) and CD371 (CLEC12A).[5–7] The leukemia initiating capacity of immunophenotypically-defined LSC has been confirmed by studies demonstrating that AML patient-derived LSC and HSC generate leukemic- and multilineage engraftment upon xenotransplantation, respectively.[8,9]

Recent studies have correlated high immunophenotypically defined $CD34^+CD38^-$ LSC frequencies at the time of diagnosis with subsequent poor prognosis.[10,11] Additionally, absence of $CD34^+CD38^-$ LSCs (in patients where no malignant $CD34^+$ cells are present; so-called CD34-negative patients) was associated with a significantly better OS.[12] We found this relevance of LSC frequency for risk of relapse and poor survival for both adult and pediatric patients. Based on these observations, immunophenotypic assessment of $CD34^+CD38^-$ LSC frequencies at diagnosis would be a good additional marker for risk

classification at diagnosis. Furthermore, the relevance of LSCs at diagnosis could pave the path for new strategies targeting surface markers for eradication of AML LSCs.[13]

As is hypothesized based on the stem cell definition, LSC are more therapy resistant than leukemic cells without stem cell characteristics, and thus likely contribute to the total frequency of MRD cells. We found (chapter 3) that the number of LSC after therapy is an independent predictive factor for relapse.[11] Moreover, patients with both MFC-MRD positivity and LSC-MRD positivity have very poor outcome. It can be suggested that in future clinical studies, allogeneic stem cell transplantations should also be considered in double positive patients within the good risk group.

As presence of any $CD34^{+}CD38^{-}$ LSC was associated with worse outcome compared to patients without LSC, it is critical that sufficient numbers of cells are acquired. However, as the immunophenotypes of LSCs are diverse, and many antigens are needed to fully grasp the total LSC load, this is a challenge.[8,14] Using an antibody panel with a higher number of fluorochromes reduces the number of tubes (and thus cells) needed. Advances in the development of new fluorescent dyes, flow cytometers capable of >20-parameter measurements and analysis software push research toward high-content MFC applications. However, for most medical laboratories, the number of fluorochromes is limited by the flow cytometers present. The 8-color LSC-tube (containing 13 different immunophenotypic markers) designed by our group[7] is currently best fit for this.

Although the prognostic impact of $CD34^{+}CD38^{-}$ LSC is shown in multiple studies by our research group, results need to be validated by others. The development of a dried ready-to-use version of the LSC-tube, and accompanying protocols for use and analysis (chapter 5) significantly contributed in dissemination to other laboratories. Within the ELN research group, the feasibility of LSC testing, and the prognostic impact of LSCs detection will now be validated both in adult and pediatric AML. For the latter group of patients, we have a longstanding collaboration with the Dutch Childhood Oncology Group (DCOG). Further studies on the prognostic value of LSC frequencies in follow-up time points, and the combination with MFC-MRD are warranted for this patient group and these are currently ongoing.

One of the important features that needs further research is the cut-off level of LSC-MRD. In the most recent LSC-MRD research (chapter 3), the cutoff of positivity is determined as 0.0000% (0 events measured). In a previous study by our research group, the cut-off of 0.0001% (1 LSC per 1,000,000 WBCs) was used.[15] While both cut-offs imply the high sensitivity of the technique, usage of these cut-offs (especially the 0.0000% cut-off) requires a very specific assay; as 'noise' (e.g. debris and background expression) could lead to false-positivity. However, similar to MFC-MRD, true absence of LSCs can never be concluded by the applied technique (even when high numbers of cells are acquired, the sample is still a very small fraction of all BM of the patient). Conceivably, the utility of the current cut-off is currently being further evaluated by our group based on the large amount of data that have been gathered over the years. One facet that we are now investigating is the gating of $CD38^{-}$ cells. As LSC markers are evaluated on $CD34^{+}CD38^{-}$ cells, correct gating of CD34

positivity and CD38 negativity is important. While $CD34^+$ cells can be relatively easily be distinguished from CD34- cells using a negative reference population (for example by using lymphocytes), no reference population for CD38 expression is known. For now, our protocols for LSCs gating include $CD38^-$ cells (MFI 10^2) and $CD38^{dim}$ (MFI 10^3) cells. For practical purposes, the prognostic value of the combination of these $CD38^-$ and $CD38^{dim}$ fractions (possibly the 'true' stem cells and the progenitor cells, respectively) is currently investigated. The challenge for this cut-off level of $CD38^-$ cells is the proof of their true biological LSC characteristics (being leukemia initiating cells), which would require novel in vivo engraftment studies in mice.

Identification of LSC with other immunophenotypes than $CD34^+CD38^-$ (i.e. $CD34^-$ or $CD34^+CD38^+$) is possible using MFC by features of drug efflux (side population) and detoxification (Aldehyde dehydrogenase activity) or with molecular techniques (17-gene stemness profile).[16] For CD34-negative AML patients (around 20% of all AML patients), efforts are ongoing to identify cell populations enriched with stem cell properties that would enable LSC tracking implemented in the current MFC assays. Our preliminary data suggest that the $CD34^-CD133^+$ population could play a crucial role in determining prognosis within these CD34-negative patients.

POTENTIAL IMPROVEMENT OF MRD BY REFINING (MFC)-MRD ASSESSMENT

Although we show the improvement of standard FCM-MRD by inclusion of LSC, the current use of MRD would also benefit from further developments. We investigated some approaches to further improve the prognostic value of MRD assessment after the second course of chemotherapy by 1) further standardization of the assay and 2) combining techniques for MRD assessment.

1) FURTHER STANDARDIZATION OF CURRENT MFC-MRD ASSAY

MRD is an independent, post diagnosis, prognostic indicator in AML. In the clinical setting, MRD is currently used to refine the apparent hematologic CR status assessed by morphology.[2,17] Especially MFC-MRD is well established as an independent prognostic factor for relapse, RFS and OS, particularly when measured before transplantation.[15,18–22] Measuring MRD at other time points can help to identify a group of patients with poor prognosis, such after the first cycle of chemotherapy[15,23,24] and after consolidation.[15,25,26]

The standard strategy to eradicate AML is intensive chemotherapy either in multiple cycles or followed by hematopoietic stem cell transplantation, with allogeneic stem cell transplantation (alloSCT) having greater anti-leukemic activity compared to autologous stem cell transplantation (autoSCT).[2,27] Given the potential toxicity, alloSCT would preferably be averted in those who do not need it.[28,29] Risk classification systems are therefore designed to allocate patients to the best suited therapy. Recent studies make

use of risk classification systems that allow MRD to guide post induction treatment.[30,31] A few published studies on intensified treatment based on presence of MRD in pediatric AML and t(8;21) patientssuggested that this type of MRD-guided therapy may improve outcome.[32,33] The results of the prospective HOVON132 study, where MRD guides treatment in the intermediate risk group (i.e. allocating MRD positive patients to (standard) alloSCT and MRD negative patients to (less intense) autoSCT) are eagerly awaited. For patients in the poorest risk category, presence of MRD conveys an extremely poor outcomes. Persistent MRD positivity in these patients may be therefore be an indication to enroll patients in early phase clinical studies or to offer palliative treatment and spare patients the potential intense toxicity of alloSCT .

Although the prognostic value of MRD allows to stratify patients accordingly to risk of relapse, the ultimate goal is developing MRD assays that can more accurately determine risk of relapse in an individual patient, allowing post-remission therapies to be individually tailored.[34] However, since 30% of MRD negative patients still relapse and 30% of the MRD positive patients does not –although remaining treatment could have eradicated the MRD-, the predictive value of the MRD status is not yet high enough for the individual patient.[35] There are several aspects of the current MRD assay that can be considered in more detail.

SAMPLE QUALITY AND HEMODILUTION

For adequate MRD measurement, the quality and quantity of the BM samples is crucial. However, these may vary greatly and thereby significantly impact the performance of the assay. When assessing the BM compartment, the first 'pull' provides a good representation of its content. As additional volume is pulled from the same spot, the likelihood of peripheral blood (PB) contaminating the sample (i.e. hemodilution) increases, potentially diluting the BM components.[36] Consequently, a sample taken after multiple pulls can underrepresent the true amount of disease burden.[37] Traditionally, the first pull has been reserved for morphology testing. However, it is now recommended that the first pull be used for MRD analysis.[38,39] In addition to quality of the BM aspirate, sampling, transport, sample processing and the time between sampling and measurement (especially in a multicenter setting) all influence the quality of the sample. Maintaining high quality samples, results in best quality results, and therefore, protocols concerning transport of material have recently been formulated.[38]

A possible alternative for circumventing underestimation of MRD due to hemodilution, is defining the balance between normal and tumor (immature) progenitor cells.[40] The denominator in standard MRD calculation is WBC (i.e. MRD= LAIP positive events/ WBC) which does not take into account the composition of the primitive blasts cells. Initial research into this so-called primitive marker MRD (PM-MRD= LAIP positive events/ primitive marker cells) revealed important prognostic impact,[41] which was retrospectively confirmed in chapter 6. Application of high level PM-MRD in MRD-negative patients results in better risk classification and in a lower number of false negatives. Additionally,

application of PM-MRD aids in standardization of MRD omitting the rather subjective step of WBC gating.[42]

STANDARDIZATION

Reproducibility remains a major concern with the use of MFC MRD, particularly between institutions in clinical trials and/or routine diagnostics. Furthermore, comparison of MFC MRD data between laboratories is complex. To a large degree, these issues likely arose due to the lack of standardized protocols and the somewhat subjective art of gating flow cytometric data.[43] To correctly interpret and use MFC MRD results, standardization and/or harmonization of MRD measurements are of vital importance. A standardized protocol describing the assay should fulfill specific requirements including adequate sample processing, instrument settings, marker/panel selection, gating strategies and clinically validated cut off points for specific time points for MRD positivity. The use of standardized protocols in other hematological malignancies (CLL, multiple myeloma and B-ALL), illustrated that the presence of these protocols resulted in great improvement of reproducibility.[44–48] These data furthermore suggest that standardization of MFC MRD is feasible, and it is not the impediments inherent to the technique, rather the lack of devotion of the researchers that hinders true standardization.[49] For AML MRD, standardization and harmonization have been emphasized in recent years.[38,40,42,50,51]

The future of MFC MRD detection likely entails the use of automated data analysis programs. Implementation of these circumvent the variability in operator skills and required expert knowledge, but comes with its own challenges, e.g. the need of an extensive database of 'normal' and regenerating BM at different time points and perhaps also for each leukemic associated immunophenotype (LAIP), each fluorochrome combination and all different (pre)treatments.[52]

Besides the better qualification of the assay, there are features of the leukemia cells and their BM niche that also influence the MRD results.

CORRECTION FOR BACKGROUND

Generally, two approaches to assess MFC MRD are employed, with each its own limitations. (I) The LAIP-based approach where the LAIP is identified at diagnosis and followed during therapy[42] and (II) the Different from Normal (DfN) approach[50] in which any aberrant pattern of cell surface markers compared to normal expression in 'normal BM' is designated as being residual leukemia. Besides differences, both methods rely on the accurate discrimination between leukemic cells and non-leukemic cells. Knowledge of background LAIP expression on normal cells is required to avoid false positive results, and most studies therefore use a threshold to distinguish MRD positivity from background expression.[39] However, different LAIPs have variable sensitivity and specificity. For example, the $CD34^{-}/CD133^{+}$ LAIP had a higher than ten-fold lower background compared to the $CD34^{+}CD15^{+}$ LAIP as shown in chapter 5. The use of one cut-off could therefore result in false negatives.[53] In our cohort of

the HOVON 102, subtraction of nBM background levels did not alter the number of MRD positive cases. Nevertheless, whether background expression is best analyzed in normal BM or regenerating BM is a topic for future research.[53–55] In general, it is accepted to assign a patient as MRD positive if residual leukemic cells are detected in the BM above a certain cut-off in most studies, or above zero in so-called 'any-MRD' studies, but MRD negativity is not as clear. The difference between nondetectable and truly negative MRD is often debated by sample/measurement quality and sensitivity inherent to the assay.

CLONAL EVOLUTION: COMBINATION WITH DFN

Since the LAIP-based approach only evaluates the dominant LAIPs detected at diagnosis, and does not evaluate the possibility of new LAIPs that arise due to clonal evolution, it inherently holds a risk of false negativity.[56] Next to that, the LAIP-based approach can only be utilized for patients that had a pre-treatment sampling (and thus, difficult for tertiary referral centers or reference laboratories). The DfN approach can be viewed as a more general approach for the detection of residual disease by comparing the patterns and shapes found with patterns and shapes in normal hematopoiesis.[50] It is important to realize that LAIPs are DfN abnormalities, and that the differences between the approaches are likely to disappear when enough events are measured at diagnosis, allowing identification of the smaller LAIPs, with sufficient large panels of antibodies.[39] Following this, antibody panels with higher number of fluorochromes in one flow cytometry tube (e.g. 18-color panels) would further enable identification of multiple LAIPs simultaneously, which perhaps make the sum of complete MRD.

In the European LeukemiaNet (ELN) MRD guidelines, both philosophies are combined into the LAIP-based DfN approach: assessing all immunophenotypic aberrancies (i.e. the dominant LAIPs at diagnosis, LAIPS that arise due to clonal evolution and LAIPs that arise due to immunophenotypic shifts after therapy). Several recent studies have been using this LAIP-based DfN approach.[23,53,57] However, comparisons of the prognostic impact of the different approaches is only explorative[58], but can be analyzed in the near future in the HOVON132 study, where we measured MRD with the LAIP method but also measured all possible upcoming LAIPs during therapy. Especially in the light of novel targeted drugs, the LAIP-based DfN approach may be essential to gain information about the therapeutic efficacy against the dominant clone.

2) COMBINATION WITH MOLECULAR MRD TECHNIQUES

Molecular MRD techniques, with reverse-transcription quantitative PCR (RT-qPCR) of NPM1 as prime example, appeared at the beginning of the current century and has rapidly proven its effectiveness in the identification of patients with residual leukemia.[59–61] The most general accepted molecular method are real-time PCR-based approaches, aiming for sensitivities between 10^{-4} and 10^{-6}. These are only applicable in the 40-50% of patients with suitable gene mutations (as NPM1 and core-binding factor

translocations).[62] Newer molecular detection techniques are digital-droplet PCR[63] and the promising next-generation sequencing (NGS)[64] (Chapter 7). Interestingly, while all techniques have their own principles and therefore their own limitations, limited studies are published in which multiple methods are applied and compared, although currently, in many laboratories MFC and molecular MRD are measured in parallel.[65–67]

More than half of the patients with cytogenetically normal AML harbor a mutation in NPM1, which remains present at relapse.[68,69] For these patients, the recommended method for MRD assessment is RT-qPCR[39], which affords a sensitivity of $1:10^{-5}$ to $1:10^{-6}$ (100- to 1000-fold greater than that achieved in FCM-MRD).[59,60,70–75] With this recommendation, clinical management of NPM1 mutated patients could appear clearly defined. However, comparison between MRD detection using NPM1 RT-qPCR and FCM is not extensively studied and the predictive value of discordant results between NPM1-MRD and FCM-MRD (especially those that are $NPM1mut^{negative}/FCM^{positive}$) is unclear.[76,77]

The development of flow cytometric methods to detect (mutant, or aberrantly cytoplasmic delocalized) NPM1 has gained interest over the last decade.[78–82] However, implementation of mutant NPM1 specific antibodies in flow cytometry is likely mostly interesting for research purposes, but unfeasible in clinical setting. The drawback of qPCR is that it can only be used for specific mutations that are only present in about 40-50% of patients. Therefore, other novel molecular techniques are also currently under investigation. The use of NGS in MRD assessment identifying persistent molecular MRD, was only recently shown as an independent predictor for survival (chapter 7). As such, guidelines regarding the use of NGS-MRD in clinical practice are far from existing, and several challenges have to be overcome before it will be routinely used. The sensitivity of NGS for MRD is coherent to the correct discrimination between truly mutated nucleotides and background 'noise'. For better sensitivity, methods to improve the error rate, and reduction of false positive calls have been developed.[83]

In our study, only mutations persistent from diagnosis, but not mutations associated with clonal hematopoiesis (CHIP) (i.e. DNMT3A, ASXL1 and TET2, collectively termed DTA), present with allele frequencies above background signals (i.e. as low as 0.02%) were associated with increased relapse risk. A method to accurately discriminate true leukemic mutations from (non-leukemic or pre-leukemic) CHIP mutations is currently lacking, but will likely contribute to better relapse prediction. Similar to MFC-MRD, debates regarding sensitivity and specificity of NGS are ongoing and discussed elsewhere.[83]

The complementary and highly predictive value for relapse of tandem targeted NGS-MRD and MFC-MRD detection after allogeneic transplantation was shown by Getta and colleagues.[84] Tandem NGS-MRD and MFC-MRD evaluation (now with a larger set of genes and in a substantially larger patient cohort) in CR showed similar complementary and predictive values for relapse. The concordance of NGS-MRD and MFC-MRD results reached almost 70%. Interestingly, this is similar to multiple myeloma,[85,86] where NGS-MRD and MFC-MRD are both regarded as appropriate techniques, with the preference based on local availability.[87] Chapter7 shows that the combination of NGS-MRD and MFC-MRD

aids in identification of false-negative MRD results when only one technique was used (41 patients were NGS-MRD^{negative} but MFC-MRD^{positive} and 64 patients were MFC-MRD^{negative} but NGS-MRD^{positive} with a 4-year relapse rate around 50% in both). However, whether the use of both techniques simultaneously *should* be advised is perhaps a more political/practical and economical question, than a biological one. Further understanding of the biology of clonal hematopoiesis (after high-dose chemotherapy) and increased sensitivity of the NGS assay are needed to resolve the discordant cases and to determine whether or not both technologies are required.

COMBINATION OF TECHNIQUES

The studies presented in chapter 3 and 7 are largely performed on the same patient cohort. The combination of NGS-MRD, MFC-MRD and LSC-MRD could thus be explored. Preliminary analysis shows that the overlap between the three techniques is 51.9% while 3.7% were positive for all three techniques simultaneously and of the 216 evaluated patients, 96 were negative for all three techniques (44.4%). Combining MRD approaches allows detection of MRD positivity in patients otherwise potentially falsely classified MRD-negative. While this suggests lower levels of false-negative cases, i.e.MRD-negative patients who develop relapse, our data showed that 30.5% of patients NGS-MRD negative, MFC-MRD negative and LSC-MRD negative nevertheless relapsed. As all three methods described here have their own pitfalls, as discussed previously, adaptation of the cut-off level (for MFC-MRD and LSC-MRD), addition of other immunophenotypic (stem) cell markers and the implementation of more sensitive sequencing methods (for example for NPM1) may lower this percentage. However, in a study by Ivey et al. where NPM1 mutated patients were screened after two cycles of chemotherapy for the presence of residual NPM1 mutations using a highly sensitive RT-qPCR method, comparable 30 percent of false-negative cases were found.[70] Perhaps, it will remain an utopia to think we can always detect a pending relapse, with many biological and environmental determinants still to be investigated.

BEST OF BOTH WORLDS

Overall, MRD detection is regarded as the most important post-therapeutic factor to identify high-risk patients and implementation in the clinic is therefore highly warranted. However, with the recent studies comparing the different methods, showing that all of these give additive information and are perhaps all needed, selection of the best MRD method for new clinical trials could be challenging, or rather be a 'cost-benefit debate'. It may be suggested that in the near future not one MRD method is universally applied for all patients. It could be most time-saving, cost-effective and sample-efficient to tailor subsets of patients to specific methods based on patients-specific or leukemia-specific characteristics. For example, for NPM1 mutated patients real-time PCR is likely the preferred method. Similarly, patients without genetic mutations at diagnosis are likely better off

with MFC-MRD or LSC-MRD. In the recently finalized HOVON132 trial, results of NPM1 detection after induction therapy were leading in the clinical decision for post-remission treatment. Simultaneously, patients without NPM1 mutation are assessed by MFC-MRD for post-remission treatment choice. However, while numbers are limited, presence of MFC-MRD in NPM1-MRD^{negative} NPM1 mutated patients is correlated with higher risk of relapse (chapter 3). Hence, selection of one technique based on patient specifics could be preferred in terms of costs, time and sample, but perhaps relinquishes in sensitivity.

ADDITIONAL USAGE OF MRD BEYOND RELAPSE RISK AFTER INDUCTION

MRD AS PREDICITVE FACTOR FOR THE INDIVIDUAL PATIENT?

MRD is now established as an important independent prognostic factor identifying a subgroup of patients with increased risk for relapse. Although, as a single factor, MRD is not a good predictive factor for the individual patient, "prediction is fundamental to therapy decisions in AML but the more important question is the extent to which incorporating results of MRD-testing improves predictive accuracy" (as stated by Estey and Gale).[88] Several models are being investigated by our group and others[89] to identify an algorithm that can more accurately predict the relapse risk for the individual patient. For challengers of MRD, other arguments include the lack of a standard assay for MRD, the question whether MRD truly represents remission status, whether BM sampling is necessary, when MRD needs to be assessed and whether one measurement suffices (and whether patients would accept repetitive bone marrow aspirations). Furthermore, the costs of a single MRD analysis is estimated to be around $350-500 and can only be justified when results are used for rational treatment decisions.[90]

PERI-TRANSPLANT MRD DETECTION

In addition to the use of MRD for tailored therapy, MRD detection may be useful to monitor disease kinetics during therapy or after transplantation using serial assessments to detect a pending relapse. [54,91-92] In contrast to a single MRD measurement, sequential measurements inform whether the burden of disease is increasing or decreasing over time. In pediatric AML, sequential qPCR MRD measurements during standard consolidation therapy showed the possibility to identify patients at high risk of relapse at an early time point.[93] Furthermore, as illustrated in the supplemental data of chapter 8, new therapeutics could accelerate disease clearance and thereby influence the optimal time point for MRD measurements. Altogether, serial MRD measurements during and after therapy may have important prognostic and therapeutic implications and could further aid in the understanding of the biology of the disease and relapse. However, repetitive BM aspirations are a considerable burden to the patient. Peripheral blood (PB) would then be an attractive alternative source for MRD assessment.[94]

Besides obvious advantageous logistic reasons, PB MFC-MRD has shown increased specificity, but lower sensitivity, compared to BM MFC-MRD.[39] While more studies are needed, recent publications support the idea that the lower sensitivity of PB MRD is outweighed by higher specificity of PB MRD.[93–95]

SURROGATE ENDPOINT FOR NOVEL THERAPIES

Another possible application of MRD is the use as a surrogate endpoint for clinical studies. As more promising targeted therapies are emerging, a relatively early endpoint is searched for to speed up the development of AML treatments. In contrast to the generally accepted outcome measures as EFS and OS, which take years, MRD can be measured as soon as the therapy is started, potentially accelerating clinical trial outcomes. To date, no studies are published that attest true surrogacy of MRD for AML (although chapter 8 could be accepted as a proof-of-principle). However, multiple studies that are currently ongoing have MRD as primary endpoint.[96] Results of these studies could prove the eligibility of MRD as a surrogate endpoint.

In conclusion, the results described in this thesis illustrate that MRD measurement is of crucial importance in the treatment of AML. As the clinical practice is eagerly anticipating the use of MRD for clinical decision making, there is an urgent need to further define and refine the utility of MRD. The MRD assay should be harmonized, standardized, qualified and validated in prospective trials, for the above mentioned clinical uses. Furthermore, while there are a number of intriguing possibilities as a result of MRD assessment (i.e. revision of risk classification and choice of SCT, selection of MRD-specific targeted therapy), a number of key questions must be resolved before MRD reaches its full clinical potential, such as the right time point, cut-off level, method for each of the utilities of MRD.

REFERENCES

1. Shah, A., Andersson, T. M.-L., Rachet, B., Björkholm, M. & Lambert, P. C. Survival and cure of acute myeloid leukaemia in England, 1971-2006: a population-based study. *Br. J. Haematol.* **162**, 509–516 (2013).

2. Döhner, H. *et al*. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017).

3. Döhner, H., Weisdorf, D. J. & Bloomfield, C. D. Acute Myeloid Leukemia. *N. Engl. J. Med.* **373**, 1136–1152 (2015).

4. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA. Cancer J. Clin.* **70**, 7–30 (2020).

5. Taussig, D. C. *et al*. Hematopoietic stem cells express multiple myeloid markers: Implications for the origin and targeted therapy of acute myeloid leukemia. *Blood* **106**, 4086–4092 (2005).

6. Kersten, B. *et al*. CD45RA, a specific marker for leukaemia stem cell sub-populations in acute myeloid leukaemia. *Br. J. Haematol.* **173**, 219–235 (2016).

7. Zeijlemaker, W. *et al*. A simple one-tube assay for immunophenotypical quantification of leukemic stem cells in acute myeloid leukemia. *Leukemia* **30**, 439–46 (2016).

8. Sarry, J. *et al*. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J. Clin. Invest.* **121**, 384–395 (2011).

9. Jan, M. *et al*. Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5009–14 (2011).

10. van Rhenen, A. *et al*. High Stem Cell Frequency in Acute Myeloid Leukemia at Diagnosis Predicts High Minimal Residual Disease and Poor Survival. *Clin. Cancer Res.* **11**, 6520–6527 (2005).

11. Terwijn, M. *et al*. Leukemic Stem Cell Frequency: A Strong Biomarker for Clinical Outcome in Acute Myeloid Leukemia. *PLoS One* **9**, e107587 (2014).

12. Zeijlemaker, W. *et al*. Absence of leukaemic CD34+ cells in acute myeloid leukaemia is of high prognostic value: A longstanding controversy deciphered. *Br. J. Haematol.* **171**, 227–238 (2015).

13. Maakaron, J. E., Rogosheske, J., Long, M., Bachanova, V. & Mims, A. S. CD33-Targeted Therapies: Beating the Disease or Beaten to Death? *J. Clin. Pharmacol.* jcph.1730 (2020). doi:10.1002/jcph.1730

14. Bachas, C. *et al.* The role of minor subpopulations within the leukemic blast compartment of AML patients at initial diagnosis in the development of relapse. *Leukemia* **26**, 1313–1320 (2012).

15. Terwijn, M. *et al*. High Prognostic Impact of Flow Cytometric Minimal Residual Disease Detection in Acute Myeloid Leukemia: Data From the HOVON/SAKK AML 42A Study. *J. Clin. Oncol.* **31**, 3889–3897 (2013).

16. Ng, S. W. K. W. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).

17. Ossenkoppele, G. J. & Schuurhuis, G. J. MRD in AML: It is time to change the definition of remission. *Best Pract. Res. Clin. Haematol.* **27**, 265–271 (2014).

18. Wang, Z. *et al.* The applicability of multiparameter flow cytometry for the detection of minimal residual disease using different-from-normal panels to predict relapse in patients with acute myeloid leukemia after allogeneic transplantation. *Int. J. Lab. Hematol.* **41**, 607–614 (2019).

19. Tan, Y. *et al.* Minimal residual disease may be an early prognostic indicator for newly diagnosed acute myeloid leukemia patients induced by decitabine-based chemotherapy. **24**, 552–558 (2019).

20. Patkar, N. *et al.* Utility of Immunophenotypic Measurable Residual Disease in Adult Acute Myeloid Leukemia—Real-World Context. *Front. Oncol.* **9**, 450 (2019).

21. Minetto, P. *et al. Early minimal residual disease assessment after AML induction with fludarabine, cytarabine and idarubicin (FLAI) provides the most useful prognostic information. British Journal of Haematology* **184**, 457–460 (Blackwell Publishing Ltd, 2019).

22. Othus, M. *et al.* Effect of measurable ('minimal') residual disease (MRD) information on prediction of relapse and survival in adult acute myeloid leukemia. *Leukemia* **30**, 2080–2083 (2016).

23. Freeman, S. D. *et al.* Measurable Residual Disease at Induction Redefines Partial Response in Acute Myeloid Leukemia and Stratifies Outcomes in Patients at Standard Risk Without NPM1 Mutations. *J. Clin. Oncol.* **36**, 1486 (2018).

24. Hoffmann, A. P. *et al.* Early achievement of measurable residual disease (MRD)-negative complete remission as predictor of outcome after myeloablative allogeneic hematopoietic cell transplantation in acute myeloid leukemia. *Bone Marrow Transplantation* **55**, 669–672 (2020).

25. Liu, J. *et al.* The significance of peri-transplantation minimal residual disease assessed by multiparameter flow cytometry on outcomes for adult AML patients receiving haploidentical allografts. *Bone Marrow Transplant.* **54**, 567–577 (2019).

26. Shah, M. V. *et al.* Early Post-Transplant Minimal Residual Disease Assessment Improves Risk Stratification in Acute Myeloid Leukemia. *Biol. Blood Marrow Transplant.* **24**, 1514–1520 (2018).

27. Versluis, J. *et al.* Graft-Versus-Leukemia Effect of Allogeneic Stem-Cell Transplantation and Minimal Residual Disease in Patients With Acute Myeloid Leukemia in First Complete Remission. *JCO Precis. Oncol.* 1–13 (2017). doi:10.1200/po.17.00078

28. Versluis, J. & Cornelissen, J. J. Risks and benefits in a personalized application of allogeneic transplantation in patients with AML in first CR. *Seminars in Hematology* **56**, 164–170 (2019).

29. Gooley, T. A. *et al.* Reduced Mortality after Allogeneic Hematopoietic-Cell Transplantation. *N. Engl. J. Med.* **363**, 2091–2101 (2010).

30. Burnett, A. K. *et al.* Curability of patients with acute myeloid leukemia who did not undergo transplantation in first remission. *J. Clin. Oncol.* **31**, 1293–1301 (2013).

31. Pfirrmann, M. *et al.* Prediction of post-remission survival in acute myeloid leukaemia: A post-hoc analysis of the AML96 trial. *Lancet Oncol.* **13**, 207–214 (2012).

32. Rubnitz, J. E. *et al.* Minimal Residual Disease-Directed Therapy for Childhood Acute Myeloid Leukemia: Results of the AML02 Multicenter Trial. *Lancet Oncol* **11**, 543–552 (2010).

33. Zhu, H. H. *et al.* MRD-directed risk stratification treatment may improve outcomes of t(8;21) AML in the first complete remission: Results from the AML05 multicenter trial. *Blood* **121**, 4056–4062 (2013).

34. Short, N. J. & Ravandi, F. How close are we to incorporating measurable residual disease into clinical practice for acute myeloid leukemia? *Haematologica* **104**, 1532–1541 (2019).

35. Othus, M., Gale, R. P., Hourigan, C. S. & Walter, R. B. Statistics and measurable residual disease (MRD) testing: uses and abuses in hematopoietic cell transplantation. *Bone Marrow Transplantation* **55**, 843–850 (2020).

36. Loken, M. R., Chu, S. C., Fritschle, W., Kalnoski, M. & Wells, D. A. Normalization of bone marrow aspirates for hemodilution in flow cytometric analyses. *Cytom. Part B - Clin. Cytom.* **76**, 27–36 (2009).

37. Helgestad, J., Rosthøj, S., Johansen, P., Varming, K. & Østergaard, E. Bone marrow aspiration technique may have an impact on therapy stratification in children with acute lymphoblastic leukaemia. *Pediatr. Blood Cancer* **57**, 224–226 (2011).

38. Cloos, J. *et al.* Comprehensive Protocol to Sample and Process Bone Marrow for Measuring Measurable Residual Disease and Leukemic Stem Cells in Acute Myeloid Leukemia. *J. Vis. Exp.* **March 5**, (2018).

39. Schuurhuis, G. J. *et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* **131**, 1275–1291 (2018).

40. Schuurhuis, G. J., Ossenkoppele, G. J., Kelder, A. A. & Cloos, J. Measurable residual disease in acute myeloid leukemia using flowcytometry: approaches for harmonization/ standardization. *Expert Rev. Hematol.* **11**, 17474086.2018.1549479 (2018).

41. Terwijn, M. *et al.* Minimal residual disease detection defined as the malignant fraction of the total primitive stem cell compartment offers additional prognostic information in acute myeloid leukaemia. *Int. J. Lab. Hematol.* **34**, 432–441 (2012).

42. Zeijlemaker, W., Kelder, A., Cloos, J. & Schuurhuis, G. J. Immunophenotypic Detection of Measurable Residual (Stem Cell) Disease Using LAIP Approach in Acute Myeloid Leukemia. *Curr. Protoc. Cytom.* **91**, (2019).

43. Maecker, H. T. *et al.* A model for harmonizing flow cytometry in clinical trials. *Nature Immunology* **11**, 975–978 (2010).

44. Rawstron, A. C. *et al.* Improving efficiency and sensitivity: European research initiative in CLL (ERIC) update on the international harmonised approach for flow cytometric residual disease monitoring in CLL. *Leukemia* **27**, 142–149 (2013).

45. Dowling, A. K., Liptrot, S. D., O'Brien, D. & Vandenberghe, E. Optimization and validation of an 8-color single-tube assay for the sensitive detection of minimal residual disease in B-cell chronic lymphocytic leukemia detected via flow cytometry. *Lab Med.* **47**, 103–111 (2016).

46. Mathis, S. *et al.* Comparison of cross-platform flow cytometry minimal residual disease evaluation in multiple myeloma using a common antibody combination and analysis strategy. *Cytom. Part B Clin. Cytom.* n/a-n/a (2014). doi:10.1002/cytob.21200

47. Björklund, E. *et al.* Quality control of flow cytometry data analysis for evaluation of minimal residual disease in bone marrow from acute leukemia patients during treatment. *J. Pediatr. Hematol. Oncol.* **31**, 406–415 (2009).

48. Dworzak, M. N. *et al.* Standardization of flow cytometric minimal residual disease evaluation in acute lymphoblastic leukemia: Multicentric assessment is feasible. *Cytom. Part B - Clin. Cytom.* **74**, 331–340 (2008).

49. Paietta, E. Consensus on MRD in AML? *Blood* **131**, 1265–1266 (2018).

50. Wood, B. L. Acute Myeloid Leukemia Minimal Residual Disease Detection: The Difference from Normal Approach. *Curr. Protoc. Cytom.* **93**, (2020).

9

51. Brooimans, R. A. *et al.* Immunophenotypic measurable residual disease (MRD) in acute myeloid leukemia: Is multicentric MRD assessment feasible? *Leuk. Res.* **76**, 39–47 (2019).

52. Buldini, B., Maurer-Granofszky, M., Varotto, E. & Dworzak, M. N. Flow-Cytometric Monitoring of Minimal Residual Disease in Pediatric Patients With Acute Myeloid Leukemia: Recent Advances and Future Strategies. *Front. Pediatr.* **7**, 1–8 (2019).

53. Sui, J. N. *et al.* Identifying leukemia-associated immunophenotype-based individualized minimal residual disease in acute myeloid leukemia and its prognostic significance. *Am. J. Hematol.* **94**, 528–538 (2019).

54. Rossi, G. *et al.* Leukemia-associated immunophenotypes subdivided in "categories of specificity" improve the sensitivity of minimal residual disease in predicting relapse in acute myeloid leukemia. *Cytom. Part B - Clin. Cytom.* **98**, 216–225 (2020).

55. Hanekamp, D., Bachas, C., van de Loosdrecht, A., Ossenkoppele, G. & Cloos, J. Re: Myeloblasts in normal bone marrows expressing leukaemia-associated immunophenotypes. *Pathology* **52**, 289–291 (2020).

56. Zeijlemaker, W., Gratama, J. W. & Schuurhuis, G. J. Tumor heterogeneity makes AML a "moving target" for detection of residual disease. *Cytom. Part B Clin. Cytom.* **86**, 3–14 (2014).

57. Zhou, Y. *et al.* Pattern Associated Leukemia Immunophenotypes and Measurable Disease Detection in Acute Myeloid Leukemia or Myelodysplastic Syndrome with Mutated NPM1. (2018). doi:10.1002/cyto.b.21744

58. Comparison of leukemia-associated immunphenotype (LAIP)-based and different-from-normal(DfN)-based analysis of measurable residual disease (MRD) in patients with AML - Google zoeken. Available at: https://www.google.com/search?q=Comparison+of+leukemia-associated+immunphenotype+(LAIP)-based+and+different-from-normal+(DfN)-based+analysis+of+measurable+residual+disease+(MRD)+in+patients+with+AML&oq=Comparison+of+leukemia-associated+immunphenotype+(LAIP)-based+and+different-from-normal+(DfN)-based+analysis+of+measurable+residual+disease+(MRD)+in+patients+with+AML&aqs=chrome..69i57.688j0j7&sourceid=chrome&ie=UTF-8. (Accessed: 13th November 2020)

59. Shayegi, N. *et al.* The level of residual disease based on mutant NPM1 is an independent prognostic factor for relapse and survival in AML. *Blood* **122**, 83–92 (2013).

60. Gorello, P. *et al.* Quantitative assessment of minimal residual disease in acute myeloid leukemia carrying nucleophosmin (NPM1) gene mutations. *Leukemia* **20**, 1103–1108 (2006).

61. Grimwade, D., Vyas, P. & Freeman, S. Assessment of minimal residual disease in acute myeloid leukemia. *Curr. Opin. Oncol.* **22**, 656–63 (2010).

62. Zhou, Y. *et al.* Pre- and post-transplant quantification of measurable ('minimal') residual disease via multiparameter flow cytometry in adult acute myeloid leukemia. *Leukemia* **30**, 1456–64 (2016).

63. Brambati, C. *et al.* Droplet digital polymerase chain reaction for DNMT3A and IDH1/2 mutations to improve early detection of acute myeloid leukemia relapse after allogeneic hematopoietic stem cell transplantation. *Haematologica* **101**, e157-61 (2016).

64. Kim, T. H. *et al.* Next-generation sequencing–based posttransplant monitoring of acute myeloid leukemia identifies patients at high risk of relapse. *Blood* **132**, 1604–1613 (2018).

65. Getta, B. M. *et al.* Measurable residual disease monitoring by ngs before allogeneic hematopoietic cell transplantation in AML. *Blood* **23**, 1703–1713 (2005).

66. Guolo, F. *et al.* Combining Flow Cytometry and Molecular Assessment Improves the Prognostic Value of Pre-Transplant Minimal Residual Disease in Acute Myeloid Leukemia. *Blood* **128**, (2016).

67. Ouyang, J. *et al.* Comparison of Multiparameter Flow Cytometry Immunophenotypic Analysis and Quantitative RT-PCR for the Detection of Minimal Residual Disease of Core Binding Factor Acute Myeloid Leukemia. *Am. J. Clin. Pathol.* **145**, 769–777 (2016).

68. Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* **352**, 254–266 (2005).

69. Schlenk, R. F. *et al.* Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N. Engl. J. Med.* **358**, 1909–1918 (2008).

70. Ivey, A. *et al.* Assessment of Minimal Residual Disease in Standard-Risk AML. *N. Engl. J. Med.* **374**, 422–433 (2016).

71. Balsat, M. *et al.* Postinduction minimal residual disease predicts outcome and benefit from allogeneic stem cell transplantation in acute myeloid leukemia with NPM1 mutation: A study by the acute leukemia French association group. *J. Clin. Oncol.* **35**, 185–193 (2017).

72. Krönke, J. *et al.* Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: A study from the German-Austrian acute myeloid leukemia study group. in *Journal of Clinical Oncology* **29**, 2709–2716 (J Clin Oncol, 2011).

73. Dillon, R. *et al.* Molecular MRD status and outcome after transplantation in NPM1-mutated AML. *Blood* **135**, 680–688 (2020).

74. Chou, W. C. *et al.* Clinical implications of minimal residual disease monitoring by quantitative polymerase chain reaction in acute myeloid leukemia patients bearing nucleophosmin (NPM1) mutations. *Leukemia* **21**, 998–1004 (2007).

75. Schnittger, S. *et al.* Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML. *Blood* **114**, 2220–2231 (2009).

76. Forghieri, F., Comoli, P., Marasca, R., Potenza, L. & Luppi, M. Minimal/measurable residual disease monitoring in NPM1-mutated acute myeloid leukemia: A clinical viewpoint and perspectives. *International Journal of Molecular Sciences* **19**, (2018).

77. Gao, M. G. *et al.* The predictive value of minimal residual disease when facing the inconsistent results detected by real-time quantitative PCR and flow cytometry in NPM1-mutated acute myeloid leukemia. *Ann. Hematol.* **99**, 73–82 (2020).

78. Du Pisani, L. A. & Shires, K. Development of a flow cytometric method to detect the presence of mutated nucleophosmin 1 in acute myeloid leukemia. *Hematol. Oncol. Stem Cell Ther.* **8**, 106–114 (2015).

79. Oelschlaegel, U. *et al.* Rapid flow cytometric detection of aberrant cytoplasmic localization of nucleophosmin (NPMc) indicating mutant NPM1 gene in acute myeloid leukemia. *Leukemia* **24**, 1813–1816 (2010).

80. Gruszka, A. M. *et al.* A monoclonal antibody against mutated nucleophosmin 1 for the molecular diagnosis of acute myeloid leukemias. *Blood* **116**, 2096–2102 (2010).

81. Tan, S. *et al. Monoclonal Antibodies against Nucleophosmin Mutants: Potentials for the Detection of Acute Myeloid Leukemia. Int. J. Med. Sci* **8**, (2011).

82. El-Gamal, R. A. E. R., Hashem, A. E. S., Habashy, D. M., Abou Elwafa, M. A. Z. & Boshnak, N. H. Flow cytometry in detection of Nucleophosmin 1 mutation in acute myeloid leukemia patients: A reproducible tertiary hospital experience. *Int. J. Lab. Hematol.* (2020). doi:10.1111/ijlh.13317

83. Levine, R. L. & Valk, P. J. M. Next-generation sequencing in the diagnosis and minimal residual disease assessment of acute myeloid leukemia. *Haematologica* **104**, 868–871 (2019).

84. Getta, B. M. *et al.* Multicolor Flow Cytometry and Multi-Gene Next Generation Sequencing Are Complimentary and Highly Predictive for Relapse in Acute Myeloid Leukemia Following Allogeneic Hematopoietic Stem Cell Transplant. *Present. 58th Am. Soc. Hematol. Annu. Meet. Expo.* **Dec**, San Diego, CA (2016).

85. Oliva, S. *et al.* Multiparameter flow cytometry (MFC) and next generation sequencing (NGS) for minimal residual disease (MRD) evaluation: Results of the FORTE trial in newly diagnosed multiple myeloma (MM). *J. Clin. Oncol.* **38**, 8533–8533 (2020).

86. Kriegsmann, K. *et al.* Comparison of ngs and mfc methods: Key metrics in multiple myeloma mrd assessment. *Cancers (Basel).* **12**, 1–15 (2020).

87. Kumar, S. *et al.* International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *The Lancet Oncology* **17**, e328–e346 (2016).

88. Estey, E. & Gale, R. P. How good are we at predicting the fate of someone with acute myeloid leukaemia? *Leukemia* **31**, 1255–1258 (2017).

89. Papaemmanuil, E. *et al*. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).

90. Sonneveld, P. *POINT-COUNTERPOINT COUNTERPOINT Should minimal residual disease negativity not be the end point of myeloma therapy?*

91. Cloos, J., Ossenkoppele, G. J. & Dillon, R. Minimal residual disease and stem cell transplantation outcomes. *Hematol. (United States)* **2019**, 617–625 (2019).

92. Cloos, J., Ossenkoppele, G. J. & Dillon, R. Minimal residual disease and stem cell transplantation outcomes. *Hematol. (United States)* **2019**, 617–625 (2019).

93. Juul-Dam, K. L. *et al*. Measurable residual disease assessment by qPCR in peripheral blood is an informative tool for disease surveillance in childhood acute myeloid leukaemia. *Br. J. Haematol.* **190**, 198–208 (2020).

94. Godwin, C. D. *et al*. Acute myeloid leukemia measurable residual disease detection by flow cytometry in peripheral blood vs bone marrow. *Blood* **137**, 569–572 (2021).

95. Rücker, F. G. *et al*. Measurable residual disease monitoring in acute myeloid leukemia with t(8;21)(q22;q22.1): Results from the AML Study Group. *Blood* **134**, 1608–1618 (2019).

96. Ngai, L. L., Kelder, A., Janssen, J. J. W. M., Ossenkoppele, G. J. & Cloos, J. MRD Tailored Therapy in AML: What We Have Learned So Far. *Front. Oncol.* **10**, (2021).

CHAPTER 10

SUMMARY & NEDERLANDSE SAMENVATTING

SUMMARY

Nearly all patients with acute myeloid leukemia (AML) receive intense chemotherapy, followed by a consolidation therapy which can be an additional cycle of chemotherapy, autologous stem cell transplantation or allogeneic stem cell transplantation. In this order, anti-leukemic efficacy increases together with toxicity.

While, fortunately, most patients achieve complete remission (CR) (leukemia cells <5% in bone marrow as determined by microscopy), unfortunately, 40-50% of patients experience a relapse. Patients who relapse have a dismal prognosis since the relapse is difficult to eradicate. A correct understanding of the risk to relapse is vital for selecting the correct therapy intensity. Risk stratification at diagnosis is based on factors such as age, white blood cell (WBC) count and genetic (mutations and cytogenetic aberrations) characteristics[1]. This risk assessment at diagnosis does not suffice for an accurate estimation of patients that relapse, therefore, more specific and sensitive methods (both by flow cytometry and molecular techniques) are widely used to assess possible residual disease during and after therapy. When this residual disease (termed measurable residual disease or minimal residual disease, MRD) is present above a critical level, patients have a higher change to experience a relapse. **Chapter 1** provides a general introduction regarding AML and its clinical features. Hereafter, the MRD and the different methods to measure MRD are introduced.

In this thesis we measured the frequency of residual leukemia cells based on the aberrant expression of cell surface markers (using multiparameter flow cytometry (MFC)) in the bone marrow of patients treated with chemotherapy. The MFC-MRD assay itself has been improved by our studies by investigating specific cell populations within the residual cells, which are important for leukemia progression. This includes besides the progenitor cells (fast proliferating cells, which populate the leukemia) also the leukemia stem cells (LSCs), which are infrequent undifferentiated cells that can initiate leukemia relapse). In **chapter 2** we provide an overview of several aspects of LSCs in AML and how these can be detected by MFC using expression of aberrant membrane surface markers.

Our studies show that a high frequency of CD34+CD38- LSCs at diagnosis is associated with poor prognosis of the patient, independent of other risk factors. We have shown this for the adult patients (18-65 years) (chapter 3) but also for the younger patient group (<20 years) (chapter 4). **Chapter 3** explored flow cytometric results acquired in the HOVON102 clinical trial. Using this, we showed that the LSC percentage after therapy (LSC-MRD) had prognostic value for outcome. In addition, we evaluated the prognostic impact of the combination LSC-MRD, and MFC-MRD. Results suggested that patients with MFC-MRD^{positive}/LSC-MRD^{positive} leukemia, should be considered as poor/very poor risk patients, independent of risk category (even for NPM1^{mutant} and NPM1$^{wild\ type}$ patients) and should preferably be treated as such. **Chapter 4** shows that approaches detecting stem cells (as defined in adult AML patients) can be applied in pediatric and younger AML patients.

High frequencies of LSCs similarly associate with poor outcome. Furthermore, absence of malignant CD34+ cells correspond with a better prognosis.

With the demonstration of the importance of MFC-MRD and LSC-MRD detection, correct performance and utilization of these assays is of utmost importance. Chapters 5 and 6 aimed to provide more insights with regards to this aspect. In **chapter 5** we evaluated the technical and analytical feasibility of a previously designed eight-color LSC tube that allows specific and sensitive detection of LSC. We created an elaborate protocol describing the set-up of machines, exact usage of the tube and analysis, to aid in correct implementation and standardization of the technique. The study was conducted in several large research centers in Europe and the United States, which had extensive flow cytometry experience, but not with the assessed LSC assay. We showed that limited training led to highly concordant results, allowing other centers to independently validate the clinical utility of LSC testing in AML.

Using a mathematical model describing all seemingly important variables, we illustrate how MFC-MRD is calculated in **chapter 6**. In most studies, MRD is calculated as the percentage of cells carrying an aberrant immunophenotype of all white blood cells measured. Retrospective analysis, using flow cytometric results acquired in the HOVON102 (chapter 3), illustrated that the prognostic impact originates from the balance between AML and normal progenitors, instead of the total leukemic load. With this so-called PM-MRD assay, we gained insight in the components of the MRD assay and when PM-MRD >10%, the patients have a dismal prognosis even when $MRD^{negative}$. We will retrospectively explore the value of integrating this PM-MRD in the MRD analysis in the upcoming HOVON study cohorts.

Next-generation sequencing (NGS) is a rapidly improving technology whose application to the monitoring of MRD is an active area of research. **Chapter 7** shows data of this application in measuring MRD in AML patients. Here, we showed that in 430 out of the 482 patients (89%) at least one mutation was detected, which made them suitable for molecular MRD detection. In 51.4% of the patients, mutations were present during complete remission in various frequencies (ranging between 0.02-47%). Correlating presence of mutation to outcome parameters revealed that targeted sequencing-based detection of molecular MRD during complete remission was associated with an increased risk of relapse or death. However, the risk of relapse and/or death was not influenced by persistence of specific mutations that are associated with age-related clonal hematopoiesis. Furthermore, in a combined MRD analysis of our MFC-MRD data with the molecular MRD data on the same samples, we showed that these techniques were highly complementary.

In **chapter 8** we evaluated the use of MFC-MRD and LSC-MRD as measure of disease response for evaluating new therapies, improving the efficiency and speed of clinical trials and regulatory approval. Within the HOVON102 study, the drug clofarabine was added to induction chemotherapy in half of the patients. Within the subgroup of patients with the ELN intermediate I risk profile, addition of clofarabine was beneficial for outcome. Evaluation of the prospectively produced MFC-MRD and LSC-MRD

results revealed lower frequencies in patients with clofarabine compared to patients in the standard arm. While patient numbers are limited, and further studies in larger cohorts with qualified MRD methods are needed, this study supports the possibility of using MRD as surrogate endpoint.

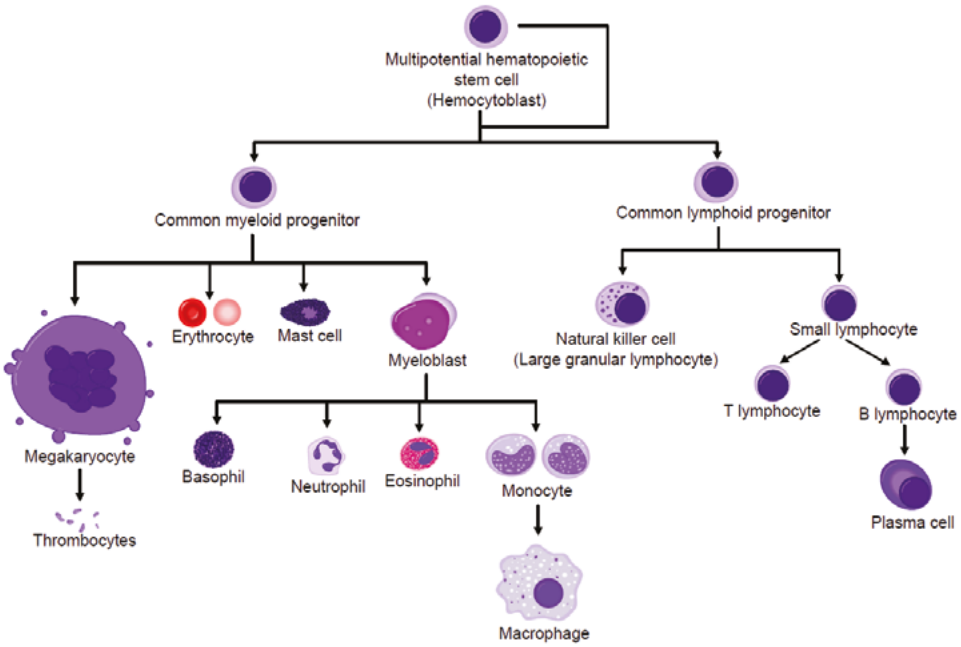
Based on findings like those presented in this thesis, MRD is currently assessed with flow cytometry, LSC flow cytometry or molecular techniques in order to refine accurate risk assessment. As a consequence of all techniques having its own principles, they all have their own limitations. In **chapter 9** we reviewed the current state-of-the art MRD techniques and evaluated their prognostic value as single and combined methods in patients where all techniques were simultaneously performed. Although combinational approaches appear to be of high importance in relapse prediction in AML patients, further refinement of the techniques to further reduce the amount of false-negative MRD results is of high importance. Furthermore, as MRD assays are essential to support clinicians to provide optimal treatment choices for patients, their results and reports should be of the best achievable quality. Approaches for refinements of the techniques, the importance of correct implementation, and thus standardization of these assay are further discussed in this last chapter.

NEDERLANDSE SAMENVATTING

MEETBARE RESTZIEKTE EN LEUKEMISCHE STAMCELLEN IN ACUTE MYELOIDE LEUKEMIE

In het bloed bevinden zich verschillende soorten cellen met ieder hun eigen functie. De verschillende bloedcellen worden gevormd in het beenmerg door het proces dat we kennen als hematopoëse. Vanuit zogenoemde hematopoëtische stamcellen (HSC) ontstaan nieuwe cellen (prolifereren). Deze nieuwe jonge cellen rijpen vervolgens uit tot verschillende cellen met verschillende functies (differentiatie). In gezonde personen ontstaat op deze manier een mix van rode bloedcellen (voor het zuurstoftransport), bloedplaatjes (nodig voor de bloedstolling) en witte bloedcellen (leukocyten genoemd, nodig voor de afweer). De leukocyten zijn onder te verdelen in verschillende klassen: de myelocyten en lymfocyten. De myelocyten bestaan uit granulocyten en monocyten, die gespecialiseerd zijn in het insluiten en het onschadelijk maken van ziekteverwekkers. De lymfocyten bestaan uit B-, T- en NK-cellen. Deze cellen rijpen uit in lymfoïde organen (zoals de milt en de lymfeklieren) en vallen cellen die besmet zijn met bacteriën en virussen aan om ze uit te schakelen.

Leukemie is een vorm van kanker waarbij het beenmerg te veel witte bloedcellen maakt. Bij acute leukemie is er sprake van een snelle woekering en ophoping van onrijpe leukocyten (blasten genoemd) in het beenmerg en ontstaan klachten meestal al binnen



Figuur 1. De bloedcellen ontstaan uit de hematopoïetische stamcel in het beenmerg.

enkele weken na de start van het woekeren van de blasten. Een tekort aan rode bloedcellen lijdt tot bloedarmoede, waarbij een patiënt last kan krijgen van vermoeidheid en bleekheid. Door een tekort aan bloedplaatjes bloeden wondjes sneller en langer waarbij patiënten blauwe plekken of bloedend tandvlees kan krijgen. Door een tekort aan normale gezonde witte bloedcellen ontstaat er een groter risico op infecties.

Dit proefschrift beschrijft onderzoek uitgevoerd in patiënten met acute myeloïde leukemie (AML). Iedereen kan AML krijgen, maar treft vooral ouderen. In Nederland wordt bij ongeveer 600 volwassenen per jaar de ziekte vastgesteld, tegenover 25 kinderen. Patiënten met AML worden behandeld met intensieve chemotherapie om snel de grote hoeveelheid aan leukemische cellen te doden. Vervolgens krijgen de meesten patiënten een vervolgtherapie (consolidatietherapie) welke kan bestaan uit wederom chemotherapie, maar ook uit autologe of allogene stamceltransplantatie. In deze volgorde neemt niet alleen de effectiviteit van de behandeling toe, maar ook de toxiciteit.

Na twee kuren chemotherapie wordt onderzocht hoe goed deze behandeling heeft gewerkt. De behandeld arts (de hematoloog) zal een beenmergpunctie nemen, om de cellen te laten onderzoeken onder de microscoop. Als de meeste leukemische cellen weg zijn (<5% blasten in het beenmerg) is de therapie succesvol geweest, en is de patiënt in complete remissie (CR). Het lijkt dan dat alsof de ziekte is verdwenen. Hoewel de meeste patiënten wel in CR komen, keert in vele gevallen de ziekte terug (recidief). Deze is moeilijker te behandelen. Afstemming van de consolidatietherapie op het risico dat een patiënt loopt op een recidief, vindt al jaren plaats op basis van bepaalde risicofactoren vastgesteld bij diagnose, zoals leeftijd, maar ook genetische factoren. Dit is echter niet afdoende om een goede inschatting te maken van het risico op een recidief. Daarom wordt tegenwoordig tijdens en na behandeling vastgesteld of de ziekte compleet weg is. De aanwezigheid van zogeheten minimale (of meetbare) restziekte (aangeduid als MRD, measurable residual disease) is geassocieerd met een groter risico op het ontwikkelen van een recidief.

De leukemiecellen in het beenmerg zijn onder andere te monitoren met behulp van microscopie en flow cytometrie. In dit proefschrift wordt vooral de flow cytometrie bestudeerd, waarbij wordt gekeken naar de aan- of afwezigheid van eiwitten op de het celoppervlak. Ieder type bloed- en beenmergcel heeft een unieke combinatie van eiwitten op zijn oppervlak, die zichtbaar gemaakt kunnen worden met antistoffen met daaraan een fluorescente marker (een soort lichtgevend vlaggetje). Elke antistof kan binden aan één specifiek eiwit. Het mengsel van cellen en de daaraan gebonden antistoffen worden door een machine gespoeld (de flowcytometer) waarin de fluorescente markers per enkele cel zichtbaar worden gemaakt door detectie van laserlicht.

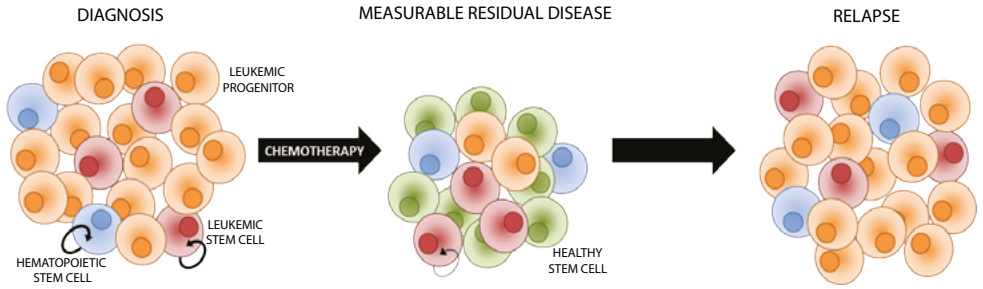
Door de profielen van de eiwitten die tot expressie komen in kaart te brengen, kan onderscheid gemaakt worden tussen verschillenden typen cellen welke aanwezig zijn in het beenmerg of bloed. Ongeveer 90% van de AML-patiënten heeft leukemische cellen met een afwijkend patroon van oppervlakte eiwitten. De combinatie van eiwitten op leukemiecellen vormen dus een unieke 'barcode' waardoor ze zich onderscheiden van

normale cellen. Deze 'barcode' noemen we een 'leukemie geassocieerd immunofenotype' (in het Engels: leukemia associated immunophenotype, afgekort tot LAIP). Door deze LAIPs op de leukemiecellen vast te stellen bij diagnose, zijn we in staat om heel specifiek op zoek te gaan naar aanwezigheid van leukemiecellen in beenmerg na chemotherapie. Middels flowcytometrie kunnen we de aanwezigheid van restziekte onderzoeken in patiënten die microscopisch in complete remissie zijn. Flowcytometrie kan 1 leukemische cel onderscheiden binnen 100.000-1.000.000 normale cellen terwijl dat voor microscopie maar 1 op de 100 is.

Door een specifieke combinatie van antistoffen te gebruiken kan met flowcytometrie ook onderscheid worden gemaakt tussen normale hematopoietische voorloper cellen (stamcellen; HSC) en leukemische stamcellen (LSC). Omdat deze LSC over het algemeen weinig celdelingen maken en verankerd zitten in hun zogenaamde stamcelniche, zijn ze vaak ongevoelig voor chemotherapie. Net als HSCs, kunnen LSCs vernieuwen en differentiëren. Ze zouden daarom een belangrijke rol kunnen spelen in het ontstaan van, en eveneens het terugkeren van, de ziekte. Patiënten waar na chemotherapie nog LSC worden gevonden, hebben een hogere kans op het ontwikkelen van een recidief.

ONDERZOEK

Van patiënten met ogenschijnlijk succesvolle chemotherapie, bleek dat degenen, waarbij nog restziekte aantoonbaar was, een hoger risico hadden op het terugkrijgen van de ziekte, dan patiënten waar geen MRD kon worden aangetoond middels flowcytometrie. Patiënten waarbij van alle gemeten witte bloedcellen, nog 0.1% de LAIP bevat zoals bij diagnose, noemen we MRD-positief. Dit proefschrift heeft onderzoek gedaan hoe de meting van MRD, en de identificatie van LSC verbeterd en gestandaardiseerd kan worden. We hebben



Figuur 2. Diagnose, restziekte en recidief en de rol van leukemische stamcellen hierin. Ten tijde van de diagnose zit het beenmerg van de patiënt bomvol met leukemische cellen (oranje), normale hematopoietische stamcellen (HSC; blauw) en leukemische stamcellen (LSC; rood). Gelukkig komen de meeste patiënten in complete remissie na behandeling met chemotherapie. In een deel van de patiënten blijft restziekte achter (welke niet met de miscroscoop kan worden herkend). Omdat in deze minimale restziekte (MRD) ook nog stamcellen aanwezig zijn, kan de ziekte opnieuw uitgroeien tot een recidief. Overgenomen van Canales et al.[2]

onderzocht hoe de resultaten van deze metingen de arts kunnen helpen om de juiste risicoschatting te hebben voor het ontwikkelen van een recidief en een onderbouwde keuze te maken tussen de verschillende opties voor de consolidatietherapie. De afweging is dat de behandeling zwaar genoeg moet zijn zodat de resterende leukemie goed aangevallen wordt, maar met zo min mogelijk toxiciteit (en een zo laag mogelijke kans om te overlijden aan de therapie zelf).

RESULTATEN

In **hoofdstuk 1** wordt een overzicht gegeven van de klinische kenmerken en de behandeling van AML met daarbij de gebruikte risicogroep indeling met een gunstig, gemiddeld, slecht of zeer slecht risicoprofiel op basis van moleculaire en cytogenetische afwijkingen, het aantal witte bloedcellen bij diagnose en de therapierespons. Hoewel er een duidelijk verband is met de prognose, kunnen ook patiënten uit de groep met een gunstig risico een recidief krijgen en overlijden. Het belang van het aantonen van MRD, en LSC, en de daarvoor beschikbare technieken wordt besproken. **Hoofdstuk 2** geeft verder uitleg over de LSC en beschrijft hun rol in AML.

Hoofdstuk 3 beschrijft de resultaten van de prospectieve HOVON/SAKK 102 studie. Met behulp van flowcytometrie is de frequentie van LSC en MRD bij 242 volwassen patiënten in morfologisch CR na chemotherapie bepaald. MRD-negatieve patiënten (<0.1% LAIP positieve cellen) ontwikkelden minder vaak recidieven en hadden een betere overleving dan MRD-positieve patiënten. LSC-negatieve patiënten hadden een beter verloop van de ziekte dan LSC-positieve patiënten. Ook wanneer rekening wordt gehouden met andere risicofactoren voor het ontwikkelen van een recidief (zoals leeftijd en de snelheid van het succes van de chemotherapie) blijkt het meten van deze LSC na chemotherapie een belangrijke meerwaarde voor het correct voorspellen van de kans op het ontwikkelen van een recidief en de kans op overleving in AML-patiënten. Opvallend resultaat was de zeer slechte prognose van patiënten die zowel MRD als LSC positief waren, ongeacht de vooraf bepaalde risicogroep. Een belangrijke toevoeging op de kennis over LSC is beschreven in **hoofdstuk 4**. Hier laten we zien dat dezelfde LSC-meting kan worden gebruikt op het beenmerg van kinderen met AML. Van de in totaal 86 onderzochte patiënten (≤20 jaar) kon in 80 beenmergen LSC worden onderzocht.

Omdat de aanwezigheid van MRD en leukemische stamcellen een belangrijke factor zijn voor het voorspellen van de recidief- en overlevingskans van patiënten, is het waarschijnlijk dat in de nabije toekomst de uitkomsten van deze metingen indicatief zijn voor de keuze van de consolidatietherapie. Het is hierdoor van belang dat de implementatie van de flowcytometrische technieken gestandaardiseerd gebeurt. De meting van LSC is tijdrovend en bewerkelijk. Dat komt ten eerste doordat LSC in lage hoeveelheden aanwezig zijn in het beenmerg, en er dus veel beenmergcellen moeten worden gemeten voor een betrouwbare uitslag. Ten tweede is niet elke LSC hetzelfde.

Hoewel we ons onderzoek focussen op CD34+CD38- zijn er verscheidende antistoffen nodig gebleken om in veel patiënten LSC aan te kunnen tonen. In eerder uitgevoerd onderzoek zijn 13 noodzakelijke antistoffen gerangschikt in één mix ('antistofpanel'). In **hoofdstuk 5** beschrijven we hoe deze mix, met bijhorende protocollen, is geïmplementeerd in verschillende grote leukemie-onderzoekslaboratoria in Europa en de Verenigde Staten. De verschillende centra, met ervaring in flowcytometrie voor AML, maar onbekend met stamcelmetingen, waren na beperkte training in staat vergelijkbare resultaten te genereren in tien opgestuurde beenmergmonsters. Deze centra kunnen nu in hun eigen klinische studies de prognostische waarde van LSC-metingen (zoals in hoofdstuk 3) valideren.

In **hoofdstuk 6** hebben we de flowcytometrie methode verbeterd door celpopulaties te bestuderen die belangrijk zijn voor de progressie van de leukemie. In de huidige MRD-metingen worden alle witte bloedcellen geteld. Wanneer meer dan 0.1% hiervan de leukemische 'barcode' (LAIP) bevat, worden patiënten geclassificeerd als MRD-positief. Omdat er, door allerlei niet-biologische redenen, variatie kan zitten in het totale aantal witte bloedcellen is geanalyseerd of de balans tussen normale jonge cellen en kwaadaardige jonge cellen, een betere voorspeller is van het risico op het ontwikkelen van een recidief en de overlevingskans. Door het mathematisch uitschrijven van de huidige MRD-rekenmethode is het nut van de verschillende factoren getest. We laten zien dat deze manier van berekenen ook patiënten identificeert met hoge kans op een recidief en mogelijk een goede toevoeging is voor de huidige MRD.

Het proces van hematopoëse wordt nauwgezet gecontroleerd. Toch kunnen er fouten optreden. Tijdens de uitrijping van stamcellen naar meer gedifferentieerde bloedcellen kunnen verschillende genetische fouten (mutaties) in het DNA ontstaan. De meeste mutaties bevinden zich uitsluitend in leukemische cellen. In **hoofdstuk 7** wordt beschreven hoe één van de recentste technieken (next-generation sequencing) kan worden gebruikt om deze mutaties te monitoren en om restziekte aan te tonen. In 482 patiënten zijn 54 verschillende mutaties gemeten en in 430 patiënten zijn mutaties gevonden. Na chemotherapie is gekeken of deze mutatie nog terug te vinden is. Wanneer dit het geval is, is het aannemelijk dat er nog leukemische cellen (MRD) aanwezig zijn. Met deze methode kon in 51.4% van de patiënten restziekte worden aangetoond. Patiënten met aanwezigheid van deze zogeheten moleculaire MRD bleken een significant grotere kans te hebben op het ontwikkelen van een recidief, en overlijden, in vergelijking met patiënten met geen moleculaire MRD. In een gecombineerde MRD-analyse van flowcytometrie gegevens met de moleculaire MRD-gegevens, laat hoofdstuk 7 zien dat beide benaderingen relevante gegevens opleveren waarbij de resultaten niet altijd overeenkomen. De twee technieken lijken juist eerder aanvullend op elkaar te zijn en met name patiënten die $MRD^{positief}$ zijn met beide technieken, hebben een zeer slechte prognose.

In **hoofdstuk 8** laten we zien hoe de percentages van LSC en MRD na chemotherapie met een experimenteel middel lager zijn dan de percentages LSC en MRD in patiënten met de standaard chemotherapie. De patiënten die behandeld werden met de nieuwe chemotherapie hadden een betere overleving dan patiënten zonder deze toevoeging.

Hoewel de patiëntengroep klein is, is deze observatie de eerste die suggereert dat het meten van de MRD en LSC na therapie de effectiviteit van de therapie weerspiegelt.

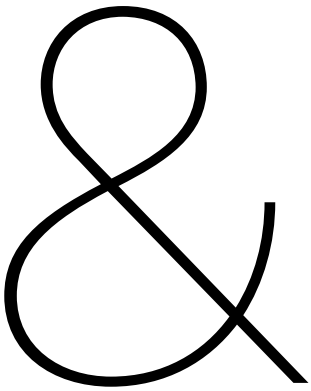
Omdat er nu meerdere technieken in staat blijken, patiënten met een hogere kans op recidief of overlijden te identificeren, bediscussiëren we in **hoofdstuk 9** de voor- en nadelen van deze verschillende technieken, hoe ze mogelijk elkaar kunnen ondersteunen voor een betere en completere identificatie van slechtere patiënten, en hoe iedere techniek zijn eigen verbeterpunten kent. Tot slot sluit dit hoofdstuk af met een samenvatting en de verwachtingen over de rol van de detectie van LSC en MRD voor de behandeling van toekomstige AML-patiënten

ANNEXES

LIST OF PUBLICATIONS

CURRICULUM VITAE

ACKNOWLEDGEMENTS



LIST OF PUBLICATIONS

Hanekamp D, Tettero JM, Ossenkoppele GJ, Kelder A, Cloos J, Schuurhuis GJ. AML/Normal Progenitor Balance Instead of Total Tumor Load (MRD) Accounts for Prognostic Impact of Flowcytometric Residual Disease in AML. Cancers (Basel). 2021 May 26;13(11):2597.

Hanekamp D*, Ngai LL*, Janssen JJWM, van de Loosdrecht A, Ossenkoppele GJ, Cloos J. Early assessment of clofarabine effectiveness based on measurable residual disease, including AML stem cells. Blood. 2021 Mar 25;137(12):1694-1697.

Hanekamp D, Snel AN, Kelder A, Scholten WJ, Khan N, Metzner M, Irno-Consalvo M, Sugita M, de Jong A, Oude Alink S, Eidhof H, Wilhelm M, Feuring-Buske M, Slomp J, van der Velden VHJ, Sonneveld E, Guzman M, Roboz GJ, Buccisano F, Vyas P, Freeman S, Bachas C, Ossenkoppele GJ, Schuurhuis GJ, Cloos J. Applicability and reproducibility of acute myeloid leukaemia stem cell assessment in a multi-centre setting. Br J Haematol. 2020 Sep;190(6):891-900.

Hanekamp D, Bachas C, van de Loosdrecht A, Ossenkoppele G, Cloos J. Re: Myeloblasts in normal bone marrows expressing leukaemia-associated immunophenotypes. Pathology. 2020 Feb;52(2):289-291.

Depreter B, Weening KE, Vandepoele K, Essand M, De Moerloose B, Themeli M, Cloos J, **Hanekamp D**, Moors I, D'hont I, Denys B, Uyttebroeck A, Van Damme A, Dedeken L, Snauwaert S, Goetgeluk G, De Munter S, Kerre T, Vandekerckhove B, Lammens T, Philippé J. TARP is an immunotherapeutic target in acute myeloid leukemia expressed in the leukemic stem cell compartment. Haematologica. 2020 May;105(5):1306-1316.

Zeijlemaker W, Grob T*, Meijer R*, **Hanekamp D***, Kelder A, Carbaat-Ham JC, Oussoren-Brockhoff YJM, Snel AN, Veldhuizen D, Scholten WJ, Maertens J, Breems DA, Pabst T, Manz MG, van der Velden VHJ, Slomp J, Preijers F, Cloos J, van de Loosdrecht AA, Löwenberg B, Valk PJM, Jongen-Lavrencic M, Ossenkoppele GJ, Schuurhuis GJ. CD34+CD38- leukemic stem cell frequency to predict outcome in acute myeloid leukemia. Leukemia. 2019 May;33(5):1102-1112.

Jongen-Lavrencic M*, Grob T*, **Hanekamp D***, Kavelaars FG, Al Hinai A, Zeilemaker A, Erpelinck-Verschueren CAJ, Gradowska PL, Meijer R, Cloos J, Biemond BJ, Graux C, van Marwijk Kooy M, Manz MG, Pabst T, Passweg JR, Havelange V, Ossenkoppele GJ, Sanders MA, Schuurhuis GJ, Löwenberg B, Valk PJM. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. N Engl J Med. 2018 Mar 29;378(13):1189-1199.

Cloos J, Harris JR, Janssen JJWM, Kelder A, Huang F, Sijm G, Vonk M, Snel AN, Scheick JR, Scholten WJ, Carbaat-Ham J, Veldhuizen D, **Hanekamp D**, Oussoren-Brockhoff YJM, Kaspers GJL, Schuurhuis GJ, Sasser AK, Ossenkoppele G. Comprehensive Protocol to Sample and Process Bone Marrow for Measuring Measurable Residual Disease and Leukemic Stem Cells in Acute Myeloid Leukemia. J Vis Exp. 2018 Mar 5;(133):56386.

Hanekamp D*, Denys B*, Kaspers GJL, Te Marvelde JG, Schuurhuis GJ, De Haas V, De Moerloose B, de Bont ES, Zwaan CM, de Jong A, Depreter B, Lammens T, Philippé J, Cloos J, van der Velden VHJ. Leukaemic stem cell load at diagnosis predicts the development of relapse in young acute myeloid leukaemia patients. Br J Haematol. 2018 Nov;183(3):512-516.

Hanekamp D, Cloos J, Schuurhuis GJ. Leukemic stem cells: identification and clinical application. Int J Hematol. 2017 May;105(5):549-557.

Goverse G, Molenaar R, Macia L, Tan J, Erkelens MN, Konijn T, Knippenberg M, Cook EC, **Hanekamp D**, Veldhoen M, Hartog A, Roeselers G, Mackay CR, Mebius RE. Diet-Derived Short Chain Fatty Acids Stimulate Intestinal Epithelial Cells to Induce Mucosal Tolerogenic Dendritic Cells. J Immunol. 2017 Mar 1;198(5):2172-2181.

Lin F, de Gooijer MC, **Hanekamp D**, Chandrasekaran G, Buil LC, Thota N, Sparidans RW, Beijnen JH, Würdinger T, van Tellingen O. PI3K-mTOR Pathway Inhibition Exhibits Efficacy Against High-grade Glioma in Clinically Relevant Mouse Models. Clin Cancer Res. 2017 Mar 1;23(5):1286-1298.

Kersten B, Valkering M, Wouters R, van Amerongen R, **Hanekamp D**, Kwidama Z, Valk P, Ossenkoppele G, Zeijlemaker W, Kaspers G, Cloos J, Schuurhuis GJ. CD45RA, a specific marker for leukaemia stem cell sub-populations in acute myeloid leukaemia. Br J Haematol. 2016 Apr;173(2):219-35.

Lin F, de Gooijer MC, **Hanekamp D**, Brandsma D, Beijnen JH, van Tellingen O. Targeting core (mutated) pathways of high-grade gliomas: challenges of intrinsic resistance and drug efflux. CNS Oncol. 2013 May;2(3):271-88.

CURRICULUM VITAE

Diana Hanekamp was born on the 19th of October, 1990 in Kampen, the Netherlands. After finishing high school, she started the bachelor Biology and Medical Laboratory Research in 2008 at the *Saxion University of Applied Science in Deventer*.

After finishing this bachelor in 2012, she started with her master Biomolecular Sciences at the *VU University Amsterdam*. In the second year of her master, she participated in an international exchange program called *CanSys* (for *can*cer and *sys*tem biology). For this program she moved to Luxembourg for a literate study at the *Université du Luxembourg* and then to Buffalo (New York, USA) to participate in the master Natural Sciences at the *University of Buffalo, State University of New York*. In Buffalo, she did her second Master internship at *Roswell Park Cancer Institute*, where she started working on acute myeloid leukemia, in the group of Dr. Eunice Wang. Here she performed research on the role of bone marrow microenvironment to leukemia cell survival, expansion and therapy resistance.

In 2014 she graduated for her masters and started her research underlying this thesis as a PhD student at the department of Hematology at the Cancer Center Amsterdam at the *VU University Medical Center*.

In 2020 Diana started as a post-doctoral fellow on a collaborative project between the A*msterdam UMC* and the *Erasmus MC*, under supervision of prof.dr. Jacqueline Cloos and dr. Peter Valk, where knowledge on both molecular measurable residual disease and immunophenotypic measurable residual disease is combined.

ACKNOWLEDGEMENTS

Met veel plezier en trots kan ik zeggen dat mijn boekje voltooid is. Zoals bij de meeste belangrijke mijlpalen in iemands leven, was de uitvoering van mijn promotieonderzoek en het schrijven van dit proefschrift absoluut onmogelijk geweest zonder de hulp en steun van mensen in mijn directe en indirecte omgeving. Ik wil iedereen die mij de afgelopen jaren van wijze raad of daad heeft voorzien via deze weg dan ook ontzettend bedanken. Ik heb beloofd het kort te houden, maar een aantal mensen wil ik echter in het bijzonder noemen.

Heel veel dank ben ik verschuldigd aan mijn promotor Jacqueline. Ik ben je eerste promovenda maar er zullen er zeker nog vele volgen. Je warme, menselijke en benaderbare persoonlijkheid valt misschien het beste te omvatten met het beschrijven van mijn sollicitatiegesprek. Terwijl ik daar zat te praten over mijn eerdere projecten en waarom ik dacht dat een promotie bij de hematologie de beste stap was, vroeg jij naar mijn hobby's en interesses; die stonden namelijk niet op mijn CV. Je bent meer dan een mentor geweest gedurende mijn PhD.

&

Beste copromotoren, beste Gert en Jeroen, jullie klinische input, verrassende vragen en opmerkingen waren van onschatbare waarde. Ik heb grote bewondering voor jullie werkzaamheden in de kliniek, en hoe deze, ondanks de echt bomvolle agenda's (respectievelijk zelfs na het pensioneren), toch te combineren bleken met het onderzoek.

Beste Gerrit Jan, door jou ben ik op het CCA terecht gekomen, en daar ben ik je dankbaar voor. Je hebt je altijd ingezet voor mijn wetenschappelijke maar ook persoonlijke groei. Dat je mij wilde begeleiden, zelfs na je pensioen, voelt nog altijd bijzonder.

Hooggeleerde leden van de leescommissie; prof.dr. Gert-Jan Kaspers, dr. Linda Smit, prof. dr. Michael Heuser, prof.dr. Gerwin Huls, prof.dr. Bart Biemond, prof.dr. Marc Raaijmakers, bedankt voor jullie deelname in de beoordelingscommissie van mijn proefschrift. Ik kijk ernaar uit om dit proefschrift tegenoverstaand van u allen te verdedigen.

Mijn paranimfen Jennifer en Evelyn, dank dat jullie naast mij willen staan op deze bijzondere dag. Jennifer, we zijn het samen begonnen (ik had me echt geen betere start kunnen wensen), en ik ben heel blij dat je me ook bij de afronding bij wilt staan, en hopelijk tot ver daarna. Evelyn, jij bent er al bij sinds de sollicitaties voor onze PhD's vanuit onze hostels/ motels/hotels in Amerika (of voor het raam; als we net voor de afgesproken skype-tijd de kamer al moesten verlaten maar toch wifi nodig hadden). Ik ben blij dat ik je één van mijn beste vriendinnen mag noemen, en ik ben blij dat jij er ook vandaag bij bent.

Ook mijn andere lieve vriendinnen en naaste (voormalige) collega's wil ik hier graag bedanken. Paulien, Dieke, Leoni, Renske, Renée, Margot, Noortje, Lisa, Hilma: sommigen meer op afstand, maar toch altijd dichtbij en betrokken. Dankjewel voor jullie vriendschap en aanmoediging.

Ik kan met gemak pagina's vullen met persoonlijke bedankjes voor iedereen van de afdeling hematologie van het CCA, maar ik houd het kort. Bedankt voor de fijne werksfeer en de goede jaren die ik bij jullie heb gehad. Het is de leukste verdieping van het CCA, maar ook ver daarbuiten. Beste mede-promovendi; bedankt voor de vele uren van hard werken, ouwehoeren en borrelen in kamer 4.04. Beste oud-promovendi, mogelijk nog meer bedankt voor de eerste uren van al dit in dat kantoor op de 3e. Beste MRDers; het is bijzonder om de grote vanzelfsprekendheid van al jullie werk te zien, maar jullie verdienen een bijzonder bedankje. Willemijn, Sander, Angèle, Yvonne, Dennis; de big five van het team, zonder jullie was geen van de hoofdstukken in dit proefschrift mogelijk. Beste studenten, bedankt voor jullie hulp en succes met jullie verdere carrière.

Samenwerken in onderzoek is cruciaal en ik wil via deze graag eenieder, coauteurs en overige, bijzonder hartelijk danken voor uw bijdragen. Dat de kloof tussen Rotterdam en Amsterdam toch niet zo groot blijkt, ondanks, of misschien wel dankzij voetbal, blijkt uit de meerdere co-auteurschappen van Tim en Peter. Bedankt voor de fijne (doorgaande) samenwerking. Babs, Jan, SKION, HOVON, ELN-netwerk, gelukkig bleken landsgrenzen en onze Nederlandse directheid geen onoverkoombare horde.

Lieve Rense, buurman, huisgenoot en eeuwige optimist. Jij hebt misschien enkel het vervelendste gedeelte van mijn promotie meegemaakt -ik was er althans echt niet gezelliger door-, maar jouw positieve steun was wat ik nodig had om dit proefschrift af te ronden.

Lieve (schoon)familie, lieve broers, bedankt voor jullie interesse en betrokkenheid de afgelopen jaren. Lieve pap en mam, bedankt voor alle mogelijkheden die jullie mij hebben gegeven. Niets is jullie te veel om mij (verder) te helpen. Zonder jullie onvoorwaardelijke steun en vertrouwen had ik dit niet tot een goed einde kunnen brengen. Ik ben jullie eeuwig dankbaar.