

## Reproducible measurable residual disease detection by multiparametric flow cytometry in acute myeloid leukemia

Röhnert, Maximilian A; Kramer, Michael; Schadt, Jonas; Ensel, Philipp; Thiede, Christian; Krause, Stefan W; Bücklein, Veit; Hoffmann, Jörg; Jaramillo, Sonia; Schlenk, Richard F; Röllig, Christoph; Bornhäuser, Martin; McCarthy, Nicholas; Freeman, Sylvie; Oelschlägel, Uta; von Bonin, Malte

DOI:

[10.1038/s41375-022-01647-5](https://doi.org/10.1038/s41375-022-01647-5)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Röhnert, MA, Kramer, M, Schadt, J, Ensel, P, Thiede, C, Krause, SW, Bücklein, V, Hoffmann, J, Jaramillo, S, Schlenk, RF, Röllig, C, Bornhäuser, M, McCarthy, N, Freeman, S, Oelschlägel, U & von Bonin, M 2022, 'Reproducible measurable residual disease detection by multiparametric flow cytometry in acute myeloid leukemia', *Leukemia*, vol. 36, no. 9, pp. 2208-2217. <https://doi.org/10.1038/s41375-022-01647-5>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

## ARTICLE OPEN



## ACUTE MYELOID LEUKEMIA

# Reproducible measurable residual disease detection by multiparametric flow cytometry in acute myeloid leukemia

Maximilian A. Röhnert <sup>1</sup>✉, Michael Kramer<sup>1</sup>, Jonas Schadt <sup>1</sup>, Philipp Ensel<sup>1</sup>, Christian Thiede <sup>1,2</sup>, Stefan W. Krause <sup>3</sup>, Veit Bücklein <sup>4,5</sup>, Jörg Hoffmann<sup>6</sup>, Sonia Jaramillo<sup>7</sup>, Richard F. Schlenk<sup>7,8</sup>, Christoph Röllig <sup>1</sup>, Martin Bornhäuser <sup>1,9</sup>, Nicholas McCarthy<sup>10</sup>, Sylvie Freeman <sup>10</sup>, Uta Oelschlägel<sup>1,11</sup> and Malte von Bonin <sup>1,11</sup>

© The Author(s) 2022

Measurable residual disease (MRD) detected by multiparametric flow cytometry (MFC) is associated with unfavorable outcome in patients with AML. A simple, broadly applicable eight-color panel was implemented and analyzed utilizing a hierarchical gating strategy with fixed gates to develop a clear-cut LAIP-based DfN approach. In total, 32 subpopulations with aberrant phenotypes with/without expression of markers of immaturity were monitored in 246 AML patients after completion of induction chemotherapy. Reference values were established utilizing 90 leukemia-free controls. Overall, 73% of patients achieved a response by cytomorphology. In responders, the overall survival was shorter for MRD<sup>pos</sup> patients (HR 3.8,  $p = 0.006$ ). Overall survival of MRD<sup>neg</sup> non-responders was comparable to MRD<sup>neg</sup> responders. The inter-rater-reliability for MRD detection was high with a Krippendorffs  $\alpha$  of 0.860. The mean time requirement for MRD analyses at follow-up was very short with 04:31 minutes. The proposed one-tube MFC approach for detection of MRD allows a high level of standardization leading to a promising inter-observer-reliability with a fast turnover. MRD defined by this strategy provides relevant prognostic information and establishes aberrancies outside of cell populations with markers of immaturity as an independent risk feature. Our results imply that this strategy may provide the base for multicentric immunophenotypic MRD assessment.

*Leukemia*; <https://doi.org/10.1038/s41375-022-01647-5>

## INTRODUCTION

Acute myeloid leukemia (AML) is a heterogeneous disease. After undergoing intensive induction chemotherapy, about 70% of eligible patients achieve a complete remission (CR). Without further treatment, 50% of the patients relapse within 6 months [1]. Post induction therapy is stratified by relapse risk and includes chemotherapy or allogeneic hematopoietic stem cell transplantation (aHSCT). The prognosis is partially determined upfront by cytogenetic and molecular genetic aberrations [2]. Remaining leukemic cells in bone marrow (BM) with <5% blasts are called measurable residual disease (MRD) and provide additive information for tailored treatment decisions and refinement of the prognosis. MRD positivity indicates residual disease and a high probability of relapse, whereas MRD negativity characterizes deep CR with low risk of relapse. At diagnosis, in at least 80% of AML patients molecular genetic aberrations are detectable by next-generation sequencing [3]. However, only some of these

aberrations can be detected with sensitive routine assays providing clinically relevant prognostic information: *mNPM1* is present in 30% [4], *CBFB::MYH11* in 5% [5] and *RUNX1::RUNX1T1* in 7% [2, 5–7] of non-acute promyelocytic leukemia patients, respectively.

MRD monitoring by multiparametric flow cytometry (MFC) has been shown to be applicable to almost all patients [8–17]. The precise workflow varies across laboratories. Sample processing and measurement differ between institutions, and there are two distinct analysis strategies to detect leukemic cell populations: the leukemia-associated immunophenotype (LAIP) and the different from normal (DfN) approach. In the LAIP concept, one or more individual LAIP are identified at diagnosis and tracked during follow-up. Depending on the antibody panel, at least one LAIP with aberrant antigen expression pattern is found in 80–95% of patients at diagnosis [18, 19]. The DfN strategy searches for aberrant immunophenotypes rarely observed in

<sup>1</sup>Department of Medicine I, University Hospital Carl Gustav Carus TU Dresden, Dresden, Germany. <sup>2</sup>Agendix GmbH, Dresden, Germany. <sup>3</sup>Department of Medicine 5, Universitätsklinikum Erlangen, Erlangen, Germany. <sup>4</sup>Department of Medicine III, University Hospital LMU Munich, Munich, Germany. <sup>5</sup>Laboratory for Translational Cancer Immunology, Gene Center, LMU Munich, Munich, Germany. <sup>6</sup>Department of Internal Medicine and Hematology, Oncology and Immunology, Philipps University Marburg and University Hospital Giessen and Marburg, Marburg, Germany. <sup>7</sup>Department of Internal Medicine V, Heidelberg University Hospital, Heidelberg, Germany. <sup>8</sup>NCT Trial Center, National Center of Tumor Diseases, German Cancer Research Center, Heidelberg, Germany. <sup>9</sup>National Center of Tumor Diseases, Dresden, Germany. <sup>10</sup>Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK. <sup>11</sup>These authors contributed equally: Uta Oelschlägel, Malte von Bonin.

✉email: Maximilian-Alexander.Roehnert@uniklinikum-dresden.de

Received: 5 January 2022 Revised: 24 June 2022 Accepted: 1 July 2022

Published online: 18 July 2022

leukemia-free BM during follow up independent of pre-treatment samples [20]. Both strategies require experienced investigators and rely partly on individualized gates. The LAIP approach is considered more sensitive, but also more susceptible to phenotypic shifts and false negative results [21, 22]. The LAIP analysis is time consuming and the individualized gating for each patient leads to a low inter-rater reliability (IRR) [18, 23, 24]. In contrast, the DfN approach may lead to false positive results due to reactive changes in hematopoiesis exposed to chemotherapy [25, 26]. In addition, differences in the analysis strategy for both concepts lead to heterogeneous MRD results with different prediction values [27, 28]. As LAIP and DfN approaches differ in their strengths and weaknesses, the European LeukemiaNet (ELN) recommends a combinatorial concept, termed LAIP-based DfN approach [29].

The HARMONIZE consortium was established in 2016 to implement standards for MFC based MRD detection within two German AML study groups (SAL, AMLCG). Here, we present a stable, fast and reproducible LAIP-based DfN analysis approach that preserves the prognostic value of MRD assessment.

## METHODS

### Study cohort

Samples from newly diagnosed AML patients at diagnosis (in median 2 days *before* start of induction therapy) and after completion of intensive induction chemotherapy (in median 34 days *after* start of the last induction cycle) were shipped from 30 centers within 24–48 h to the laboratory of the AML Registry of the Study Alliance Leukemia (SAL) in Dresden (institutional review board Dresden, 98032010; clinicaltrials.gov, NCT03188874). Both, BM aspirates and peripheral blood were suitable for MFC at diagnosis. At follow-up, only BM aspirates were utilized. The clinical data was extracted from the registry.

### Reference cohort

All leukemia-free controls (LFC,  $n = 90$ ) were treated at the University Hospital Dresden (Supplementary Table 1) and analyzed by three independent investigators. The aberrant subpopulations as described below ( $n = 32$ ) were also observed in LFC with different frequencies. The upper limit of the one-sided 97.5% reference range for the percentage of each aberrant population among  $CD45^+$  events was set as reference value. LFC included BM aspirates of BM donors ( $n = 30$ ), of patients with acute lymphoblastic leukemia in molecular CR (ALL molCR) ( $n = 19$ ) with a prior exposure to chemotherapy (median 40 days *after* start of the last chemotherapy cycle), of hip surgery patients ( $n = 32$ ) representing older patients and of patients with untreated primary central nervous system lymphoma (PCNSL,  $n = 9$ ).

### Sample preparation and acquisition

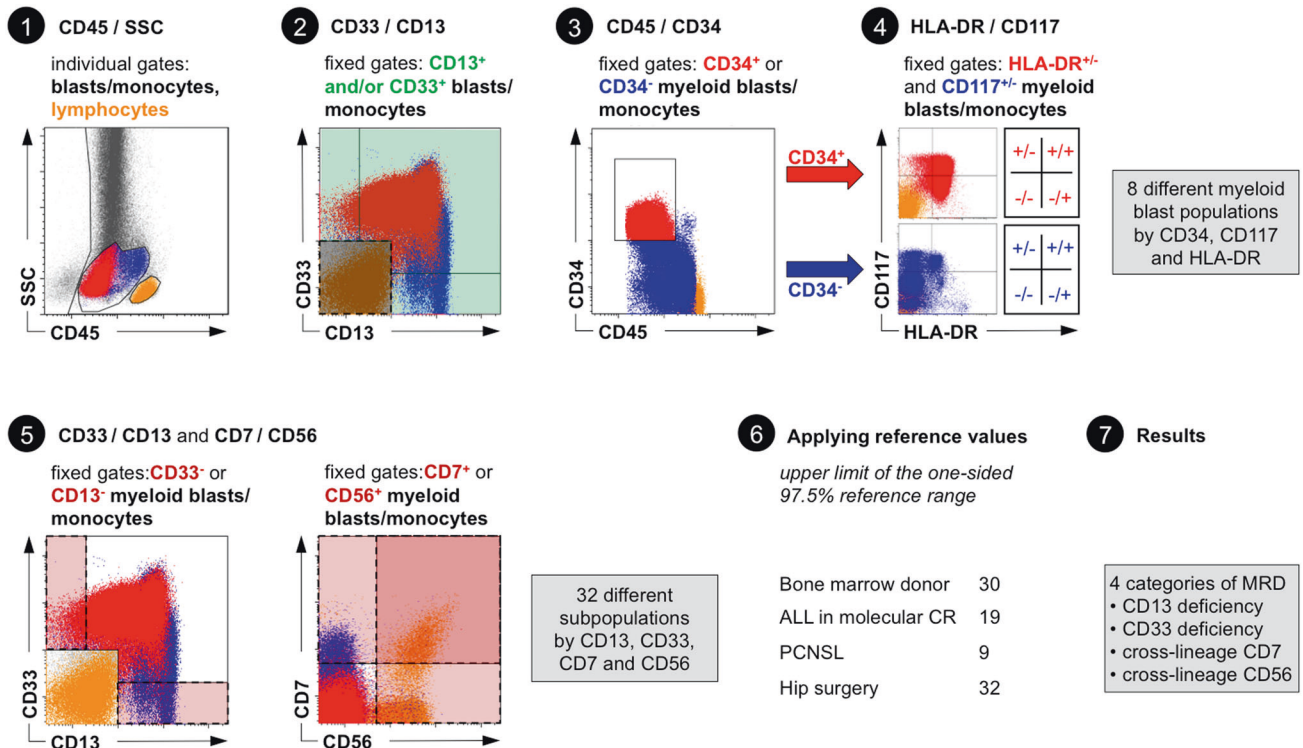
The antibody panel (Supplementary Table 2) consists of eight monoclonal antibodies (mAb) and was designed in 2016 by the HARMONIZE consortium [30–32]. The ELN also recommends the targeted antigens as mandatory core MRD markers [33]. In addition, a comparable mAb panel is used by HOVON/SAKK [29, 34, 35]. However, both panels propagate different mAb clones with divergent fluorochromes.

At least 500,000 events were acquired per tube. The samples were measured centrally and analyzed by at least one of three different investigators. Further details concerning cell preparation and acquisition can be found in the Supplementary material.

### Gating strategy

Kaluza 2.1 software (Beckman Coulter) was used for analysis. An SQL server and Excel (all by Microsoft) served to analyze, store and process the data.

Our proposed LAIP-based DfN analysis is based on a hierarchical gating strategy with fixed gates. The investigators adjusted only the gates for doublet discrimination, exclusion of debris, leukocytes, progenitors/monocytes (P/M) and lymphocytes. The P/M population had to express at least one of the myeloid markers CD13 or CD33 (myP/M) and was subdivided afterward by expression of CD34, CD117 and HLA-DR (the



**Fig. 1 Detailed gating strategy for the proposed LAIP-based DfN MRD approach.** (1) individual gating for blasts/monocytes and lymphocytes using CD45/SSC (2) fixed gates for the expression of at least one myeloid marker (CD13 and/or CD33) on events in the blasts/monocytes gate (3) fixed gate for CD34 on myeloid blasts/monocytes (4) fixed gates to distinguish 8 different myeloid blast populations by CD34, CD117 and HLA-DR (5) fixed gates to define events within the myeloid blasts/monocytes with a deficiency of CD13 or CD33 and a cross-lineage expression of CD7 or CD56 leading to 32 subpopulations (6) applying reference values for each subpopulations (7) MRD assessment by four different categories. Not shown: exclusion of doublets and debris.

backbone markers recommended by the EuroFlow Consortium) [36] resulting in 8 myP/M main populations. These main populations were further characterized by four distinct aberrant categories: deficiency of CD13 or CD33, cross-lineage expression of CD7 or CD56 leading to 32 subpopulations. Subpopulations that exceeded their reference values were used to calculate the MRD load. The difference between measurement and reference value represented disease burden (percentage of CD45<sup>+</sup> cells). Disease burden of subpopulations with identical aberrant category were summed up leading to an aggregated size for deficiency of CD13 or CD33, cross-lineage expression of CD7 or CD56. Only the aberrant category with the highest sum was used to quantify the MRD load.

In our approach, the gates for the myeloid markers (CD13, CD33), the backbone markers (CD34, CD117 and HLA-DR) and the cross-lineage markers (CD7, CD56) were fixed (Fig. 1). The positioning of the fixed gates was driven by review of reference measurements (leukemia-free controls) and internal controls within AML samples (in particular lymphocytes). The leukemic cell population itself never guided the definition of the fixed gates. Consequently, gates sometimes cut through leukemic cell populations and did not follow visual demarcations within the blast population. The gates in our strategy identify pronounced aberrant features only. Deficiency of CD13 and CD33 represent a true absence of these antigens rather than a weak expression. The cross-lineage expression of CD7 and CD56 represents a strong rather than a weak expression.

### Definition of MRD in AML samples

At diagnosis, an aberrant category was defined as LAIP (aLAIP) when  $\geq 10\%$  myP/M were affected. At follow-up only subpopulations with at least  $\geq 20$  events were analyzed. MRD<sup>pos</sup> was defined by at least one aberrant subpopulation exceeding its reference value. When this aberrant category was already detectable at diagnosis this reoccurrence was defined as MRD<sup>pos</sup> by aLAIP. MRD<sup>pos</sup> by post-treatment DfN (ptDfN) was defined as de-novo appearance of an aberrant category.

As ELN recommends a flat 0.1% cut-off for MRD [2, 33], we interpreted our data in two additional ways. First, the analysis was restricted to aberrant subpopulations expressing at least one marker of immaturity and the cut-off for those populations was uniformly set to 0.1% (MRD<sup>posELN</sup> vs. MRD<sup>negELN</sup>). Second, MRD<sup>pos</sup> was subdivided according to the MRD load in MRD<sup>posLo</sup> (our reference values were exceeded by  $< 0.1$ ) and MRD<sup>posHi</sup> (excess of the reference values by  $\geq 0.1$ ).

Many MFC approaches rely on populations expressing markers of immaturity (e.g. CD34<sup>+</sup> and/or CD117<sup>+</sup>) whereas monopoietic cells (identified by SSC and CD45) are frequently excluded. Monocytic AMLs often present without immunophenotypically immature populations [37]. Therefore, we performed additional analyses selecting only aberrant subpopulations expressing at least one marker of immaturity ( $n = 24$ , MRD<sup>posImmOnly</sup> vs. MRD<sup>negImmOnly</sup>).

### Cross-validation

To validate the proposed MRD approach, the results were compared with two already published and established methods to analyze MFC data for presence of MRD: a traditional (manual) flow cytometry approach based on the conventional detection of an aberrant LAIP (convLAIP) [19, 38] and an unsupervised computational approach (Unsup) [39, 40]. The convLAIP approach was restricted to the samples used to calculate the IRR of the proposed approach ( $n = 117$ , see below). Three investigators independently analyzed these samples. The Unsup approach encompassed all follow-up samples.

Furthermore, we compared the results of the proposed MRD approach also with molecular MRD results. Established and decisive molecular markers (*CBFB::MYH11*, *mNPM1* and *RUNX1::RUNX1T1*) as well as other clonal aberrations (e.g. *mRUNX1*, *m/DH1*) were used. A cut-off of 0.1% variant allele frequency was utilized as recommended recently by ELN to distinguish MRD positivity (Mol<sup>pos</sup>) and MRD negativity (Mol<sup>neg</sup>) [33].

Particular attention was given to patients rated MRD<sup>pos</sup> only by ptDfN as this cohort was regarded vulnerable to misinterpretation.

### Inter-rater reliability (IRR)

Three independent investigators analyzed the first 117 follow-up and all LFC samples to define the IRR as quality parameter of the proposed MRD approach. Krippendorffs  $\alpha$  (Ka) as value for the IRR was calculated for two parameters: (I) percentage of CD45<sup>+</sup> events for each of the 32 subpopulations within the LFC and (II) the final MRD status within the AML samples.

### Time requirements for sample analysis

The time to perform the different analysis steps was independently evaluated in 10 samples of the LFC and 10 samples of patients with AML at diagnosis and at follow-up by three investigators, respectively. Different work steps were evaluated: (I - gating) Import of MFC files into Kaluza software and adjusting the non-fixed gates; (II - export) Export of raw data into the SQL database; (III - report) evaluating the MRD status using Excel and creating a MRD report using an Access database.

### Statistical analysis

To define IRR, Ka, a reliability coefficient ranging from 0 to 1, with 1 representing perfect agreement between multiple raters [41], was calculated.

To quantitatively compare different models in their ability to predict clinical outcome, the Akaike information criterion (AIC) was used [42, 43].

The Kaplan–Meier method was used to estimate survival probabilities. Survival curves were compared utilizing the Cox regression model. Multi-variable Cox regression models were used to describe the effect of different variables on survival. A  $p < 0.05$  was regarded as statistically significant. Overall survival (OS) was defined as the time from diagnosis to death from all causes, relapse free survival (RFS) as the time from response to AML relapse or death. In this regard, response was characterized by achievement of complete remission (CR), CR with incomplete hematologic recovery (CRi), or morphologic leukemia-free state (MLFS) [2]. Hematologic relapse, molecular relapse (2 consecutive positive samples for *NPM1mut/ABL*  $> 1\%$  in a previously for *mNPM1* MRD negative patient) or a drop in overall chimerism  $< 80\%$  after aHSC were consistent with relapse [44]. Event free survival (EFS) was defined as time from diagnosis to death from any cause, relapse or allogeneic hematopoietic stem cell transplantation  $> 180$  days after completion of intensive induction therapy, whatever occurred first.

## RESULTS

### Reference values

Reference values were in the range of 0.001% of CD45<sup>+</sup> for the aberrant subpopulation CD34<sup>+</sup>CD117<sup>+</sup>HLA-DR<sup>+</sup>CD56<sup>+</sup> up to 1.992% for the subpopulation CD34<sup>+</sup>CD117<sup>+</sup>HLA-DR<sup>+</sup>CD13<sup>-</sup>. They were substantially influenced by the heterogeneity of the LFC cohorts. Fifteen of the 32 aberrant subpopulations (47%) were mainly influenced by ALL in molecular CR, 10 (31%) by BMD, 5 (16%) by patients undergoing hip surgery and only 2 (6%) by PCNSL (Supplementary Table 3). E.g. cross-lineage expression of CD56 and deficiency of CD13 were mostly seen in ALL, while CD33 deficiency was observed in patients undergoing hip surgery. In general, ALL samples showed the largest variance for most aberrant categories. Due to the minimum population size ( $\geq 20$  events for AML samples), an aberrant subpopulation with a very low reference value can turn MRD positive (MRD<sup>pos</sup>) only when a large number of CD45<sup>+</sup> events is acquired. For example, an aberrant subpopulation barely exceeding its reference value of 0.001%, at least 2,000,000 CD45<sup>+</sup> events would be necessary to obtain  $\geq 20$  relevant events. Four aberrant subpopulations (CD34<sup>+</sup>CD117<sup>+</sup>HLA-DR<sup>+</sup>CD13<sup>-</sup>; CD34<sup>+</sup>CD117<sup>+</sup>HLA-DR<sup>+</sup>CD7<sup>+</sup>; CD34<sup>+</sup>CD117<sup>+</sup>HLA-DR<sup>+</sup>CD56<sup>+</sup>; CD34<sup>+</sup>CD117<sup>+</sup>HLA-DR<sup>+</sup>CD56<sup>+</sup>) were affected by this phenomenon at the targeted acquisition of 500,000 events.

### Inter-rater reliability (IRR) of the leukemia-free controls (LFC)

In the LFC cohort, Ka for the 8 main populations was 0.757–0.990. All but one main population presented with Ka  $\geq 0.800$ . The populations with the lowest contingency (CD34<sup>+</sup>CD117<sup>+</sup>) only differed in the expression of HLA-DR. Even for the subpopulations with aberrant features ( $n = 32$ ), there was a considerable high IRR for the percentage of CD45<sup>+</sup> events with a Ka  $> 0.900$  for the deficiency of CD13, followed by Ka for the deficiency of CD33, the cross-lineage expression of CD7 and the cross-lineage expression of CD56 with  $> 0.800$ ,  $> 0.700$ , and  $> 0.600$ , respectively (Supplementary Table 3).

### Time requirements for sample analysis

The mean time for gating (I) and export of the results (II) was 01:18 min and 00:47 min, without significant differences between



**Table 1.** Time requirement (minutes:seconds) of the working steps at diagnosis and follow-up, leukemia-free controls (LFC).

	I - gating Mean (SD)	II - export Mean (SD)	III - report Mean (SD)	overall Mean (SD)
Diagnosis	01:03 (00:12)	00:46 (00:09)	02:33 (00:16)	04:17 (00:33)
Follow-up	01:31 (00:30)	00:51 (00:11)	02:09 (00:20)	04:31 (00:49)
LFC	01:20 (00:17)	00:45 (00:08)	not intended	not intended
overall	01:18 (00:24)	00:47 (00:10)		

<sup>a</sup>SD standard deviation, LFC leukemia-free controls.

diagnosis, follow-up and LFC samples. Overall time for analysis, data transfer and generation of a report was 04:17 min and 04:31 min for diagnosis and follow-up samples, respectively (Table 1).

### MRD assessment

Our analysis included 246 patients with AML (non-APL) who were treated with intensive induction therapy and for whom reliable clinical data and at least one suitable MRD analysis at follow-up was available (Table 2).

Diagnosis and follow-up samples were available for 216/246 patients (88%). At diagnosis, at least one aberrant category affecting  $\geq 10\%$  or  $\geq 5\%$  of the myP/M could be detected in 152/216 (70%) or 179/216 patients (83%), respectively. In the following, the 10% threshold for the presence of a LAIP was used as formerly recommended by ELN [2, 23, 45]. In 56/152 (37%) of patients the aLAIP was defined exclusively by aberrant subpopulations with markers of immaturity and in 72/152 (47%) without markers of immaturity. The most common aberrant category at diagnosis was deficiency of CD13 in 75/152 patients (49%). Cross-lineage expression of CD56, CD7 and deficiency of CD33 was detectable in 44/152 (29%), 44 (29%), and 41 patients (27%), respectively. One, two, or three categories were simultaneously detectable in 103/152 (68%), 46 (30%), and 3 (2%), with the most common combination of CD13 deficiency plus cross-lineage expression of CD56 in 21/49 (43%) of the patients.

At follow-up, in total 157/246 patients (64%) were MRD<sup>pos</sup> (Fig. 2A). They were classified as MRD<sup>pos</sup> by aLAIP<sup>only</sup>, ptDfN<sup>only</sup> and aLAIP/ptDfN in 33/157 (21%), 80 (51%) and 44 (28%) cases, respectively. In MRD<sup>pos</sup> patients, deficiency of CD13 or CD33, cross-lineage expression of CD7 or CD56 was observed in 76/157 (48%), 66 (42%), 74 (47%) and 87 (55%). One, two, three and four aberrant categories could be simultaneously detected in 71/157 (45%), 41 (26%), 30 (19%), and 15 (10%). Most of the patients of the MRD<sup>pos</sup> group had undergone immunophenotyping at diagnosis (137/157; 87%).

In the subgroup of responders ( $n = 180$ ) the proportion of MRD<sup>pos</sup> patients was significantly lower compared to non-responders ( $n = 66$ ): 99/180 (55%) versus 58/66 (89%;  $p < 0.0001$ ). For most of the MRD<sup>pos</sup> responders a measurement at diagnosis (86/99; 87%) was available and MRD<sup>pos</sup> was classified by aLAIP<sup>only</sup>, ptDfN<sup>only</sup> and aLAIP/ptDfN in 24/86 (28%), 45 (52%) and 17 (20%), respectively (Fig. 2B).

MRD assessment by three independent investigators for a cohort of 117 consecutive samples showed that 107/117 (92%) cases were classified concordantly leading to a Ka of 0.86.

The median follow-up time for the entire cohort was 18.9 months (IQR 10.9–29). Compared to MRD<sup>neg</sup> patients, the OS was significantly shorter in the MRD<sup>pos</sup> group (HR 5.6, CI: 2.2–14.1, Fig. 2A). Thus, the 2-year OS was 92% (CI: 86–99%) for MRD<sup>neg</sup> and 63% (CI: 55–73%) for MRD<sup>pos</sup>, respectively. MRD status retained its importance on OS in patient cohorts stratified according to response (CR/CRi/MLFS and RD/PR, Fig. 2B). In responders, the 2-year OS was 91% (CI: 84–99%) for MRD<sup>neg</sup> and 68% (CI: 57–81%) for MRD<sup>pos</sup> (HR 3.8; CI: 1.5–10.0;  $p = 0.006$ ). The

OS for the MRD<sup>neg</sup> non-responders was comparable with MRD<sup>neg</sup> responders. The MRD status also retained its impact on OS after stratifying patients according to ELN risk category (Supplementary Fig. 1A–C).

Regarding RFS and EFS, MRD<sup>pos</sup> was associated with a significantly shorter RFS and EFS (HR 2.6, CI: 1.4–5.0; and HR 2.7, CI: 1.5–5.0, Fig. 2C, D).

Most importantly, in a multivariable Cox regression model, MRD<sup>pos</sup> retained its significant prognostic impact on OS for all patients, OS for responders, RFS and EFS (Table 3).

As ELN recommends a 0.1% cut-off [29], we interpreted our data in two additional ways. First, the analysis was restricted to aberrant subpopulations expressing at least one marker of immaturity (CD34<sup>+</sup> and/or CD117<sup>+</sup>) and the appendant reference values were uniformly set to 0.1%. In this context, evidence of MRD was termed MRD<sup>posELN</sup>. Of the 246 patients, only 24% ( $n = 60$ ) fulfilled MRD<sup>posELN</sup> criteria. The MRD<sup>ELN</sup> analysis still showed significant prognostic relevance, however, the discriminatory power was less compared to the original strategy (Supplementary Fig. 2A). Second, MRD<sup>pos</sup> was subdivided according to the MRD load in MRD<sup>posLo</sup> (the reference value was exceeded by  $< 0.1$ ) and MRD<sup>posHi</sup> (excess of the reference value by  $\geq 0.1$ ). Of the 157 MRD<sup>pos</sup> patients, one third was assigned to the MRD<sup>posLo</sup> cohort. The MRD load (MRD<sup>posLo</sup> vs. MRD<sup>posHi</sup>) did not provide further prognostic information regarding OS (Supplementary Fig. 2B).

As many analysis strategies are focused on aberrancies in the immature compartment, the data was further analyzed utilizing a strategy restricted to aberrant subpopulations expressing at least one of those markers (24 subpopulations) and compared to the strategy encompassing all aberrant subpopulations (32 subpopulations). Of the 246 patients, 49% ( $n = 121$ ) were classified as MRD<sup>posImmOnly</sup> providing slightly less prognostic impact compared to the proposed MRD approach utilizing all subpopulations. Differential analysis of MRD<sup>pos</sup> by aLAIP<sup>only</sup>, ptDfN<sup>only</sup> or aLAIP/ptDfN did not improve the prediction of outcome (Supplementary Fig. 2C, D, Table 5).

With respect to clinical characteristics (Table 2), variables known to negatively impact patient outcome were enriched in MRD<sup>pos</sup> patients (age, ELN risk category, karyotype, FLT3 mutation status, MRC score and morphological response). Accordingly, the frequency of aHSCT was higher in the MRD<sup>pos</sup> group (MRD<sup>neg</sup> 55% (49/89) versus MRD<sup>pos</sup> 77% (121/157),  $p < 0.0001$ ).

### Cross-validation

The convLAIP approach was applicable to  $n = 106$  cases with measurements at diagnosis and follow-up. In 99% of the pre-therapeutic samples at least one traceable aLAIP could be detected by the convLAIP approach, in contrast to 78% by the proposed methodology. The aLAIP detected by both approaches typically shared similar features (in 99% of cases). Ka for MRD assessment by convLAIP was 0.59. The MRD status of 69% of follow up samples was rated concordantly by the convLAIP approach and the proposed approach (convLAIP<sup>pos</sup>/MRD<sup>pos</sup> 39%, convLAIP<sup>neg</sup>/MRD<sup>neg</sup> 30%). There was disagreement on the MRD status in 31% of follow up samples, almost always as convLAIP<sup>neg</sup>/MRD<sup>pos</sup> constellation

**Table 2.** Patient and disease characteristics, MRD classification according to the proposed MRD approach.

		Overall <i>n</i> = 246	MRD <sup>neg</sup> <i>n</i> = 89 (36%)	MRD <sup>pos</sup> <i>n</i> = 157 (64%)	
Age					
Median (IQR)	Years	56 (46–63)	50 (36–60)	58 (51–64)	<i>p</i> < 0.0001
>65 years	<i>n</i> (%)	36 (15)	7 (8)	29 (19)	<i>p</i> = 0.0381
Sex					
Female	<i>n</i> (%)	104 (42)	40 (45)	64 (41)	<i>p</i> = 0.6438
ECOG					
Miss	<i>n</i>	29	15	14	
0–1	<i>n</i> (%)	202 (93)	70 (95)	132 (92)	<i>p</i> = 0.7284
2–4	<i>n</i> (%)	15 (7)	4 (5)	11 (8)	
AML type					
Miss	<i>n</i>	27	13	14	<i>p</i> = 0.1783
De-novo	<i>n</i> (%)	185 (75)	68 (76)	117 (75)	
sAML	<i>n</i> (%)	19 (8)	6 (7)	13 (8)	
tAML	<i>n</i> (%)	15 (6)	2 (2)	13 (8)	
BM blasts in %					
Miss	<i>n</i>	28	14	14	
Median (IQR)		60 (39–80)	65 (50–80)	55 (33–80)	<i>p</i> = 0.1022
WBC in GPT/l					
Miss	<i>n</i>	27	13	14	
Median (IQR)		9.7 (2.9–43.8)	17.9 (4.3–43.6)	7.6 (2.7–43.3)	<i>p</i> = 0.0870
Complex karyotype					
Miss	<i>n</i>	50	19 (21)	31 (20)	
Yes	<i>n</i> (%)	15 (8)	1 (1)	14 (11)	<i>p</i> = 0.0306
FLT3					
Miss	<i>n</i>	12	7	5	
FLT3-ITD	<i>n</i> (%)	64 (27)	32 (39)	32 (21)	<i>p</i> = 0.0053
NPM1					
Miss	<i>n</i>	11	5	6	
NPM1 <sup>mut</sup>	<i>n</i> (%)	83 (35)	36 (43)	47 (31)	<i>p</i> = 0.0967
ELN 2017					
Miss	<i>n</i>	4	0	4	
Favorable	<i>n</i> (%)	86 (36)	44 (49)	42 (28)	<i>p</i> = 0.0005
Intermediate	<i>n</i> (%)	89 (37)	31 (35)	58 (38)	
Adverse	<i>n</i> (%)	67 (28)	14 (16)	53 (35)	
MRC index					
Miss	<i>n</i>	55	21	34	
Good	<i>n</i> (%)	12 (6)	9 (13)	3 (2)	<i>p</i> < 0.0001
Standard	<i>n</i> (%)	73 (38)	34 (50)	39 (32)	
Poor	<i>n</i> (%)	106 (56)	25 (37)	81 (66)	
Induction					
2 cycles	<i>n</i> (%)	128 (52)	52 (58)	76 (48)	<i>p</i> = 0.1680
Consolidation					
Chemo	<i>n</i> (%)	76 (31)	40 (45)	36 (23)	<i>p</i> < 0.0001
alloHSCT <6 month	<i>n</i> (%)	107 (43)	34 (38)	73 (46)	
alloHSCT >6 month	<i>n</i> (%)	63 (26)	15 (17)	48 (31)	
Response at follow-up					
RD	<i>n</i> (%)	23 (9)	2 (2)	21 (13)	<i>p</i> < 0.0001
PR	<i>n</i> (%)	43 (18)	6 (7)	37 (24)	
CR/CRi/MLFS	<i>n</i> (%)	180 (73)	81 (91)	99 (63)	

<sup>a</sup>IQR interquartile range, ECOG Eastern Cooperative Oncology Group Performance Status Scale, sAML secondary AML, tAML therapy related AML, BM bone marrow, WBC white blood cell, ELN 2017 European LeukemiaNet risk stratification by genetics 2017, MRC Index Medical Research Council Index for risk stratification of AML, alloHSCT allogeneic hematopoietic stem cell transplantation, RD refractory disease, PR partial remission, CR complete remission, CRi complete remission with incomplete count recovery, MLFS morphological leukemia-free state.

(Supplementary Table 4). Nevertheless, the convLAIP provided prognostic power regarding overall survival (Supplementary Fig. 3A), but convLAIP<sup>neg</sup>/MRD<sup>pos</sup> patients showed a comparable outcome to patients rated MRD positive by both approaches (convLAIP<sup>pos</sup>/MRD<sup>pos</sup>, Supplementary Fig. 3B).

The Unsup pipeline was applicable to 244/246 (99%) of follow-up measurements. The Unsup pipeline and the proposed approach showed a slightly higher concordance on MRD rating (Unsup<sup>pos</sup>/MRD<sup>pos</sup> 45%, Unsup<sup>neg</sup>/MRD<sup>neg</sup> 28%) compared to the conventional LAIP approach. This time, inconsistent results were spread to both conflicting categories (Unsup<sup>neg</sup>/MRD<sup>pos</sup> 19%, Unsup<sup>pos</sup>/MRD<sup>neg</sup> 8%) (Supplementary Table 4). The Unsup pipeline also provided significant prognostic power (Supplementary Fig. 3C).

In 99/246 patients (40%), decisive molecular markers for MRD monitoring were available at diagnosis (*mNPM1*: 78, *RUNX1::RUNX1T1*: 8, *CBFB::MYH11*: 13). For 85 of these 99 patients (*mNPM1*: 67, *RUNX1::RUNX1T1*: 6, *CBFB::MYH11*: 12) molecular MRD results at follow-up were available. In addition, tracking of less established molecular markers (biallelic *mCEBPA*: 7, *mCEBPA-TAT*: 1, *mDNMT3A*: 1, *FLT3-ITD*: 4, *FLT3-TKD*: 1, *mIDH1*: 1, *mIDH2*: 7, *KMT2A::MLLT3*: 2, *KMT2A-PTD*: 10, *PICALM::MLLT10*: 1, *mRUNX1*: 2, *mSRSF2*: 1, *mTET2*: 1, *mTP53*: 1, *mUTAF1*: 1) was done. Discordant results (Mol<sup>neg</sup>/MRD<sup>pos</sup> or Mol<sup>pos</sup>/MRD<sup>neg</sup>) were observed in 15/126 (12%) and 26/126 (21%), respectively (Supplementary Table 4). A shorter OS was observed for Mol<sup>pos</sup> compared to Mol<sup>neg</sup> without reaching statistical significance (Supplementary Fig. 3D).

As patients classified as MRD<sup>pos</sup> by ptDfN<sup>only</sup> were regarded vulnerable to misinterpretation this cohort was analyzed in more detail. A measurement at diagnosis was available in 59/80 (74%) cases. In 43/59 (73%) of these patients, a minor subclone (<10% of myP/M) with identical aberrant category was already detectable at diagnosis (affecting in median 1.9% of myP/M, IQR 4.2%). The reduction of the population size to define an aLAIP at diagnosis considerably lowered the number of patients classified as MRD<sup>pos</sup> by ptDfN<sup>only</sup>;  $\geq 10\%$ : 59,  $\geq 5\%$ : 49,  $\geq 2.5\%$ : 41, and  $\geq 1\%$ : 30. Only in 16/59 (27%) patients the aberrant category of ptDfN<sup>only</sup> was not detectable at all at diagnosis (cross-lineage expression of CD56: 10, deficiency of CD13: 1, and deficiency of CD33: 5). For MRD<sup>pos</sup> by ptDfN<sup>only</sup> patients a simultaneous molecular MRD testing was available in 11/80 (14%) cases. Concordant results were observed in 64% of these cases (*mNPM1*: 4, *CBFB::MYH11*: 1, *mIDH2*: 1, *FLT3-ITD*: 1). MRD<sup>pos</sup> by ptDfN<sup>only</sup> patients were rated convLAIP<sup>pos</sup> in 31% of the cases. In 13/80 (16%) CR/CRI samples obtained at later time points (during/after conventional consolidation) from patients rated MRD<sup>pos</sup> by ptDfN<sup>only</sup> post-induction, the same “de-novo post treatment aberrant category” could be detected again in 62% of cases. A measurement at relapse was available for 14/80 (18%) patients rated MRD<sup>pos</sup> by ptDfN<sup>only</sup> at post-induction. The same “de-novo post treatment aberrant category” could be observed in 71% of relapse samples.

## DISCUSSION

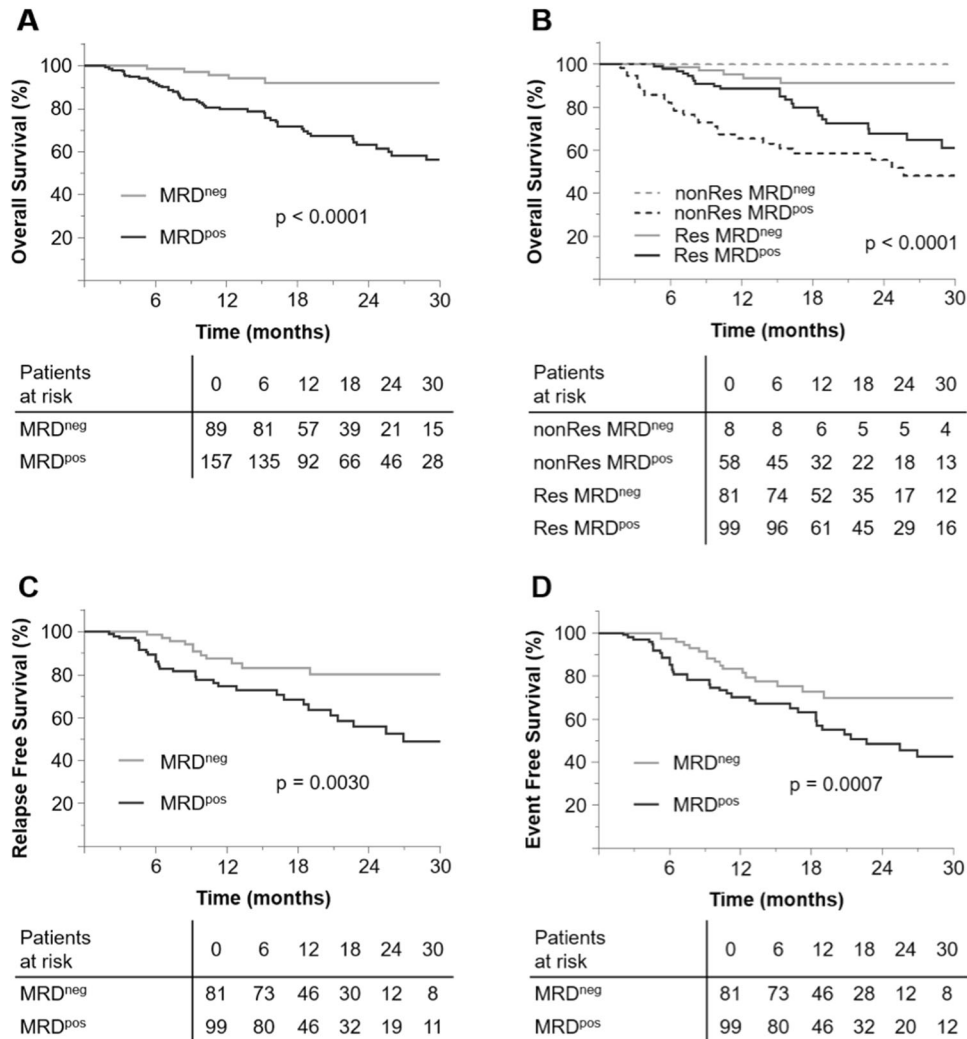
Even though MRD assessment by MFC is technically available for the majority of patients with AML, its broad applicability is still hampered due to the lack of standardization. The focus of this work was to develop a robust, fast and reproducible LAIP-based DfN analysis strategy to evaluate MRD by MFC.

The analysis strategy focused separately on two kinds of abnormalities: reduced expression of myeloid antigens and cross-lineage expression of lymphoid antigens [46]. The reduced expression of CD13 and CD33 as part of a LAIP has been variably described in 10–22% and 18–36% of AML cases, respectively [47–49]. Also, the frequency of the cross-lineage expression of CD7 and CD56 varies substantially with 17–43% [45, 49, 50]. This variability is not only explainable by the aberration-defining gate

itself, but also preceding gating steps and the reference population have a major impact on the observation frequency.

Deviating from most analysis strategies for MRD-assessment, we decided to establish one single tube, but augmented the number of populations to be analyzed. The progenitor cell gate was expanded to include monocytes (P/M). P/M cells were required to express at least CD13 or CD33 (myP/M). At diagnosis, this myeloid assignment was negligible as in median 94% (IQR 18.1%) of cells in the P/M gate fulfilled this criterion. At follow-up in median only 84% (IQR 20.3%) of P/M cells met this specification. As unique selling point our MRD analysis includes also the CD34<sup>-</sup>CD117<sup>-</sup> compartment within myP/M largely representing monopietic cells. Indeed, acute monoblastic/monocytic leukemia represents approximately 12% of the AML patients [51] and shows an expression of CD34<sup>+</sup> or CD117<sup>+</sup> only in 7.7% and 19.8% of cases, respectively [37]. Within our approach, the reference values for the aberrant populations without markers of immaturity were higher compared to aberrant populations expressing either CD34 or CD117 (Supplementary Table 5). For other entities like MDS, the evaluation of monocytes using e.g., CD56 is part of various diagnostic scores [52, 53]. In addition, CD56 expression has been described to distinguish clonal monocytes within CMML from reactive monocytosis [54–56]. These observations have led us to also analyze aberrations outside of the CD34<sup>+</sup>CD117<sup>+</sup> compartment. In fact, exclusion of aberrant populations without expression of markers of immaturity mostly led to a decline in the informative value of the here proposed MRD approach as calculated by the AIC, which supports the assumption that populations beyond phenotypically immature cells also contain prognostic information. Most MRD<sup>pos</sup> patients ( $n = 84$ ) showed aberrancies in both compartments. Of note, 36 patients were classified as MRD<sup>pos</sup> solely by aberrations within the compartment without expression of markers of immaturity. Some of these patients were at the same time also HLA-DR negative. The most common aberrant categories were cross-lineage expression of CD56 and deficiency of CD33 (each 44%). A cross-lineage expression of CD7 or a deficiency of CD13 was not found in these cases. This observation fits well with previously published data reporting that leukemic immature monocytes used for MRD monitoring by a LAIP-approach were frequently characterized by decreased expression of HLA-DR and increased expression of CD56 and CD13 [57]. Again, the assignment of these cells to the monopietic compartment and their maturity remained somewhat speculative as the panel did not allow a proper categorization as neither monocytic markers nor other antigens associated with immaturity e.g., CD133 were evaluated [58].

In only 28% of MRD<sup>pos</sup> responders, the rating was based solely on the detection of an aberrant category already evident at diagnosis, whereas ptDfN<sup>only</sup> defined MRD<sup>pos</sup> in 51% of cases. This unexpected high rate of ptDfN<sup>only</sup> is partly related to the availability of measurements at diagnosis and the LAIP definition used in our approach. By modifying the aLAIP definition to  $\geq 5\%$  of myP/M, in 83% of patients at least one LAIP was detectable at diagnosis and the MRD<sup>pos</sup> rate by ptDfN<sup>only</sup> dropped from 38% to 31%. A minor subclone with identical aberrant category was already detectable at diagnosis in 73% of these patients. In addition, the assumption that ptDfN<sup>only</sup> mostly represents “true” MRD than phenotypic shifts, was supported by simultaneous molecular MRD testing that showed 64% concordant results. Furthermore, the same “de-novo post treatment aberrant category” could be observed in 71% of relapse samples. Indeed, selection pressure by chemotherapy can change the original composition and initially existing but rather small populations expand, as documented for (molecular) genetics [59] and immunophenotypes [21]. However, phenotypic shifts might have also contributed. This conception was supported by the observation that 32/105 (30%) patients were assessed discordantly at follow-up using the proposed MRD approach and the convLAIP



**Fig. 2 Outcome stratified by the proposed LAIP-based DfN MRD approach. A** Overall survival for all patients, **B** Overall survival stratified by cytomorphologic response and MRD status, **C** Relapse free survival, **D** Event free survival.

approach (convLAIP<sup>neg</sup>/MRD<sup>pos</sup>). Features of the aLAIP defined by the convLAIP at diagnosis were detected in 99% by the proposed MRD approach. Leukemic cells can undergo phenotypic shifts during the course of the disease due to an evolution (emergence of not previously present immunophenotypes). At relapse, a gain in the expression of immaturity markers is frequently described [21]. Phenotypic shifts as a result of selection pressure or clonal evolution are usually not clearly distinguishable, but nevertheless representing both “true” MRD<sup>pos</sup>. But abnormalities in immunophenotype have been also observed leukemia-independent (age and treatment related) and might result in false positive results [20, 25, 60]. Clonal hematopoiesis is also suggested to be associated with phenotypic aberrations [61, 62].

The reference values for the aberrant populations with markers of immaturity correspond well with previously published sensitivity levels of MFC methods ( $10^{-4}$ – $10^{-5}$ ) [49, 63]. To smoothen the heterogeneity in reference values, ELN recommended a flat 0.1% cut-off ( $10^{-3}$ ) as this level had been of prognostic relevance in most publications and is at least one log above the published sensitivity level for MRD by MFC [63]. In our approach, 7 out of 32 aberrant populations presented with reference values  $>10^{-3}$ , so these subpopulations (all expressed neither of CD34 nor CD117) had to be excluded from the analysis with this uniform cut-off (termed MRD<sup>posELN</sup>). In the end, the different approaches reduced the informative value of our gating strategy.

The quality of MRD assessment by MFC is considerably affected by the reference values. Although, the cohort of ALL molCR only represented 21% of the LFC samples, roughly this cohort determined 50% of the reference values. BMD, the most commonly used LFC cohort with predominantly young subjects, represented 30% of the LFC cohort and only established 31% of reference values. This distribution pattern underpins the necessity to include a broad range of different LFC cohorts as various factors (e.g., prior exposition to chemotherapy and age) substantially influence the frequency of certain expression profiles.

Most importantly, our LAIP-based DfN analysis strategy (including cell compartments with and without expression of markers of immaturity) provided significant prognostic information on clinical outcome after intensive induction treatment. MRD<sup>pos</sup> patients showed a significantly shorter OS and a higher relapse risk, both in univariable as well as multivariable regression models. The 2017 ELN genetic risk stratification is frequently used for pretreatment risk assessment [2]. Of note, our MRD results helped to further segregate the prognosis within each ELN risk category. The MRD status was most predictive for outcome in patients with favorable and adverse risk. The importance in the adverse risk category was not surprising, as MRD status pre-transplant has been described to be of prognostic significance [46]. Whether allogeneic transplantation and intensity of the conditioning regimen can have an influence on the MRD-associated prognosis is a matter of debate



**Table 3.** Multivariable Cox regression model of the proposed MRD approach; ELN 2017: ordinal scale; age: rational scale; ECOG: ordinal scale; sAML: nominal scale; tAML: nominal scale.

	<b>OS – all patients</b> <b>HR (CI; p-value)</b>	<b>OS - responder</b> <b>HR (CI; p-value)</b>	<b>RFS</b> <b>HR (CI; p-value)</b>	<b>EFS</b> <b>HR (CI; p-value)</b>
MRD <sup>POS</sup>	3.7 (1.4–9.5; 0.007)	3.0 (1.1–8.6; 0.036)	2.1 (1.0–4.3; 0.047)	2.3 (1.2–4.5; 0.018)
ELN fav	1.3 (0.7–2.5; 0.410)	1.1 (0.4–2.8; 0.841)	1.0 (0.5–2.0; 0.996)	0.8 (0.4–1.5; 0.458)
ELN adv	0.6 (0.3–1.4; 0.251)	1.7 (0.7–4.4; 0.275)	1.1 (0.5–2.3; 0.786)	0.9 (0.5–1.9; 0.846)
age	1.0 (1.0–1.1; 0.134)	1.0 (1.0–1.0; 0.906)	1.0 (1.0–1.0; 0.878)	1.0 (1.0–1.0; 0.739)
ECOG 2–4	2.1 (0.8–5.4; 0.128)	1.0 (0.2–4.3; 0.993)	1.3 (0.5–3.7; 0.645)	2.6 (1.1–6.3; 0.034)
sAML	1.0 (0.4–2.4; 0.988)	0.6 (0.1–4.4; 0.590)	0.3 (0.0–2.3; 0.249)	0.3 (0.0–1.9; 0.182)
tAML	1.1 (0.5–2.7; 0.795)	2.0 (0.7–5.5; 0.176)	1.7 (0.7–4.1; 0.210)	1.8 (0.8–4.2; 0.155)

<sup>a</sup>OS overall survival, RFS relapse free survival, EFS event free survival, HR hazard ratio, CI confidence interval, ELN fav/ adv favorable/ adverse risk according to European LeukemiaNet risk stratification by genetics 2017, ECOG Eastern Cooperative Oncology Group Performance Status Scale, sAML secondary AML, tAML therapy related AML.

[39, 64, 65]. The discriminatory power of the MRD status remained valid when established baseline prognostic variables were considered. MRD<sup>POS</sup> was associated with adverse outcome in responding as well as non-responding patients. However, in non-responders, the MRD status did not reach statistical significance due to low patient numbers. Nevertheless, our data suggest that the proposed approach can reliably distinguish vigorous hematopoietic regeneration with an increase in normal progenitors from persistence of leukemic cells. This is of particular importance as both scenarios are associated with a totally different prognosis. Summarizing, our analysis strategy could confirm the prognostic significance of the MRD status after intensive induction treatment.

In contrast to previous reports, we explicitly focused on applicability of the MRD assessment within clinical routine. Current protocols with manual gating are time-consuming (no published data available), they rely on the expertise of the individual investigator and are therefore prone to inter-rater variations [20, 34]. The proposed MRD approach is fast and shows a very promising IRR. Artificial intelligence is established as a research tool in order to circumvent these disadvantages [39, 66, 67], but has not been implemented as diagnostic test in the daily clinical routine yet. The introduction of fixed gates within our approach resulted in a high inter-rater reliability with respect to both, LFC and AML samples and in short analysis time.

## CONCLUSION

We present a hierarchical gating strategy, combining the LAIP and DfN analysis approaches, which allows a high level of MFC standardization and a promising inter-rater reliability in MRD detection. Our standardized MFC approach is implementable at other laboratories and enables standardized multicentric immunophenotypic MRD assessment. Such standardization is an important step towards individualized treatment decisions within routine AML therapy and MFC may thus also serve as a biomarker within prospective clinical trials.

## REFERENCES

- Büchner T, Urbanitz D, Hiddemann W, Rühl H, Ludwig WD, Fischer J, et al. Intensified induction and consolidation with or without maintenance chemotherapy for acute myeloid leukemia (AML): Two multicenter studies of the German AML Cooperative Group. *J Clin Oncol*. 1985;3:1583–9.
- Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129:424–47.
- Papaemmanuil E, Ph D, Gerstung M, Ph D, Bullinger L, Gaidzik VI, et al. Genomic classification and prognosis in acute myeloid. *Leukemia*. 2016;374:2209–21.
- 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;374:422–33.
- Liu Yin JA, O'Brien MA, Hills RK, Daly SB, Wheatley K, Burnett AK. Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: Results of the United Kingdom MRC AML-15 trial. *Blood*. 2012;120:2826–35.
- Rücker FG, Agrawal M, Corbacioglu A, Weber D, Kapp-Schwoerer S, Gaidzik VI, et al. Measurable residual disease monitoring in acute myeloid leukemia with t(8;21)(q22;q22.1): Results from the AML Study Group. *Blood*. 2019;134:1608–18.
- Willekens C, Blanchet O, Renneville A, Cornillet-Lefebvre P, Pautas C, Guieze R, et al. Prospective long-term minimal residual disease monitoring using RQ-PCR in RUNX1-RUNX1T1-positive acute myeloid leukemia: Results of the French CBF-2006 trial. *Haematologica*. 2016;101:328–35.
- Ossenkoppele GJ, Schuurhuis GJ. MRD in AML: It is time to change the definition of remission. *Best Pract Res: Clin Haematol*. 2014;27:265–71.
- Buccisano F, Maurillo L, Spagnoli A, Del Principe ML, Fraboni D, Panetta P, et al. Cytogenetic and molecular diagnostic characterization combined to post-consolidation minimal residual disease assessment by flow cytometry improves risk stratification in adult acute myeloid leukemia. *Blood*. 2010;116:2295–303.
- Buccisano F, Dillon R, Freeman SD, Venditti A. Role of minimal (measurable) residual disease assessment in older patients with acute myeloid leukemia. *Cancers*. 2018;10:205.
- 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.
- Loken MR, Alonzo TA, Pardo L, Gerbing RB, Raimondi SC, Hirsch BA, 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.
- Terwijn M, Kelder A, Huijgens PC, Dräger AM, Oussoren YJM, Scholten WJ, 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.
- Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukemia. *Leukemia*. 2017;31:1482–90.
- Walter RB, Gooley TA, Wood BL, Milano F, Fang M, Sorror ML, et al. Impact of pretransplantation minimal residual disease, as detected by multiparametric flow cytometry, on outcome of myeloablative hematopoietic cell transplantation for acute myeloid leukemia. *J Clin Oncol*. 2011;29:1190–7.
- Rautenberg C, Stölzel F, Röllig C, Stelljes M, Gaidzik V, Lauseker M, et al. Real-world experience of CPX-351 as first-line treatment for patients with acute myeloid leukemia. *Blood Cancer J*. 2021;11:164.
- 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.
- Schuurhuis GJ, Ossenkoppele GJ, Kelder A, Cloos J. Measurable residual disease in acute myeloid leukemia using flow cytometry: approaches for harmonization/standardization. *Expert Rev Hematol*. 2018;11:921–35.
- Köhnke T, Bücklein V, Rechkemmer S, Schneider S, Rothenberg-Thurley M, Metzeler KH, et al. Response assessment in acute myeloid leukemia by flow cytometry supersedes cytomorphology at time of aplasia, amends cases without molecular residual disease marker and serves as an independent prognostic marker at time of aplasia and post-induction. *Haematologica*. 2019;104:E510–3.
- Wood BL. Acute myeloid leukemia minimal residual disease detection: the difference from normal approach. *Curr Protoc Cytom*. 2020;93:e73.

21. Zeijlemaker W, Gratama JW, Schuurhuis GJ. Tumor heterogeneity makes AML a 'moving target' for detection of residual disease. *Cytometry B Clin Cytom.* 2013;86:3–14.
22. Oelschlägel U, Nowak R, Schaub A, Köppel C, Herbst R, Mohr B, et al. Shift of aberrant antigen expression at relapse or at treatment failure in acute leukemia. *Cytometry.* 2000;42:247–53.
23. Feller N, Van Der Velden VHJ, Brooimans RA, 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.
24. Brooimans RA, van der Velden VHJ, Boeckx N, Slomp J, Preijers F, te Marvelde JG, et al. Immunophenotypic measurable residual disease (MRD) in acute myeloid leukemia: Is multicentric MRD assessment feasible? *Leuk Res.* 2019;76:39–47.
25. Camburn AE, Petrasich M, Ruskova A, Chan G. Myeloblasts in normal bone marrows expressing leukaemia-associated immunophenotypes. *Pathology.* 2019;51:502–6.
26. Eckel AM, Cherian S, Miller V, Soma L. CD33 expression on natural killer cells is a potential confounder for residual disease detection in acute myeloid leukemia by flow cytometry. *Cytometry B Clin Cytom.* 2020;98:174–8.
27. Rosso A, Juliusson G, Lorenz F, Lehmann S, Derolf Å, Deneberg S, et al. Is there an impact of measurable residual disease as assessed by multiparameter flow cytometry on survival of AML patients treated in clinical practice? A population-based study. *Leuk Lymphoma.* 2021;62:1973–81.
28. Paiva B, Vidriales MB, Sempere A, Tarín F, Colado E, Benavente C, et al. Impact of measurable residual disease by decentralized flow cytometry: a PETHEMA real-world study in 1076 patients with acute myeloid leukemia. *Leukemia.* 2021;35:2358–70.
29. Schuurhuis GJ, Heuser M, Freeman S, Béne MC, Buccisano F, Cloos J, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood.* 2018;131:1275–91.
30. Röhnert M, von Bonin M, Kramer M, Ensel P, Holtschke N, Röllig C, et al. Standardized identification of measurable residual disease (MRD) by multicolor flow cytometry (MFC) in patients with acute myeloid leukemia (AML). *EHA Libr.* 2020;06/12/20:EP566.
31. Röhnert M, von Bonin M, Bücklein V, Krause S, Völkl S, Rieger M, et al. Comparison of leukemia-associated immunophenotype (LAIP)-based and different-from-normal (DfN)-based analysis of measurable residual disease (MRD) in patients with AML. *Oncol Res Treat.* 2019;42:87–88.
32. Röhnert M, von Bonin M, Kramer M, Ensel P, Holtschke N, Röllig C, et al. Identification of prognostic immunophenotypes at first diagnosis in patients with acute myeloid leukemia (AML) by a standardized multicolor flow cytometry (MFC) panel originally designed to detect measurable residual disease (MRD) at follow-up. *Blood.* 2020;136:35.
33. Heuser M, Freeman SD, Ossenkoppele GJ, Buccisano F, Hourigan CS, Ngai LL, et al. 2021 update measurable residual disease in acute myeloid leukemia: european leukemianet working party consensus document. *Blood.* 2021;138:2753–67.
34. 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:e66.
35. 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;30:1769–72.
36. Van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, Van Der Velden VHJ, Flores-Montero J, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia.* 2012;26:1908–75.
37. Gorczyca W. *Flow Cytometry in Neoplastic Hematology: Morphologic-Immunophenotypic Correlation* (2nd ed.). (CRC Press, 2010).
38. Köhnke T, Sauter D, Ringel K, Hoster E, Laubender RP, Hubmann M, et al. Early assessment of minimal residual disease in AML by flow cytometry during aplasia identifies patients at increased risk of relapse. *Leukemia.* 2015;29:377–86.
39. Craddock C, Jackson A, Loke J, Siddique S, Hodgkinson A, Mason J, et al. Augmented reduced-intensity regimen does not improve postallogeic transplant outcomes in acute myeloid leukemia. *J Clin Oncol.* 2021;39:768–78.
40. Mccarthy N, Loke J, Andrew G, Jackson A, Hodgkinson A, Mason J et al. Validation and application of an unsupervised analysis approach to measurable residual disease testing in acute myeloid leukemia. *EHA Libr.* 2021;06/09/21:EP432.
41. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Measures.* 2007;1:77–89.
42. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control.* 1974;19:716–23.
43. Portet S. A primer on model selection using the Akaike Information Criterion. *Infect Dis Model.* 2020;5:111–28.
44. Platzbecker U, Middeke JM, Sockel K, Herbst R, Wolf D, Baldus CD, et al. Measurable residual disease-guided treatment with azacitidine to prevent haematological relapse in patients with myelodysplastic syndrome and acute myeloid leukaemia (RELAZA2): an open-label, multicentre, phase 2 trial. *Lancet Oncol.* 2018;19:1668–79.
45. Feller N, van der Pol MA, 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;18:1380–90.
46. Ngai LL, Kelder A, Janssen JJWM, Ossenkoppele GJ, Cloos J. MRD tailored therapy in AML: what we have learned so far. *Front Oncol.* 2021;10:603636.
47. Buccisano F, Maurillo L, Gattei V, Del Poeta G, Del Principe MI, Cox MC, et al. The kinetics of reduction of minimal residual disease impacts on duration of response and survival of patients with acute myeloid leukemia. *Leukemia.* 2006;20:1783–9.
48. 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;36:636–49.
49. Al-Mawali A, Gillis D, Hissaria P, Lewis I. Incidence, sensitivity, and specificity of leukemia-associated phenotypes in acute myeloid leukemia using specific five-color multiparameter flow cytometry. *Am J Clin Pathol.* 2008;129:934–45.
50. 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;94:528–38.
51. Haferlach T, Schoch C, Schnittger S, Kern W, Löffler H, Hiddemann W. Distinct genetic patterns can be identified in acute monoblastic and acute monocytic leukaemia (FAB AML M5a and M5b): a study of 124 patients. *Br J Haematol.* 2002;118:426–31.
52. Wells DA, Benesch M, Loken MR, Vallejo C, Myerson D, Leisenring WM, et al. Myeloid and monocytic dyspoiesis as determined by flow cytometric scoring in myelodysplastic syndrome correlates with the IPSS and with outcome after hematopoietic stem cell transplantation. *Blood.* 2003;102:394–403.
53. Ravandi F, Jorgensen J, Borthakur G, Jabbour E, Kadia T, Pierce S, 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.
54. Lacronique-Gazaille C, Chaury MP, Le Guyader A, Faucher JL, Bordessoule D, Feuillard J. A simple method for detection of major phenotypic abnormalities in myelodysplastic syndromes: Expression of CD56 in CMML. *Haematologica.* 2007;92:859–60.
55. Subira D, Font P, Villalón L, Serrano C, Askari E, Góngora E, et al. Immunophenotype in chronic myelomonocytic leukemia: is it closer to myelodysplastic syndromes or to myeloproliferative disorders? *Transl Res.* 2008;151:240–5.
56. Feng R, Bhatt VR, Fu K, Pirruccello S, Yuan J. Application of immunophenotypic analysis in distinguishing chronic myelomonocytic leukemia from reactive monocytosis. *Cytometry B Clin Cytom.* 2018;94:901–9.
57. Zhou Y, Moon A, Hoyle E, Fromm JR, Chen X, Soma L, et al. Pattern associate-dleukemia immunophenotypes and measurable disease detection in acute myeloidleukemia or myelodysplastic syndrome with mutated NPM1. *Cytometry B Clin Cytom.* 2019;72:67–72.
58. Gallacher L, Murdoch B, Wu DM, Karanu FN, Keeney M, Bhatia M. Isolation and characterization of human CD34-Lin- and CD34+Lin- hematopoietic stem cells using cell surface markers AC133 and CD7. *Blood.* 2000;95:2813–20.
59. Vosberg S, Greif PA. Clonal evolution of acute myeloid leukemia from diagnosis to relapse. *Genes Chromosomes Cancer.* 2019;58:839–49.
60. Van Lochem EG, Van Der Velden VHJ, Wind HK, Te Marvelde JG, Westerdaal NAC, Van Dongen JJM. Immunophenotypic differentiation patterns of normal hematopoiesis in human bone marrow: Reference patterns for age-related changes and disease-induced shifts. *Cytometry B Clin Cytom.* 2004;60:1–13.
61. Soerensen JF, Aggerholm A, Kerndrup GB, Hansen MC, Ewald IKL, Bill M, et al. Clonal hematopoiesis predicts development of therapy-related myeloid neoplasms post-autologous stem cell transplantation. *Blood Adv.* 2020;4:885–92.
62. Loghavi S, DiNardo CD, Furudate K, Takahashi K, Tanaka T, Short NJ, et al. Flow cytometric immunophenotypic alterations of persistent clonal haematopoiesis in remission bone marrows of patients with NPM1-mutated acute myeloid leukaemia. *Br J Haematol.* 2021;192:1054–63.
63. Zhou Y, Wood BL. Methods of detection of measurable residual disease in AML. *Curr Hematologic Malignancy Rep.* 2017;12:557–67.
64. Hourigan CS, Dillon LW, Gui G, Logan BR, Fei M, Ghannam J, et al. Impact of conditioning intensity of allogeneic transplantation for acute myeloid leukemia with genomic evidence of residual disease. *J Clin Oncol.* 2020;38:1273–83.
65. Venditti A, Piciocchi A, Candoni A, Melillo L, Calafiore V, Cairoli R, et al. GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. *Blood.* 2019;134:935–45.

66. Ko BS, Wang YF, Li JL, Li CC, Weng PF, Hsu SC, et al. Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome. *EBioMedicine*. 2018;37:91–100.
67. Lacombe F, Lechevalier N, Vial JP, Béné MC. An R-derived FlowSOM process to analyze unsupervised clustering of normal and malignant human bone marrow classical flow cytometry data. *Cytom A*. 2019;95:1191–7.

## ACKNOWLEDGEMENTS

The authors thank the other members of the HARMONIZE consortium for their support: Marion Subklewe, Simon Völkl, Cornelia Brendel, Benjamin Tast, and Michael Rieger. We acknowledge the assistance of Helena Jambor in preparation of the manuscript. The authors thank Jana Bornhäuser, Cornelia Hoffmann, Claudia Klotsche, and Nadja Kubitz for their excellent technical assistance, Annett Engmann and Katrin Peschel for their contribution to project management, and all SAL centers for their commitment for the registry.

## AUTHOR CONTRIBUTIONS

UO, MvB, and MR designed the research and analyzed data. MR, JS, and PE wrote the paper. MK performed the bioinformatical analysis. NM and SF provided constructive comments and discussion. VB trained UO, MvB and MR for the conventional LAIP approach. SK provided technical support for the conventional LAIP approach. RS and CT provided molecular MRD data. SK, VB, JH, SJ, CR, and MB provided critical feedback and helped to shape the research, analysis and manuscript.

## FUNDING

This work was supported by Deutsche Forschungsgesellschaft (PERDAM, #318488004), Technische Universität Dresden (MeDDrive, intramural funding, #60466) and Wilhelm Sander-Stiftung (Minimal, #2021.035.1). Open Access funding enabled and organized by Projekt DEAL.

## COMPETING INTERESTS

CT is CEO and part ownership of AgenDix GmbH, a company performing molecular diagnostics. All other authors have nothing to disclose.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41375-022-01647-5>.

**Correspondence** and requests for materials should be addressed to Maximilian A. Röhnert.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022