

Comprehensive genetic diagnosis of acute myeloid leukemia by next-generation sequencing

Elisabeth K.M. Mack,¹ André Marquardt,^{1*} Danny Langer,^{1*} Petra Ross,¹ Alfred Ultsch,² Michael G. Kiehl,³ Hildegard I.D. Mack,⁴ Torsten Haferlach,⁵ Andreas Neubauer¹ and Cornelia Brendel¹

¹Department of Hematology, Oncology and Immunology, Philipps-University Marburg, and University Hospital Gießen and Marburg, Marburg, Germany; ²Databionics, Department of Mathematics and Informatics, Philipps-University Marburg, Germany; ³Department of Internal Medicine, Frankfurt (Oder) General Hospital, Frankfurt/Oder, Germany; ⁴Institute for Biomedical Aging Research, Leopold-Franzens-University Innsbruck, Austria and ⁵MLL, Munich Leukemia Laboratory, Germany

*These authors contributed equally to this work.



Haematologica 2019
Volume 104(2):277-287

ABSTRACT

Differential induction therapy of all subtypes of acute myeloid leukemia other than acute promyelocytic leukemia is impeded by the long time required to complete complex and diverse cytogenetic and molecular genetic analyses for risk stratification or targeted treatment decisions. Here, we describe a reliable, rapid and sensitive diagnostic approach that combines karyotyping and mutational screening in a single, integrated, next-generation sequencing assay. Numerical karyotyping was performed by low coverage whole genome sequencing followed by copy number variation analysis using a novel algorithm based on *in silico*-generated reference karyotypes. Translocations and DNA variants were examined by targeted resequencing of fusion transcripts and mutational hotspot regions using commercially available kits and analysis pipelines. For the identification of *FLT3* internal tandem duplications and *KMT2A* partial tandem duplications, we adapted previously described tools. In a validation cohort including 22 primary patients' samples, 9/9 numerically normal karyotypes were classified correctly and 30/31 (97%) copy number variations reported by classical cytogenetics and fluorescence *in situ* hybridization analysis were uncovered by our next-generation sequencing karyotyping approach. Predesigned fusion and mutation panels were validated exemplarily on leukemia cell lines and a subset of patients' samples and identified all expected genomic alterations. Finally, blinded analysis of eight additional patients' samples using our comprehensive assay accurately reproduced reference results. Therefore, calculated karyotyping by low coverage whole genome sequencing enables fast and reliable detection of numerical chromosomal changes and, in combination with panel-based fusion- and mutation screening, will greatly facilitate implementation of subtype-specific induction therapies in acute myeloid leukemia.

Introduction

Acute myeloid leukemia (AML) has recently been the object of thorough investigations at a molecular level, including whole genome next-generation sequencing (NGS) studies.¹ According to current guidelines from the European Leukemia Network, recommended genetic testing in adult AML is predominantly directed towards risk stratification in order to identify an appropriate strategy for consolidation therapy.² Genetic markers comprise t(8;21)(q22;q22.1)/*RUNX1-RUNX1T1*, inv(16)(p13.1;q22) or t(16;16)(p13.1;q22)/*CBFB-MYH11*, t(15;17)/*PML-RARA*, t(9;11)(p21.3;q23.3)/*MLL3-KMT2A*, other translocations involving the *KMT2A*

Correspondence:

brendelc@staff.uni-marburg.de

Received: May 26, 2018.

Accepted: September 5, 2018.

Pre-published: September 6, 2018.

doi:10.3324/haematol.2018.194258

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/104/2/277

©2019 Ferrata Storti Foundation

Material published in *Haematologica* is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>.

Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



gene, t(6;9)(p23;q34.1)/*DEK-NUP214*, inv(3)(q21.3;q26.2) or t(3;3)(q21.3;q26.2)/*GATA2*, *MECOM*, loss of chromosome 5/5q, 7, or 17/17p, mutations in *CEPBA* (biallelic), *NPM1*, *RUNX1*, *ASXL1* and *TP53*, and internal tandem duplications (ITD) in the *FLT3* gene.² Additionally, AML with *IDH2*^{R172} mutations alone and AML with mutations in chromatin regulators or splicing factors such as *DNMT3A*, *TET2*, *SRSF2* and *SF3B1* have been proposed recently as distinct genomic subclasses of AML.³

In spite of the considerable genetic heterogeneity of the disease, chemotherapy with cytarabine and anthracyclines has been the backbone of induction treatment for adults with AML throughout the last 30 years.^{4,5} Only acute promyelocytic leukemia with the hallmark translocation t(15;17)/*PML-RARA* has been shown to be highly curable by all-*trans* retinoic acid and arsenic trioxide.⁶ Immediate chemotherapy-free first-line treatment of acute promyelocytic leukemia is possible because this specific entity can be diagnosed within just a few hours by peripheral blood smear or bone marrow cytology and targeted reverse transcriptase polymerase chain reaction (PCR) analysis for *PML-RARA*. In contrast, discrimination of all other AML subtypes requires multiple molecular and cytogenetic analyses. In particular, AML karyotyping often requires shipping of samples to specialized laboratories, thus precluding completion within 5-7 days, as recommended.² Here, we developed and evaluated an integrated NGS platform for numerical karyotyping and focused screening for translocations and mutations in AML-related genes which enables fast identification of the majority of genetic alterations in AML with prognostic and therapeutic implications.

Methods

Patients and cell lines

The 33 patients' samples analyzed in this study were obtained from the repository of the Clinic for Hematology, Oncology and Immunology or the Munich Leukemia Laboratory with the patients' informed consent. Cell lines were purchased from DSMZ. This study was approved by the clinical ethics committee at the Faculty of Medicine, Philipps-University Marburg (N. 38/16).

Library preparation and sequencing

Whole genome libraries were constructed using the NEBNext[®] Ultra II kit (New England Biolabs, Ipswich, MA, USA). Fusion libraries were prepared using the FusionPlex[®] Heme v1 or v2 panels (ArcherDX, Boulder, CO, USA). Variant libraries were generated using the TruSight[®] Myeloid panel (Illumina, San Diego, CA, USA) or the QIAseq[™] Human Myeloid Neoplasms panel (Qiagen, Hilden, Germany). Sequencing was performed on an Illumina MiSeq instrument. The experimental procedures are detailed in the *Online Supplementary Methods*.

Bioinformatics

Copy number variations (CNV) were analyzed using a proprietary algorithm designated CAI[N] (chromosomal aberration identifier [numerical]) which was implemented in Python. Fusion calling was performed using the Archer[®] Analysis pipeline (version 4.1). For variant calling, the BaseSpace[®] TruSight[®] Myeloid App (Illumina) and ITD-seek⁷ or, respectively, smCounter⁸ (Qiagen) were used as appropriate. For identification of *KMT2A*-partial tandem duplications (PTD) in amplicon libraries, a novel Python algorithm (PTDi:

PTD identifier) was developed. Details are given in the *Online Supplementary Methods*.

Results

Design of an integrated next-generation sequencing platform for comprehensive genetic analyses in acute myeloid leukemia

Our NGS platform for comprehensive genetic characterization of AML samples was designed to enable fast and reliable detection of genetic aberrations that are of critical importance for diagnosis, prognosis and therapy in adult AML.^{1-3,9} The complete workflow including three NGS-library preparations and data analysis by five different algorithms can be completed within 5 days (Figure 1A).

In order to limit the required sequencing resources to the capacities of a benchtop sequencing device, we addressed AML-relevant translocations on the level of RNA using anchored multiplex PCR¹⁰ for targeted enrichment of chimeric transcripts. RNA-based detection of common gene fusions in AML and DNA-based mutational screening are already available through predesigned commercial kits (*Online Supplementary Tables S1-S4*) with associated analysis software. Thus, we included numerical karyotyping into our platform by a strategy that does not require specific target enrichment. In particular, we performed low coverage whole genome sequencing (lc-WGS), which has been shown previously to enable robust detection of CNV.^{11,12} For data analysis, we developed novel algorithms for the detection of CNV (Figure 1B) and *KMT2A*-PTD (CAI[N] and PTDi, respectively). Moreover, we added ITD-seek⁷ for the identification of *FLT3*-ITD to the available bioinformatics pipelines for fusion- or mutation calling. Depending on the size of the mutation panel, one or two samples can be analyzed at a time on a standard (maximum 15 x 10⁶ reads) flowcell (Figure 1C). Operational costs per sample are comparable to the total expenses for conventional cytogenetics, fluorescence *in situ* hybridization (FISH) and mutation analysis. Taken together, our integrated NGS approach rapidly and economically delivers clinically meaningful insights into AML genomes, opening up the possibility to inform treatment decisions early based on molecular features and calculated cytogenetic information.

Principles of the CAI[N] algorithm and stability of *in silico*-generated reference karyotypes

In order to facilitate clinical interpretation, we modified the concept of "virtual" or "digital" karyotypes^{13,14} and constructed "calculated karyotypes" from NGS data which resemble cytogenetic karyotypes. CNV in the range from cytogenetic bands to whole chromosomes are conveniently identified using a read depth approach¹⁵ and require only 5-10% genome coverage for detection with >90% sensitivity and specificity,^{16,17} corresponding to 1-2x10⁶ reads. CAI[N] compares read frequencies in 1 Mb fixed genomic windows to *in silico*-generated normal reference karyotypes and maps amplified/deleted regions to cytogenetic bands so that chromosomal gains or losses can be reported using cytogenetic notation (Figures 1B and 2A). Centromeres are not covered by our NGS karyotyping method as they include repetitive sequences that prevent unique alignment of sequencing reads.

To examine the stability of *in silico* reference karyotypes,

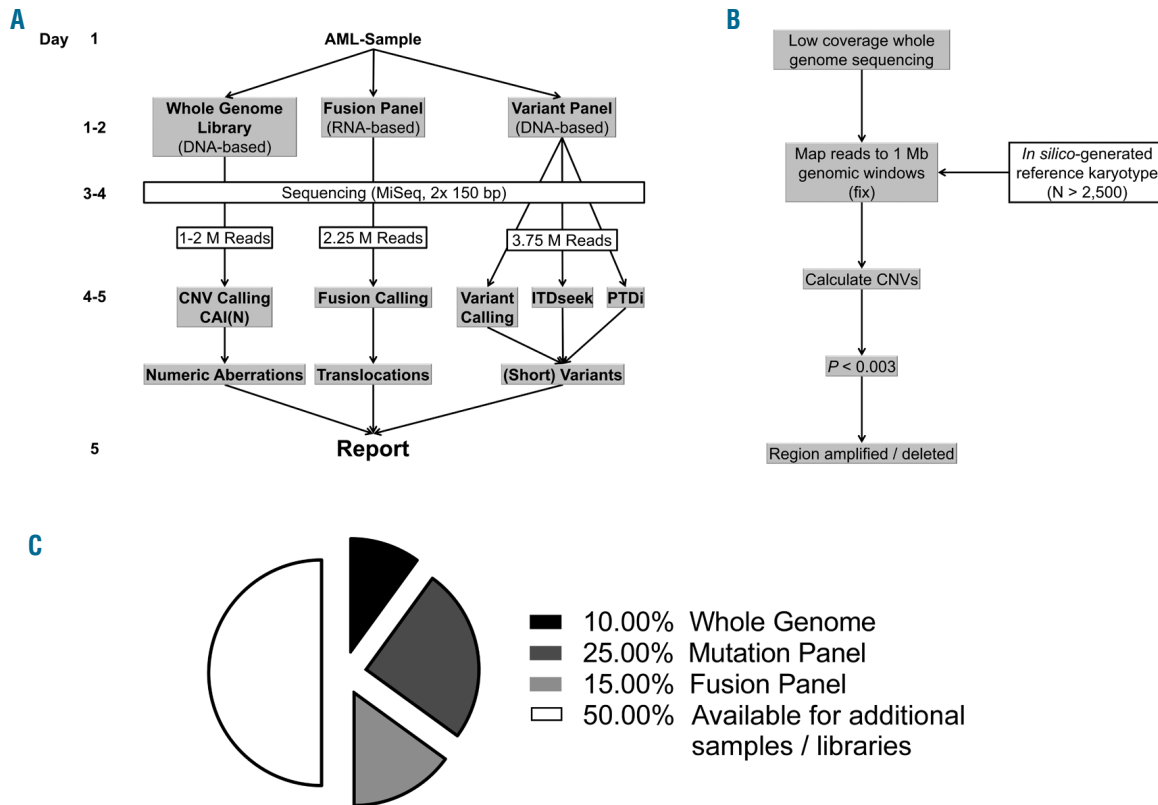


Figure 1. Comprehensive genetic diagnosis of acute myeloid leukemia by next-generation sequencing. (A) Outline of the workflow: each sample is subjected to preparation of three sequencing libraries. Libraries are indexed separately for sequencing on the same flowcell. Data are analyzed using five distinct algorithms for the detection of CNV, fusions, and DNA variants. The whole workflow can be completed within 5 days if performed by one person; times to perform individual steps of the composite assay are indicated on the right. (B) Outline of the CAI[N] algorithm for CNV analysis. Reads are mapped to 1 Mb fixed genomic windows and read distributions are compared to the average of more than 2,500 normal karyotypes ($N_{\text{female}}=2,819$, $N_{\text{male}}=2,605$) generated by random sampling of 150-250 bp reads from the reference genome. A region is called amplified or deleted if the observed read number in a window differs significantly ($P < 0.003$) from the average of *in silico*-generated karyotypes. (C) Flow cell occupancy by three sequencing libraries. Two samples can be analyzed in parallel in one sequencing run on a standard MiSeq v2 flowcell when libraries are sequenced with the read numbers indicated in (A).

we analyzed read distributions on whole chromosomes and in 1 Mb windows for random normal female and male karyotypes. Read frequencies showed very narrow variances and more reads mapped to autosomes in male karyotypes than in female ones, consistent with fewer reads mapping to the Y chromosome compared to a second X (Figure 2B). Of note, the Y chromosome appears smaller than its actual size, also due to repetitive sequences. To further investigate whether lc-WGS data resemble the results of *in silico* random experiments, we sequenced two libraries from healthy female donors at 1.4×10^6 reads. Read distribution patterns matched the *in silico* reference at all read depths examined (Figure 2C). These results confirm that lc-WGS can be accurately simulated computationally, allowing us to use random normal karyotypes as a stable reference for CNV analyses.

Detection of chromosomal gains and losses by copy number variation karyotyping

After evaluation of CAI[N] for consistency with normal karyotypes, we determined its capacity to detect numerical aberrations. First, we examined an individual with Down syndrome (T21) and the benign meningioma cell line BEN-MEN-1¹⁸ by lc-WGS and CAI[N] analysis. Both trisomy 21

in the T21 proband and loss of chromosome 22 in BEN-MEN-1 cells were identified correctly (Figure 3).

Next, we investigated deletions or additions of chromosome parts in three AML patients' samples exhibiting loss of the long arm of chromosome 5 (Table 1, *Online Supplementary Figures S1-S3*). CAI[N] recovered 5q deletions with different breakpoints that closely matched reference laboratory results (Figure 4A, Table 1). Moreover, a gain of chromosome 1p was detected in patient AML-1, consistent with a previously reported partial trisomy 1p (Figure 4B, *Online Supplementary Figure S1*, Table 1).

Finally, to test the capability of our approach to identify chromosomal gains or losses that are not readily detected by cytogenetic banding, we performed CNV karyotyping on two AML cell lines, HL-60 and NB-4. We observed complex patterns of copy number alterations in both cell lines, including massive overrepresentation of 8q24.21 (containing the *MYC* locus) with loss of the remaining parts of chromosome 8 (*Online Supplementary Figures S4A,B*, *S5* and *S6*, *Online Supplementary Tables S5* and *S6*), as described previously.¹⁹⁻²⁶ Amplification of *MYC* was validated by quantitative PCR in HL-60 and NB-4 cells and in patient AML-2, in whom a copy number gain of 8q24 had not been observed by cytogenetics. Similarly, quantitative

Table 1. Patients' samples and karyotypes.

Pat. ID	Age	FAB	Sample type	Blasts (Cyto)	Blasts (FACS)	Reference karyotype	CNV karyotype	Total reads	Mappable reads
AML-1 ¹	68	M2	BM	n.a.	n.a.	45,XX, +der(1)t(1;11)(q21;q13); del(5)(q13q33), del(7)(q22q36), -17,del(17)(p12)[26] FISH: partial trisomy 1 (1p), trisomy 11q	47,XX, add(1)(p36.32p21.3), del(5)(q14.3q34), del(7)(q22.1q36.3), +11, del(15)(q11.2q25.1), del(17)(p13.3p11.2), del(17)(q11.1q21.33)	4,994,165	3,958,833
AML-2	19	M2 (after ALL)	BM	90%	n.a.	45,Y,-X, del(5)(q15q31), del(9)(p24p22), del(17)(p13p11.2), -20, +mar [10]; 46,XY,idem, del(7)(q22q36)[11]	46,XY, del(5)(q14.3q35.3), add(8)(q24), del(9)(p22.1p21.2), del(17)(p13.3p11.1), del(20)(q11.22q13.33)	1,290,882	1,026,834
AML-3	71	M6	BM	60%	26%	46-49,XX,+X,+X der(3)t(3;17)(p12;?), del(5)(q13q33), +del(5)(q13q33), der(7)t(3;7)(?:p22), +10, der(11)amp(11q)t(11;20), der(17;22)(q10;q10), -18, der(20)t(20;22)[cp15]	47,XX,+X, del(3)(p21.33p14.3), del(5)(q14.3q34), add(11)(q13.4q25), del(17)(p13.3p13.1), -18, der(20)del(p12.3p12.1) add(p11.23q11.22) del(q11.23q13.31), +22	4,974,979	3,840,179
AML-4 ²	53	M2	PB	100% CD34 -pos. fraction		n.a	46,XX, del(7)(q11.22q36.3)	1,537,585	1,216,416
AML-5	34	M4	BM	>90%	47%	46,XX,t(6;11)(q27;q23)[19]; 46,XX[1]	46,XX	2,078,039	1,634,883
AML-6 ³	36	M4	PB	n.a.	30%	n.a	46,XX	2,079,920	1,646,231
AML-7	82	M5	BM	30%	40%	46,XX,t(X;16)(p11;q21)[19]; 46,XX [1]	46,XX	2,652,806	2,113,075
AML-8	57	M0	BM	80%	83%	46,XX [15]	46,XX	1,180,039	933,365
AML-9	65	M1	PB	90%	86%	46,XY[20]	46,XY	1,980,007	1,554,048
AML-10	61	M2	BM	80%	91%	46,XX, der(10)t(10;11)(p12;q13) inv(11)(q13q23), der(11)t(10;11)(p12;q13) [20]; nuc ish 11q23(5'MLLx2,3'MLLx2) (5'MLL sep 3'MLLx1) [60], nuc ish 11q23(5'MLLx2,3'MLLx2) (5'MLL sep 3'MLLx0) [40]	46,XX	1,741,918	1,379,261
AML-11a	74	M4	BM	65%	76%	46,XX, t(8;21)(q22;q22) [4]; 47,XX, +4, t(8;21)(q22;q22) [2]	47,XX,+4	1,722,984	1,363,640
AML-11b	74	M4	PB	90%	20%	n.a	47,XX,+4	1,690,417	1,339,197
AML-12 ⁴	72	M5	BM	30%	20%	n.a	46,XY, del(7)(q21.1q36.1), add(13)(q11q14.2)	1,798,644	1,408,808
AML-13	43	M4eo	PB	59%	44%	46,XY, inv(16)(p13q22) [20]	46,XX	2,170,329	1,699,480
AML-14	53	M5a (t-AML)	PB	n.a.	n.a.	n.a. (female pat. after allo-HSCT from male donor; chimerism 100%)	46,XY	1,627,409	1,273,816
AML-15	70	M2	BM	90%	98%	n.a (NPM1_A, FLT3-ITD)	46,XY	1,622,690	1,274,059
AML-16	58	M2	BM	50%	60%	46,XX,t(8;21)(q22;q22) [6], 45,X,-X,t(8;21)(q22;q22) [12], 46,XX [2]	45,X,-X	2,053,046	1,629,435
AML-17	52	M1/ M4/ M2	BM	n.a	51%	46,XX [21]	46,XX	2,394,150	1,886,189

continued on the next page

continued from the previous page

AML-18 ^s	45	M5	PB	89%	97%	n.a	46,XY	2,037,251	1,604,338
AML-19	68	M3v	BM	80%	90%	46,XY, t(15;17)(q22;q12)[2]; 46,XY,t(1;6;7) (p36;p21;q22), t(15;17)(q22;q12)[11]	46,XY	2,241,454	1,772,271
AML-20	61	n.a	BM	100%	98%	n.a (FLT3-ITD, NPM1_D, DNMT3A R882H, IDH1 R132H)	46,XY	3,603,346	2,838,440
AML-21	18	M2	PB	n.a.	>30%	46,XX, del(7q),t(8;21)(q22;q22)	46,XX,del(7)(q33q36.2)	1,104,847	875,403
AML-22	31	M5	BM	n.a.	58%	48-49,XY, +19, +der(21)X1-2[cp4]; 46,XY[7]	48,XY,+19,+21	2,042,269	1,602,693
AML-23	48	M4eo	PB	80%	63%	46,XY [6]; 46,XY, inv. (16)(p13q22) [18]	46,XY	1,940,532	1,523,373
AML-24	41	RAEB-t	BM	10%	11%	45,XY,-7 [7], 46,XY,-7 +21 [9], 46,XY [5]	46,XY,-7,+21	1,661,351	1,297,049
AML-25	57	M1	PB	70%	60%	46,XX [20]	46,XX	2,326,960	1,848,467
AML-26	72	M0/M1	BM	n.a.	n.a.	nuc ish 5p15(htERTx3), 5q31(CDC25C,EGR1x3), 8cen(D8Z2x4), 17p13(TP53x3), 17q11(NF1x3)	poor DNA quality; evidence for del(5p), del(5)(q31), add(8p), +17	920,209	643,682
AML-27	66	M4	BM	n.a.	n.a.	45,XY,-7[9]; 45,XY,-7.ish del(12)(p13p13) (3'ETV6-) [3]; 45,XY,-7.ish del(12)(p13p13) (3'ETV6-,5'ETV6-) [6], 46,XY[4]	poor DNA quality/ too few reads; evidence for, -7, del(12p13.2)	590,092	472,009
AML-28	54	M5a	BM	n.a.	n.a.	46,XY,t(9;11)(p22;q23)	47,XY, ⁺¹⁹	867,272	687,922
CML-1	74	CML	PB	n.a.	n.a.	46,XX, t(9;22)(q34;q11) [20]	n.d	n.a.	n.a.
HES-1	73	HES	BM	n.a.	n.a.	46,XY [15]	46,XY	2,072,387	1,636,629
ALL-1	62	Common B-ALL	BM	80%	70%	47,XX, t(1;9)(q22;q34),+17 [3], 46,XX,t(1;9)(q22;q34),der(6) t(6;17)(p23;q21) [2]	n.d.	n.a.	n.a.

^sSample AML-1: chromosome 15 was not covered by FISH analyses performed by the reference laboratory (probes included WCP1/WCP18,WCP7,EGR1,WCP5,ATM,p53,IGH BCL2).
⁵Sample AML-4: cytogenetic analyses were not performed at the time the sample was taken for NGS karyotyping: 12 months earlier (BM): 45,X,der(X;7)(q10;p10)[7]; 46,XX,add(7)(q21)[6],46,XX [3],46,XY[4].nuc ish Xp11.1-q11.1(DXZ1x2),Yq12(DYZ1x0)[50]/ Xp11.1-q11.1(DXZ1x1),Yq12(DYZx1)[50],7cen(D7Z1x2),7q31(DS486x1)[31/100]; 3 months later (pB): 45,der(X;7)(q10;p10)[2], 46,XX,der(7)t(X;7)(?q27;q11)[3], 46,XX,der(7)t(7;18)(p13;q22)t(X;7)(q27;q11),der(18)t(7;18)(p13;q22)[13],46,XX[3]. nuc ish 7cen(D7Z1x2),7q31(DS486x1)[99/100]; Read counts (total/mapped) are for the 100% CD34-positive fraction, see *Online Supplementary Figure S7*.
⁶Sample AML-6: the sample was taken after third-line treatment, which led to a partial remission. Karyotype before treatment (presumed persisting clones in bold; see also *Online Supplementary Table S7*): **46,XX,inv(3)(q21q26)[1]**/idem,der(5)t(5;16)(q13;q21),der(12)t(12;15)(p12;q21),-15,-16[11]/**46,idem,t(12;22)(p12;q12)[8]**.
⁷Sample AML-12: cytogenetic analyses were not performed at the time the sample was taken for NGS karyotyping: approximately 6 weeks earlier: 46,XY [5], nuc ish 5p15(htERTx2),5q31(CDC25C,EGR1x2)[100], 7cen(D7Z1x2),7q31(D7S486x2)[100],13q14(DLEUx3),17p13(TP53x2)[20/100].
⁸Sample AML-18: cytogenetic analyses were not performed at the time the sample was taken for NGS karyotyping (relapse): approximately 15 months earlier at primary diagnosis: normal karyotype 46,XY,NPM1 mut,FLT3-ITD positive. BM: bone marrow; PB: peripheral blood.

PCR analysis of three loci on chromosome 7q (*ARHGEF5*, *PIK3CG*, *VKORC1L1*) confirmed that this region was amplified in NB-4 cells and not lost in AML-2 (*Online Supplementary Figures S2* and *S4C*, Table 1).

In summary, our results demonstrate that lc-WGS followed by CAI[N] analysis correctly identifies copy number changes with high resolution and allows specific genes to be linked directly to amplified or deleted regions.

Sensitivity of CAI[N] copy number variation karyotyping

To test the sensitivity of our karyotyping approach, we performed lc-WGS and CAI[N] analysis on a dilution series

of BEN-MEN-1 DNA in healthy donor DNA. Moreover, we investigated samples with different blast contents which were prepared after enrichment of CD34-positive cells from the peripheral blood of patient AML-4 by magnetic bead separation.²⁷ Loss of chromosome 22 was readily detectable in mixtures containing as little as 10% BEN-MEN-1 DNA (Figure 5). Deletion of chromosome 7q was recovered by CNV karyotyping for blast contents $\geq 20\%$ with almost identical breakpoints (*Online Supplementary Figure S7*, Table 1). As loss of chromosome 7q was not present in all cells in the CD34-positive blast population (Table 1), the detection limit for this aberration was slightly higher than for monosomy

22, which is found in all BEN-MEN-1 cells. These findings indicate that the sensitivity of CAI[N]-CNV karyotyping is sufficient to detect highly prevalent chromosomal aberrations in AML samples with a blast count of at least 20% without prior enrichment of the blast population.

Fusion gene and DNA variant detection

Our composite assay relies on predesigned amplicon panels for the detection of fusion transcripts and DNA variants. Amplification strategies implemented in these kits¹⁰ or, respectively, specific panels, have already been extensively evaluated.^{7,28} Thus, we focused our studies on the detection of all subclass-defining translocations and all major types of clinically relevant DNA variants in adult AML. To investigate coverage of important fusion genes, we analyzed cell lines and patients' samples harboring or lacking frequent chimeric transcripts. All expected fusions were identified in KASUMI-1,²⁹ ME-1,³⁰ NB-4,²¹ AML-5, AML-6 and CML-1, including two variants³¹ of *BCR-ABL1* in the last sample. On the other hand, no fusions were detected in HL-60¹⁹ cells and in a patient with hyper-

osinophilic syndrome (HES-1), as reported by the reference laboratory (*Online Supplementary Figure S8A, Online Supplementary Table S7*). In a pool of the four cell lines, all fusions were recovered, but in a 1:25 dilution thereof, only the *RUNX1-RUNX1T1* fusion transcript was identified. This finding underlines that RNA-based fusion detection is expression-dependent, so that the sensitivity of the assay varies for different samples and fusions.

Moreover, we exemplarily tested the TruSight[®] Myeloid panel (Illumina) and the QIASeq[™] Myeloid Neoplasms panel (Qiagen), which incorporates molecular barcodes for PCR-error correction,³² as screening tools to identify short DNA variants in AML genomes. All single nucleotide variants detected in HL-60, NB-4, ME-1, MV4-11 and SKNO-1 cells by the TruSight[®] Myeloid panel were consistent with COSMIC³³ data or confirmed by Sanger sequencing, and the QIASeq[™] panel uncovered all reported mutations in samples from two patients with AML (*Online Supplementary Table S8*). Sequencing a dilution series of MV4-11 DNA revealed detection limits for the two p53 mutations of 1% and 10% with the TruSight[®] and

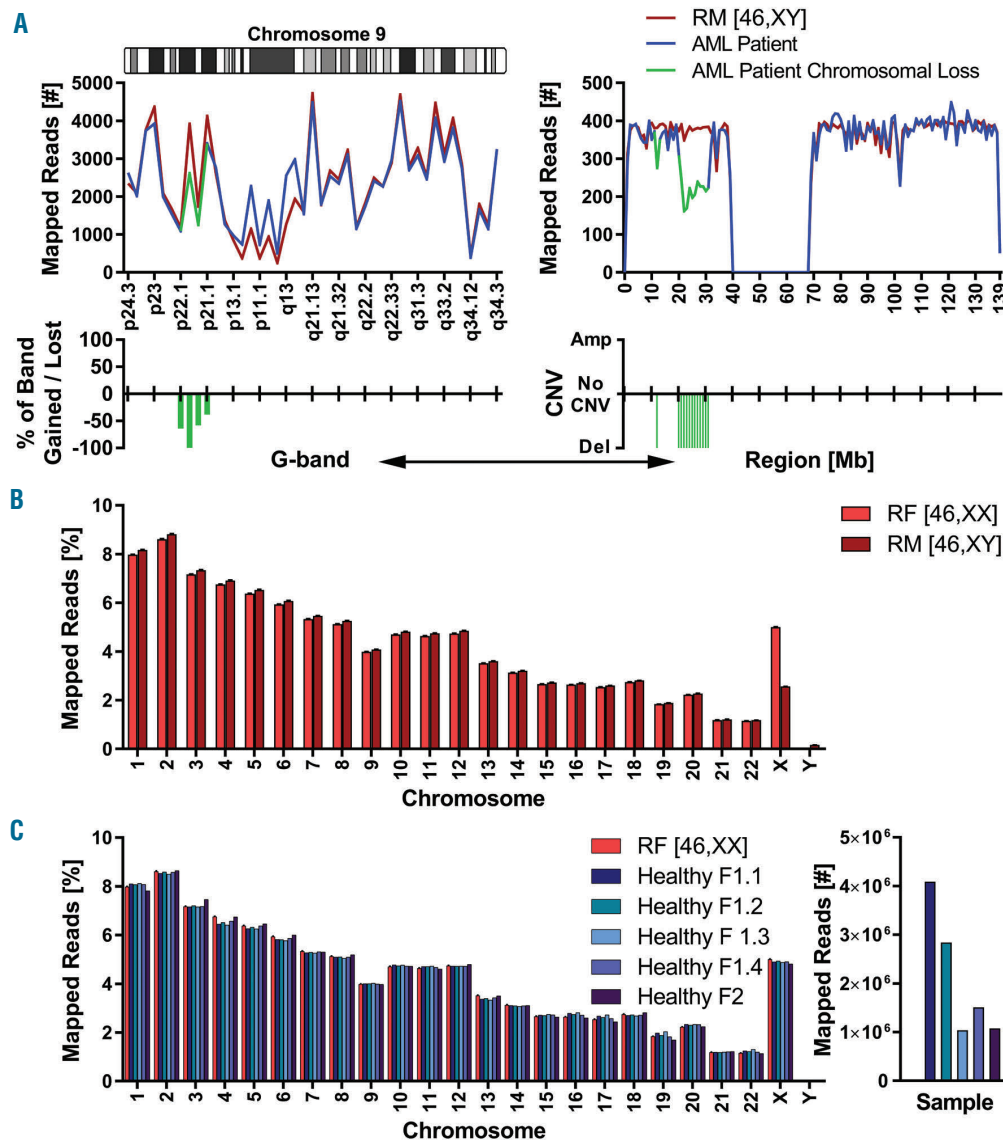


Figure 2. Calculated chromosome banding and *in silico*-generated reference karyotypes. (A) Calculated chromosome banding by CAI[N] analysis of lc-WGS data. Read distributions in genomic windows along chromosome 9 of an AML patient (AML-2, Table 1) are indicated (left: cytogenetic bands, right: 1 Mb windows). (B) Frequencies of uniquely mapped reads on whole chromosomes for *in silico*-generated normal karyotypes. RF: random female (N=2,819), RM: random male (N=2,605). Error bars represent the standard deviation (below visibility; <0.01%). Note that in (A) the centromere of a chromosome is not covered and in (B) the Y chromosome appears smaller than its actual size because of repetitive DNA sequences, which prevent unique alignment of sequencing reads. (C) Scalability of the CAI[N] algorithm: Four whole genome libraries from two healthy female donors were sequenced with different read numbers in multiplexed sequencing runs (right panel). Healthy F1.1-4: four runs of the same library.

QiaSeq™ panels, respectively (Online Supplementary Figure S8B). Therefore, amplicon sequencing enables variant detection with analytical sensitivities that, in well-covered regions, are equivalent to those of Sanger sequencing, the reference method for clinical mutation testing.^{34,35}

Larger DNA sequence variants such as *FLT3*-ITD and *KMT2A*-PTD require special attention in NGS data analysis as they may be missed by common variant calling tools.^{36,37} Thus, previous authors have added ITD-seek to the TruSight® Myeloid analysis pipeline.⁷ Applying this tool to our MV4-11 dilution series, we identified a 34 bp insertion in *FLT3* exon 14, which closely matched published results,³⁸ with an analytical sensitivity of 10% (Online Supplementary Figure S9A). The *FLT3*-ITD was not detected in the 10% MV4-11 sample using the QiaSeq™ panel with smCounter analysis, presumably because of suboptimal coverage achieved in our sequencing runs. However, given that our work with commercial kits aimed at clarifying their principal applicability for diagnostic purposes, we did not repeat these experiments.

For the identification of *KMT2A*-PTD in amplicon sequencing data, we developed PTDi by adapting a tool that had been used previously with a capture-based targeted sequencing approach.³⁶ PTDi analysis revealed amplification of exons 3-8 in patient AML-7 with a known e3e9 *KMT2A*-PTD (Online Supplementary Figure S9B, Online

Supplementary Table S9). Thus, not only are short DNA variants detectable by amplicon sequencing, but also difficult ITD and PTD. Taken together, our results clearly confirm that combining RNA- and DNA-based amplicon panels allows all major translocations and all types of clinically relevant mutations in AML to be uncovered by NGS.

Evaluation of the comprehensive next-generation sequencing platform for the diagnosis of acute myeloid leukemia in a clinical setting

After testing the performance of all sequencing modules and bioinformatics procedures, we evaluated the clinical utility of our platform as a diagnostic tool. As our karyotype studies in cell lines clearly showed that NGS detects a higher number of numerical aberrations than chromosome banding (Online Supplementary Tables S5 and S6), we first investigated a potential need for manual review of the raw CAI[N] output in order to avoid overestimation of karyotype complexity and enable appropriate risk stratification (Table 1, Online Supplementary Tables S10 and S11). We performed lc-WGS on additional patients' samples and reconstructed CNV karyotypes by cross comparison of CAI[N] results and known cytogenetic findings. All non-complex karyotypes, including three samples with presumably normal karyotypes in which cytogenetic analysis had failed, were identified correctly (AML-5, -7-11a/b, -13,

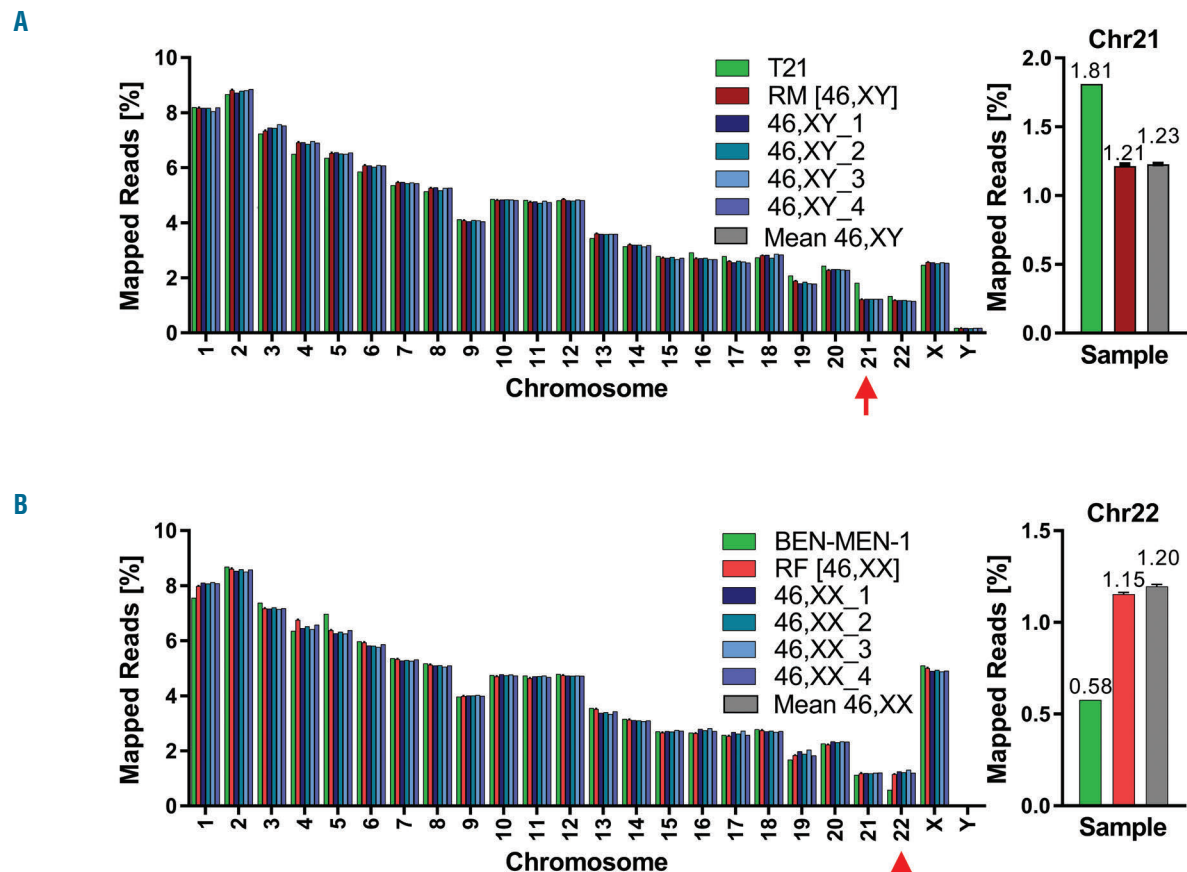


Figure 3. Detection of whole chromosome gains and losses by copy number variation karyotyping. Whole genome libraries from (A) an individual with Down syndrome (T21) and (B) the BEN-MEN-1 cell line were sequenced with low coverage and analyzed by CAI[N]. RF: random female (N=2,819), RM: random male (N=2,605). Error bars represent the standard deviation (below visibility).

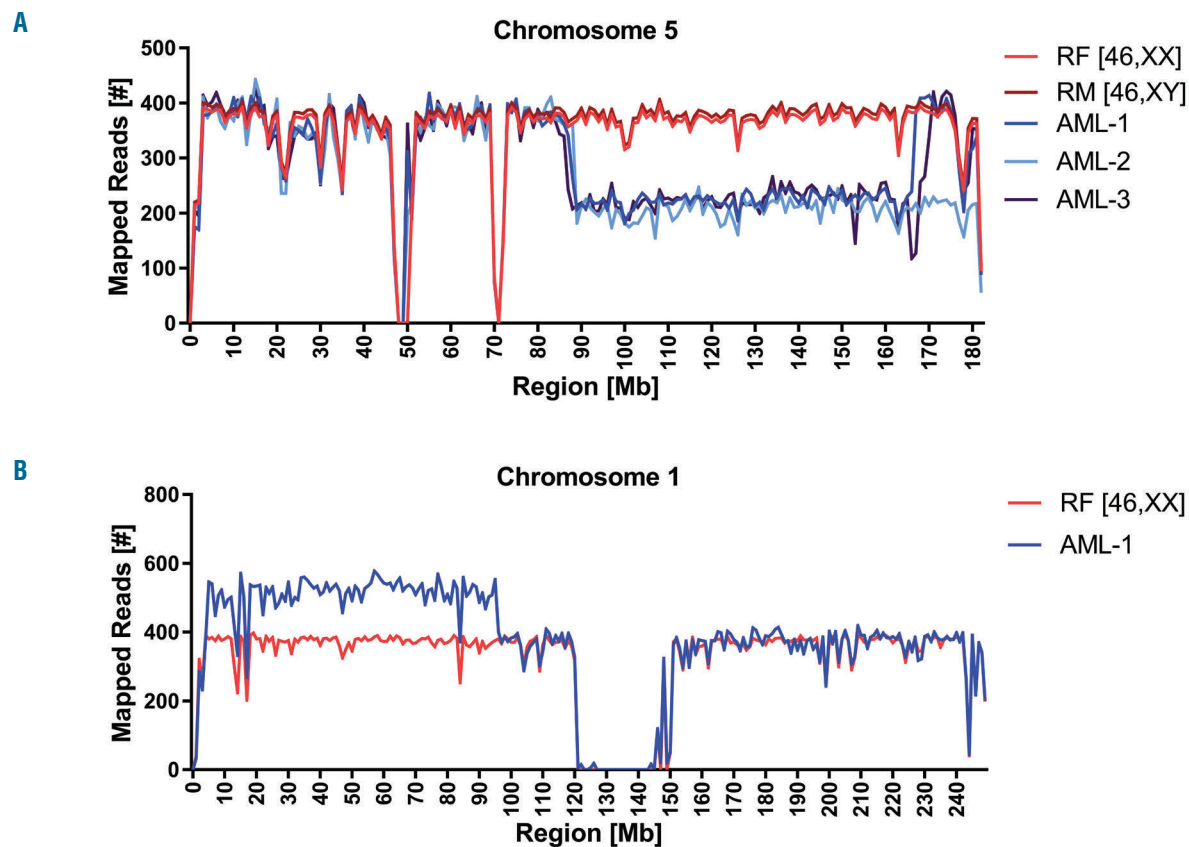


Figure 4. Detection of partial chromosome losses and gains by copy number variation karyotyping. Whole genome libraries from three AML patients' samples were sequenced with low coverage and analyzed by CAI[N]. (A) Region plots for chromosome 5. (B) Region plot of chromosome 1 for patient AML-1. Read numbers in 1 Mb windows were normalized to 1×10^6 total reads. RF: random female (n=2,819), RM: random male (n=2,605). See also *Online Supplementary Figures S1-S3*.

-16, -17, -19, HES-1 / AML-14, -15, -20). In two samples that had not been characterized extensively by FISH, CNV karyotyping apparently identified marker chromosomes or detected additional aberrations (AML-2 / AML-1). Moreover, CAI[N] revealed copy number changes for five of six chromosomes involved in translocations in a highly complex karyotype (AML-3), but missed loss of chromosome 10, which had been detected in <10% of cells by interphase FISH. In four patients, for whom cytogenetics had not been performed when samples were taken, NGS recovered at least a subset of aberrant clones or a normal karyotype as reported at initial diagnosis (AML-4, -6, -12 / AML-18, respectively). Taken together, overall karyotype complexity was determined correctly in all cases and 13/13 samples without risk-defining translocations were accurately assigned to prognostic groups based on CNV karyotyping alone (*Online Supplementary Table S10*). Thus, we did not specifically validate discordant results between conventional and NGS karyotyping.

Next, to study patients' samples in an unbiased manner, a group of four of the authors performed blinded analysis of CNV, fusion genes and mutations on eight additional AML samples and one acute lymphocytic leukemia sample (AML-21-28, ALL-1,³⁹ Table 1, *Online Supplementary Tables S7-S9*). Two samples were excluded before unblinding because of insufficient read coverage resulting from poor DNA quality (AML-26, -27). In the remaining samples, CNV karyotyping uncovered all expected numerical

changes and one additional aberration, gain of chromosome 19 in patient AML-28. Fusion analysis identified all translocations previously described in these patients. Variant analysis revealed no single nucleotide variants, insertions, deletions or ITD in this series of samples, in agreement with reference results. These findings further underscore the potential diagnostic value of our assay for the clinical management of AML.

Discussion

Here we developed and validated an integrated diagnostic platform that exclusively utilizes state-of-the-art NGS technology to obtain comprehensive clinically relevant insights into AML genomes. Our targeted approach covers all genetic features that define subclasses of AML with recurrent genetic abnormalities and/or prognostic groups as well as potentially actionable mutations^{2,9} and thus can serve as a potential alternative for diverse classical cytogenetics and molecular biology assays in certain laboratory settings. In contrast to a previously described single-run NGS assay for AML diagnosis,⁴⁰ we did not include detection of copy number neutral loss of heterozygosity,^{41,42} as its prognostic value in AML⁴³ is not fully established.²

We examine numerical aberrations, translocations, short sequence variants, *FLT3*-ITD and *KMT2A*-PTD in a rapid, robust and reliable composite assay. Our platform uses a

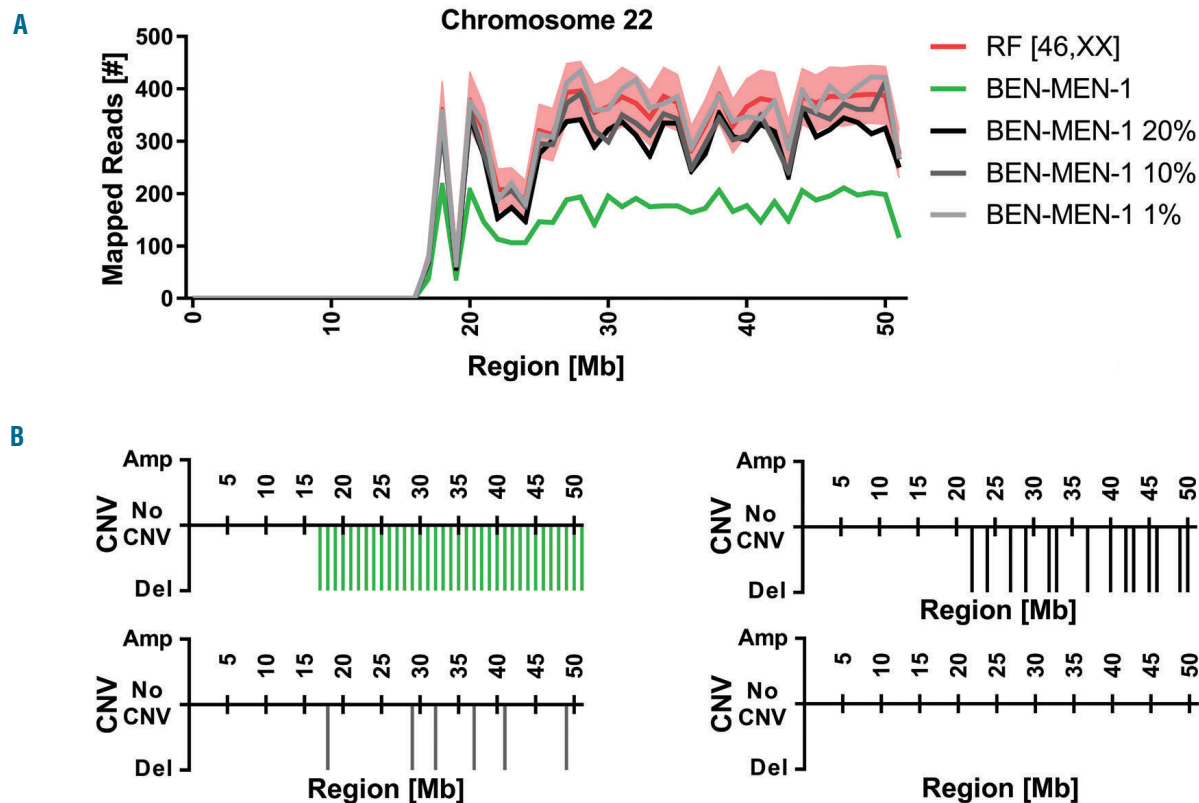


Figure 5. Sensitivity of copy number variation karyotyping. Genomic DNA from the BEN-MEN-1 cell line (monosomy 22) was diluted in healthy donor DNA (Healthy F1, Figure 2) in different ratios and subjected to lc-WGS and CAI[N] analysis. (A) Region plots for chromosome 22. The range ± 3 standard deviations around the mean is indicated in pale red. (B) CNV decision plots. Read numbers in 1 Mb windows were normalized to 1×10^6 total reads. RF: random female (N=2,819). Color coding in (B) as in (A).

benchtop sequencing device and commercially available reagents and kits, thus facilitating implementation even at smaller diagnostic centers that currently cannot offer the full spectrum of molecular and cytogenetic analyses required for a complete genetic work-up of AML. NGS results differ from diagnostic reports obtained by current standard techniques in two major aspects: first, NGS karyotyping by lc-WGS and fusion transcript analysis does not reveal insights into the potential clonal heterogeneity of the disease, and second, DNA-based mutation testing yields variant allele frequencies rather than allelic ratios. While the observed frequency of cytogenetic abnormalities in a certain number of metaphases is not relevant for risk stratification, mutational burden is highly prognostically relevant for *FLT3*-ITD. Allelic ratios calculated from fragment analysis are not equivalent to variant allele frequencies, but based on our data presented here we cannot propose a conversion from variant allele frequencies to the established risk stratification parameter.

In the method described here, sequencing resources are most effectively reduced by fusion detection on the level of RNA using anchored multiplex PCR for target enrichment. Anchored multiplex PCR requires previous knowledge on only one fusion partner, which is targeted by a gene-specific primer included in the sequencing panel,¹⁰ so that even novel translocations involving commonly rearranged genes can be detected using a relatively low-complexity primer pool. Moreover, if a translocation does not result in a fusion transcript, but rather in transcriptional activation of

a target gene, such as *MECOM* overexpression in AML with *inv(3)(q21.3;q26.2)* or *t(3;3)(q21.3;q26.2)*,⁴⁴ the underlying chromosomal rearrangement may be deduced from relative expression analyses via molecular barcode quantification. In contrast, translocation detection on the level of DNA needs to cover large intronic breakpoint regions and therefore requires higher throughput sequencing equipment or limitation of the assay to a subset of AML-relevant translocations.^{40,45,46}

Most importantly, our approach provides an easy to use tool for “numerical” karyotyping, that – differently from single nucleotide polymorphism-karyotyping methods⁴⁰ – enables CNV analysis in a completely unbiased manner without the need for specific capture probes. Notably, the CAI[N] algorithm does not uncover absolute ploidy, but classifies gains and losses even in strongly altered hypo- or hyperdiploid cases (e.g. HL-60 and NB-4 cells).

Despite mapping to chromosome bands, CNV karyotyping by lc-WGS and CAI[N] analysis identifies chromosomal gains and losses at higher resolution than conventional cytogenetics (1 Mb versus 5-10 Mb). Subcytogenetic CNV have been observed previously by array-based comparative genomic hybridization in both complex and normal karyotype AML, including recurrent aberrations with potential prognostic impact, such as gain of 8q24.^{32,41,47} Moreover, NGS karyotyping does not require short-term culture of the sample material and therefore eliminates a major technical challenge of classical cytogenetics and potential biases in clonally heteroge-

neous populations. For example, gain of chromosome 19 has been reported to be missed frequently by conventional banding techniques,⁴⁶ as observed here in patient AML-28. Nevertheless, we did not verify this finding by an alternative detection method. Given that NGS technology is increasingly used by diagnostic laboratories, the lc-WGS strategy described here opens up the possibility of performing high-resolution numerical karyotyping of AML samples on a routine basis. On the other hand, unlike classical chromosome banding or FISH analyses, CNV karyotyping combined with targeted resequencing of chromosomal fusions will not identify balanced translocations that are not covered by the sequencing panel. These features include some abnormalities defining AML with myelodysplasia-related changes, which is not, however, exclusively a genetic diagnosis.⁹ Hence, larger cohorts of AML patients will have to be examined in order to develop a new risk stratification system that incorporates the gain of information obtained by lc-WGS (or array-based comparative genomic hybridization) compared to cytogenetics and potential gaps of knowledge that may emerge with a particular CNV karyotyping strategy. Future studies also need to include more extensive validation of the sensitivity of our CNV karyotyping

assay to detect chromosomal gains and losses in regions that have not been specifically investigated here.

In summary, to the best of our knowledge, our work represents the fastest and most comprehensive analysis platform for diagnosing AML developed so far. Combining molecular genetics and cytogenetics in one NGS run will pave the way for differentiated management of AML patients not only in clinical trials, but also in a standard-of-care setting, as is currently emerging with midostaurin as the first targeted agent in induction therapy for patients with *FLT3* alterations.⁴⁹

Funding

This work was supported by grants from Rhön Klinikum AG (RKA n. 64 to EKMM, CB, MGK), Deutsche Forschungsgemeinschaft KFO210 (BR2232/2 to CB; Ne310/18-1 to AN), Deutsche José Carreras Leukämie-Stiftung (AH06-01 to AN, 18R/2016 to EKMM) and Stiftung PE Kempkes (06/2014 to EKMM).

Acknowledgments

We thank Lisa-Marie Koch, Ute Niebergall and Kathleen Stabla for excellent technical assistance and Prof. Edison T. Liu for helpful advice.

References

- Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059–2074.
- 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(4):424–447.
- Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med.* 2016;374(23):2209–2221.
- Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N Engl J Med.* 2015;373(12):1136–1152.
- Grimwade D, Ivey A, Huntly BJP, et al. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood.* 2016;127(1):29–41.
- Lo-Coco F, Avvisati G, Vignetti M, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med.* 2013;369(2):111–121.
- Au CH, Wa A, Ho DN, Chan TL, Ma ES. Clinical evaluation of panel testing by next-generation sequencing (NGS) for gene mutations in myeloid neoplasms. *Diagn Pathol.* 2016;11:11.
- Xu C, Nezami Ranjbar MR, Wu Z, DiCarlo J, Wang Y. Detecting very low allele fraction variants using targeted DNA sequencing and a novel molecular barcode-aware variant caller. *BMC Genomics.* 2017;18(1):5.
- Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood.* 2016;127(20):2391–2405.
- Zheng Z, Liebers M, Zhelyazkova B, et al. Anchored multiplex PCR for targeted next-generation sequencing. *Nat Med.* 2014;20(12):1479–1484.
- Liu S, Song L, Cram DS, et al. Traditional karyotyping vs copy number variation sequencing for detection of chromosomal abnormalities associated with spontaneous miscarriage. *Ultrasound Obstet Gynecol.* 2015;46(4):472–477.
- Heitzer E, Ulz P, Belic J, et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* 2013;5(4):30.
- Hagenkord JM, Parwani AV, Lyons-Weiler MA, et al. Virtual karyotyping with SNP microarrays reduces uncertainty in the diagnosis of renal epithelial tumors. *Diagn Pathol.* 2008;3:44.
- Wang T-L, Maierhofer C, Speicher MR, et al. Digital karyotyping. *Proc Natl Acad Sci U S A.* 2002;99(25):16156–16161.
- Pirooznia M, Goes FS, Zandi PP. Whole-genome CNV analysis: advances in computational approaches. *Front Genet.* 2015;6:138.
- Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics.* 2009;10:80.
- Li X, Chen S, Xie W, et al. PSCC: sensitive and reliable population-scale copy number variation detection method based on low coverage sequencing. *PLoS One.* 2014;9(1):e85096.
- Puttmann S, Senner V, Braune S, et al. Establishment of a benign meningioma cell line by hTERT-mediated immortalization. *Lab Invest.* 2005;85(9):1163–1171.
- Gallagher R, Collins S, Trujillo J, et al. Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia. *Blood.* 1979;54(3):713–733.
- Karnan S, Tsuzuki S, Kiyoi H, et al. Genomewide array-based comparative genomic hybridization analysis of acute promyelocytic leukemia. *Genes Chromosom Cancer.* 2006;45(4):420–425.
- Lanotte M, Martin-Thouvenin V, Najman S, Balerini P, Valensi F, Berger R. NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood.* 1991;77(5):1080–1086.
- Shiple J, Weber-Hall S, Birdsall S. Loss of the chromosomal region 5q11-q31 in the myeloid cell line HL-60: characterization by comparative genomic hybridization and fluorescence in situ hybridization. *Genes Chromosom Cancer.* 1996;15(3):182–186.
- Mozziconacci MJ, Rosenauer A, Restouin A, et al. Molecular cytogenetics of the acute promyelocytic leukemia-derived cell line NB4 and of four all-trans retinoic acid-resistant subclones. *Genes Chromosom Cancer.* 2002;35(3):261–270.
- Liang JC, Ning Y, Wang RY, et al. Spectral karyotypic study of the HL-60 cell line: detection of complex rearrangements involving chromosomes 5, 7, and 16 and delineation of critical region of deletion on 5q31.1. *Cancer Genet Cytogenet.* 1999;113(2):105–109.
- Collins S, Groudine M. Amplification of endogenous myc-related DNA sequences in a human myeloid leukaemia cell line. *Nature.* 1982;298(5875):679–681.
- Von Hoff DD, Forseth B, Clare CN, Hansen KL, VanDevanter D. Double minutes arise from circular extrachromosomal DNA intermediates which integrate into chromosomal sites in human HL-60 leukemia cells. *J Clin Invest.* 1990;85(6):1887–1895.
- Hoffmann JC, Stabla K, Burchert A, et al. Monitoring of acute myeloid leukemia patients after allogeneic stem cell transplantation employing semi-automated CD34+ donor cell chimerism analysis. *Ann Hematol.* 2014;93(2):279–285.
- Luthra R, Patel KP, Reddy NG, et al. Next-generation sequencing-based multigene mutational screening for acute myeloid

- leukemia using MiSeq: applicability for diagnostics and disease monitoring. *Haematologica*. 2014;99(3):465–473.
29. Asou H, Tashiro S, Hamamoto K, Otsuji A, Kita K, Kamada N. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood*. 1991;77(9):2031–2036.
 30. Yanagisawa K, Horiuchi T, Fujita S. Establishment and characterization of a new human leukemia cell line derived from M4E0. *Blood*. 1991;78(2):451–457.
 31. Meissner R de V, Covas DT, Dias PMB, Job F, Leite M, Nardi NB. Analysis of mRNA transcripts in chronic myeloid leukemia patients. *Genet Mol Biol*. 1999;22:475–479.
 32. Rucker FG, Bullinger L, Schwaenen C, et al. Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J Clin Oncol*. 2006;24(24):3887–3894.
 33. Forbes SA, Tang G, Bindal N, et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res*. 2010;38(Database issue):D652–D657.
 34. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet*. 2013;14(6):415–426.
 35. Strom SP. Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biol Med*. 2016;13(1):3–11.
 36. Conte N, Varela I, Grove C, et al. Detailed molecular characterisation of acute myeloid leukaemia with a normal karyotype using targeted DNA capture. *Leukemia*. 2013;27(9):1820–1825.
 37. Spencer DH, Abel HJ, Lockwood CM, et al. Detection of FLT3 internal tandem duplication in targeted, short-read-length, next-generation sequencing data. *J Mol Diagn*. 2013;15(1):81–93.
 38. Quentmeier H, Reinhardt J, Zaborski M, Drexler HG. FLT3 mutations in acute myeloid leukemia cell lines. *Leukemia*. 2003;17(1):120–124.
 39. Frech M, Jehn LB, Stabla K, et al. Dasatinib and allogeneic stem cell transplantation enable sustained response in an elderly patient with RCSD1-ABL1-positive acute lymphoblastic leukemia. *Haematologica*. 2017;102(4):e160–e162.
 40. McKerrell T, Moreno T, Ponstingl H, et al. Development and validation of a comprehensive genomic diagnostic tool for myeloid malignancies. *Blood*. 2016;128(1):e1–9.
 41. Akagi T, Ogawa S, Dugas M, et al. Frequent genomic abnormalities in acute myeloid leukemia/myelodysplastic syndrome with normal karyotype. *Haematologica*. 2009;94(2):213–223.
 42. Walter MJ, Payton JE, Ries RE, et al. Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proc Natl Acad Sci U S A*. 2009;106(31):12950–12955.
 43. Gronseth CM, McElhone SE, Storer BE, et al. Prognostic significance of acquired copy-neutral loss of heterozygosity in acute myeloid leukemia. *Cancer*. 2015;121(17):2900–2908.
 44. Wang H-Y, Rashidi HH. The new clinicopathologic and molecular findings in myeloid neoplasms with inv(3)(q21q26)/t(3;3)(q21;q26.2). *Arch Pathol Lab Med*. 2016;140(12):1404–1410.
 45. Duncavage EJ, Abel HJ, Szankasi P, Kelley TW, Pfeifer JD. Targeted next generation sequencing of clinically significant gene mutations and translocations in leukemia. *Mod Pathol*. 2012;25(6):795–804.
 46. He J, Abdel-Wahab O, Nahas MK, et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. *Blood*. 2016;127(24):3004–3014.
 47. Tybakiñoja A, Elonen E, Piippo K, Porkka K, Knuutila S. Oligonucleotide array-CGH reveals cryptic gene copy number alterations in karyotypically normal acute myeloid leukemia. *Leukemia*. 2007;21(3):571–574.
 48. Alvarez S, MacGrogan D, Calasanz MJ, Nimer SD, Jhanwar SC. Frequent gain of chromosome 19 in megakaryoblastic leukemias detected by comparative genomic hybridization. *Genes Chromosom Cancer*. 2001;32(3):285–293.
 49. Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med*. 2017;377(5):454–464.