

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.*

©2019 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol..2018.194258

Received: May 26, 2018.

Accepted: September 5, 2018.

Pre-published: September 6, 2018.

Correspondence: brendelc@staff.uni-marburg.de

Comprehensive Genetic Diagnosis of Acute Myeloid Leukemia by Next Generation Sequencing

Elisabeth K.M. Mack, André Marquardt, Danny Langer, Petra Ross, Alfred Ultsch, Michael G. Kiehl, Hildegard I.D. Mack, Torsten Haferlach, Andreas Neubauer and Cornelia Brendel

Supplementary Materials

Contents

Supplementary Material main file (.pdf)

- Supplementary Methods
- Supplementary References
- Supplementary Figures S1-S10
- Supplementary Tables S1-S12

Supplementary Methods

Preparation of sequencing libraries

Bone marrow or peripheral blood samples were lysed in hypotonic buffer (150 mM NH₄Cl, 10 mM KHCO₃, 0.1 mM EDTA) and DNA and RNA were extracted using the QIAamp[®] DNA and RNeasy[®] Mini kits (Qiagen, Hilden, Germany). DNA and RNA were isolated from cell lines using the NucleoSpin[®] Tissue and RNA Mini kits (Macherey Nagel, Düren, Germany). For the preparation of whole genome libraries, 1-3 µg of genomic DNA from bone marrow or peripheral blood samples in 24 µl nuclease-free water were mixed with 3 µl NEBNext[®] dsDNA Fragmentase[®] (New England Biolabs, Ipswich, MA) and 3 µl fragmentase reaction buffer v2 and incubated at 37°C for 25 minutes. Fragmentation was stopped by adding 7.5 µl EDTA (pH 8.0) and fragmented DNA was diluted 1:1 in ultra-pure water. Double size selection of approximately 300 bp fragments was performed using 50 µl of the

diluted fragmented DNA and Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN) in a ratio of 0.7 (35 μ l) in the first step and 0.2 (10 μ l) in the second step. Beads were washed using 80% ethanol and DNA was eluted in 45 μ l 10 mM TRIS-HCl, pH 8.0. The size distribution of the DNA fragments was analyzed on a TapeStation 2200 system (Agilent, Santa Clara, CA) with a high sensitivity D1000 screen tape and concentration was determined using a Qubit™ 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA) with the Qubit™ dsDNA HS assay. 10-30 ng of fragmented size selected DNA in 50 μ l 10 mM TRIS-HCl, pH 8.0 were subjected to end repair, adapter ligation, library amplification and dual indexing using the NEBNext Ultra™ II DNA Library Prep Kit and NEBNext® Multiplex Oligos for Illumina® as recommended by the manufacturer. Eight amplification cycles were performed for >20 ng input and nine for \leq 20 ng. Construction of TruSight® Myeloid libraries (Illumina, San Diego, CA) used 50-100 ng unfragmented genomic DNA as input and followed the manufacturer's instructions for hybridization of the oligo pool, removal of unbound oligos, extension and ligation of bound oligos and library amplification. Library normalization was skipped. QIAseq™ Myeloid libraries (Qiagen, Hilden, Germany) were prepared as recommended by the manufacturer with 20-80 ng genomic DNA input. Size distributions of DNA-based sequencing libraries were checked on a D1000 screen tape. Preparation of FusionPlex® Heme v1 or v2 libraries (ArcherDX, Boulder, CO) started with 100-200 ng total RNA and was conducted as outlined in the manufacturer's protocol. Libraries were quantified by qPCR using the NEBNext® Library Quant Kit for Illumina® as recommended on a CX96 qPCR cycler (BioRad, Hercules, CA). Concentrations were calculated from C_t values using mean library sizes determined by TapeStation analysis for whole genome and mutation libraries or, respectively, 200 bp for fusion libraries. Each library was quantified in a 1:10⁵ and 1:10⁶ dilution and calculated concentrations were averaged.

Next Generation Sequencing

For sequencing, libraries were diluted to 4 nM (whole genome libraries), 3 nM (mutation libraries) or 3.5 nM (fusion libraries) and pooled in a ratio of 1:2.5:2 (2:5:4 μ l). 5 μ l of the library pool were denatured by adding 5 μ l 0.2 M NaOH and incubating the mixture at room temperature for 5 min. Denaturation was stopped by 990 μ l ice-cold HT1 incorporation buffer, yielding a concentration of 20 pM. For runs with mutation or fusion libraries only, denatured pools were further diluted to a final concentration of 15 pM, or, respectively, 17.5 nM. The denatured and diluted library pool was supplemented with 1% of 12.5 μ M denatured PhiX v3 control library (Illumina) and 600 μ l of the final sample were loaded into the sequencing reagent cartridge. Sequencing was performed on a MiSeq instrument with standard v2 flowcells and a read length of 2x 150 bp with two index reads. The workflow parameter was set to *Generate FastQ* and the assay parameter to *TruSeq HT*, and the standard Illumina sequencing primer was used in runs that included more than one type of sequencing library or whole genome libraries only. For runs that included fusion libraries only, the assay parameter was changed to *Nextera*, runs with TruSight[®] Myeloid libraries only used the workflow *TruSeq Amplicon*, and runs that included QIAseq[™] mutation libraries only were performed using the assay parameter *Nextera v2* with a custom sequencing primer for read 1 (provided with the panel).

Quantitative PCR

Amplification of the *MYC* gene and loss of chromosome 7q were validated in selected samples by quantitative PCR using primers targeting intronic sequences in the regions of interest: *MYC_F* 5'-gtgtctttctccattcctg-3', *MYC_R* 5'-ttttccctctgccttctcc-3', *ARHGEF5_F* 5'-agccaccatttctatgacc-3', *ARHGEF5_R* 5'-tgtgcctctctggtatctttg-3', *PIK3CG_F* 5'-acaggcaacgtgaagcaag-3', *PIK3CG_R* 5'-acagttaagggtcaagtcccag-3', *VKORC1L1_F* 5'-acagcgaagccttgaaac-3', *VKORC1L1_R* 5'-gcagatgaccaacccaaag-3', *ACTG1_F* 5'-tgcccggcttgattctgac-3', *ACTG1_R* 5'-tgccagtgtgatgtggag-3' (Eurofins Genomics, Ebersberg, Germany). Reactions were set up using the QuantiTect[®] SYBR[®] Green PCR

mastermix (Qiagen) as recommended by the manufacturer and the following temperature conditions on a CFX96 qPCR instrument (Biorad): 95 °C / 15 min, 40 cycles 95 °C / 20 sec – 56 °C / 30 sec – 72°C / 30 sec – plate read, melt curve 65-95 °C / 30 sec. All qPCR assays were performed in triplicates, CNVs relative to a normal karyotype control were calculated according to the $2^{\Delta\Delta Ct}$ method using *ACTG1* as a reference gene.

Description of the CAI[N] algorithm and analysis pipeline

CNV analysis was performed using a proprietary algorithm designated CAI[N] (chromosomal aberration identifier [numerical]). In brief, normal karyotypes were constructed by randomly sampling 1.5×10^6 reads (150-250 bp) from the hg38 reference genome (UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) assembly)¹. Compiling the reference for CNV calculations *in silico* allowed us to construct a larger and thus statistically more robust control dataset than sequencing normal karyotype samples (which in this order of magnitude would not have been feasible at our center).

Reads were re-aligned using Bowtie 2 (version 2.2.5)² with standard parameters and mapped to 1 Mb fixed genomic windows. Read counts in each window were normalized to 1×10^6 to obtain a conservative estimate for the number of mappable reads in sequencing experiments. Normal female (N=2,819) and male (N=2,605) karyotypes were averaged in order to generate a reference for CNV calculations. For experimental samples, the R1 reads were extracted from *fastq* sequence files and processed as described for *in silico*-reads. Only R1 reads were used. Differences of normalized read numbers from the average of normal *in silico* karyotypes ($\Delta = n_{\text{sample}} - \text{Mean}(n_{\text{reference}})$) were compared to the standard deviations for the reference and windows were called significantly over- or underrepresented if $\Delta \geq 3 \cdot \text{SD}_{\text{reference}}$ (corresponding to $P < 0.003$). Significantly over- or underrepresented windows were mapped to cytogenetic bands using positional information from the UCSC Genome Browser Database. Gains or losses of cytogenetic regions were reported in case of significant over- or underrepresentation of (a) 100% of windows in one band or (b) at least 34% of windows in three adjacent bands.

Initially, this threshold was set arbitrarily in order to increase confidence of CNV calling on the level of chromosome bands. It was kept as a decision rule throughout the course of our work as it performed very well in karyotype identification. Using a Mean \pm 3 \cdot SD measure for the identification of CNVs assumed that read counts in 1 Mb windows are normally distributed. It has been reported previously that read frequencies in genomic windows follow a Poisson distribution that converges towards a normal distribution with increasing coverage and/or increasing window size.³⁻⁶ With a window size of 1 Mb and a total read count of 1×10^6 we expect about 320 reads per window (in regions with mappable reads), which is large enough to approximate the actual distribution of read counts by a normal distribution. Normality tests of mapped read frequencies in 1 Mb windows in our *in silico* reference dataset confirmed this assumption for $\geq 97.9\%$ of windows (Table S12). Normality tests were performed using the functions *normaltest* and *shapiro* from the SciPy⁷ Python-based statistical functions module (release 1.1.0, www.scipy.org).

Raw CAI[N] karyotypes were reviewed with the help of graphical output in order to decide whether a gain/loss was likely detectable by cytogenetics. Small deletions towards chromosome ends or copy number alterations in noisy background were excluded from the final reports. Only a subset of *in silico*-generated references was used for the construction of these plots in Figures S1-S3, S5 and S6 ($N_{\text{female}}=2,007$, $N_{\text{male}}=1,398$).

CAI[N] was implemented in Python (version 2.7; www.python.org). The source code is available from the authors upon request. CAI[N] uses only R1 reads. GC-correction was omitted based on the following considerations: (a) PCR amplification during library preparation has been found to be the major cause of GC bias⁸, and we here used the highest fidelity polymerase that is currently available^{9,10}. (b) 150 bp reads sampled randomly from the reference genome do not exhibit extreme GC-content and the GC-content of reads obtained in actual sequencing experiments lies in the same range (Fig. S10a upper panel). (c) The GC-content of reads that uniquely map back to the reference genome lies in the same range for

experimental and *in silico*-sampled reads (Fig. S10a, lower panel. (d) 1 Mb windows in the human genome do not display extremely high or extremely low GC-content (Fig. S10b).

The following outline describes the steps of the CAI[N] algorithm.

1. Construction of reference karyotypes

1.1. Prepare file

1.1.1. Download the hg38 reference genome

hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.2bit

1.1.2. Filter to remove lines and sequences containing “ChrUn”.

1.1.3. Concatenate the sequence information for each chromosome to obtain one sequence file for each chromosome.

1.2. Simulate whole genome library preparation and low coverage whole genome sequencing:

1.2.1. Expect 1.5×10^6 reads for each experimental sample (10% flowcell capacity).

1.2.2. Introduce 12.2×10^6 random cuts into the genome to obtain approximately 1.5×10^6 non-overlapping fragments of the desired read length $150 \text{ bp} \leq x \leq 250 \text{ bp}$ (i.e. simulate DNA fragmentation by fragmentase).

1.2.3. Extract all fragments of the desired length and map them to the genome using bowtie2 (cf. below).

1.2.4. Save results in a .csv file.

1.2.5. Repeat the simulation of normal karyotypes as desired. Higher numbers of reference karyotypes result in smoother region plots.

2. Prepare sample reads

2.1. Unzip the R1 reads and determine the number of reads obtained by sequencing.

3. Map reads

3.1. Use bowtie2 with standard parameters and hg38 as reference genome to map the obtained reads:

```
bowtie2 -x hg38_human_reference -U reads.fq -S bowtie_output.txt
```

3.2. Filter for uniquely mapping reads and clean file for easier handling:

```
sed '/XS:/d' bowtie_output.txt > uniquely_mapped_reads.txt
```

```
sed '/^@/d' uniquely_mapped_reads.txt > newmapping_test.txt
```

3.3. Count the number of mapped reads for each chromosome, for example for chromosome 1:

```
grep -c P 'chr1\t'
```

3.4. Divide the chromosomes of the reference genome in 1 Mb fixed windows.

3.5. Assign all uniquely mapped reads to genomic windows according to their starting position.

3.6. Normalize the number of reads to a total read count of 1×10^6 .

4. Comparison patient – in silico normal reference karyotype

4.1. Determine the mean read number and standard deviation for each 1 Mb window in the *in silico*-generated reference dataset.

4.2. Check for each 1 Mb window in the reference dataset and the experimental sample (patient) whether the number of normalized reads is smaller than or equal to 10% of the mean read count per window on the whole chromosome.

4.3. If $n \leq 0.1 \cdot \text{Mean}(n_{\text{chromosome}})$ for the reference AND the experimental sample, set $n = \text{NaN}$ for this window.

4.4. Compare normalized read numbers for each 1 Mb window between the experimental sample and the *in silico*-generated normal reference karyotype for the same gender (random female / random male).

4.5. Call a region (1 Mb window) significantly over- or, respectively, underrepresented if

$$|\Delta = n_{\text{sample}} - \text{Mean}(n_{\text{reference}})| > 3 \cdot \text{SD}_{\text{reference}} \text{ (corresponding to } P < 0.003)$$

5. Mapping to cytogenetic bands

5.1. Download the cytoBand file from

hgdownload.soe.ucsc.edu/goldenPath/hg38/database/cytoBand.txt.gz

5.2. Map all 1 Mb windows in the experimental sample to the corresponding cytogenetic band using the zero-corrected normalized read count file (cf. step 4.3).

5.3. For each band, determine the percentage of windows that is over- or, respectively, underrepresented. Call the band over-/underrepresented if at least 34% of windows are over-/underrepresented.

5.4. Combine all adjacent bands affected by the same kind of CNV. Note the first and the last over-/underrepresented band.

6. Output

6.1. List (a) all bands that are 100% over- or underrepresented irrespective of total region size and (b) all cytogenetic regions (first to last band) that are at least 34% over-/underrepresented and span at least three 1 Mb windows.

6.2. Plot for each chromosome (a) read counts in 1 Mb windows for the sample and the *in silico* reference karyotype (“region plot”) and (b) the interpretation of the difference between sample and reference (significant at $P < 0.003$) for each window (“decision plot”).

Region and decision plots in figures 2, 4, 5, S4, S7 were prepared for publication using GraphPad Prism 7 (GraphPad Software, La Jolla, CA).

Analysis of amplicon panel sequencing data

Fusion sequencing results were analyzed using the Archer[®] Analysis pipeline (version 4.1 or 5.0 for the FusionPlex[®] Heme v1 or v2 panels, respectively) with standard or advanced settings (Error Correction: *ON*, Read Depth Normalization: 3,500,000) and, apart from that, default parameters. SNV and Indel calling in TruSight[®] Myeloid data used the BaseSpace[®] TruSight[®] Myeloid App (Illumina), which included the following analysis tools: Isis Smith-Waterman-Gotoh 2.6.21.7 (Aligner), SAMtools 1.2, Somatic Variant Caller 4.0.13.1, Isis 2.6.21.7 (Analysis Software) and IONA 1.0.10.37 (Annotation Service) with the Illumina

Annotation Dataset version 72.5. Variants were filtered with BaseSpace[®] Variant Studio 2.2. Analysis parameters / filters included: Exon Padding: 20 bp, Context: not *Intergenic*, Consequence: not *synonymous variant*, *intron variant* or *in-frame deletion / insertion*, COSMIC or ClinVar annotation, AltVariantFreq >0.03 (>0.001 in dilution studies), Filters: *Pass*, Quality >99, TotalDepth >100, AlleleFreqEur <0.05 (allele frequency in European population from 1,000 genomes), SIFT/Polyphen: not *benign* AND *tolerated*. QIASeq[™] datasets were analyzed using smCounter¹¹ (Qiagen, <http://www.qiagen.com/de/shop/genes-and-pathways/data-analysis-center-overview-page/qiaseq-targeted-dna-panels-analysis-center/>) with standard settings. Variants were filtered in Microsoft Excel applying the following filter parameters: ID: *includes "COSM"*, Filter: *Pass*, UMT (unique molecular tags) >40, VMF (variant molecular tags frequency frequency) >0.05, Annotation: not *upstream/downstream gene variant*, *synonymus variant* or *intron variant* (except for *splice region variant*), Annotation Impact: *High* or *Moderate*. Variants passing these filters were further examined manually for their clinical significance by checking their annotations in COSMIC (<http://cancer.sanger.ac.uk/cosmic>) and, if applicable, *JAX Clinical Knowledge Base* (<https://ckb.jax.org>). *FLT3*-ITDs in TruSight[®] Myeloid libraries were analyzed using ITD-seek¹² (source code and documentation downloaded from <https://github.com/tommyau/ITDseek> on 05/09/2016) with SAMtools version 1.13.1. For the detection of *KMT2A*-PTDs we developed PTDi. The algorithm was adapted from Conte et al. (2013)¹³ and implemented in Python 2.7. In brief, PTDi performs the following steps:

1. Merge corresponding forward and reverse reads.
2. Extract all unique reads, i.e unique combinations of a molecular tag and a gene specific primer.
3. Identify all reads mapping to *KMT2A* based on the gene-specific primer sequence.
4. Assign reads to individual exons based on the gene-specific primer sequence and count.

5. Normalize the read count per exon of the *KMT2A* gene to the total number of unique reads for the sample.
6. Compare normalized read counts per exon in the experimental sample to a PTD-negative control.

Supplementary References

1. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 2015;43(D1):D670–D681.
2. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–359.
3. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature* 2008;456(7218):53–59.
4. Vardhanabhati S, Jeng XJ, Wu Y, Li H. Parametric modeling of whole-genome sequencing data for CNV identification. *Biostatistics* 2014;15(3):427–441.
5. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 2009;1080.
6. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 2009;19(9):1586–1592.
7. Millman KJ, Aivazis M. Python for Scientists and Engineers. *Comput Sci Eng* 2011;13(2):9–12.
8. Aird D, Ross MG, Chen W-S, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 2011;12(2):R18.
9. Potapov V, Ong JL. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS One* 2017;12(1):e0169774.
10. Stasik S, Schuster C, Ortlepp C, et al. An optimized targeted Next-Generation Sequencing approach for sensitive detection of single nucleotide variants. *Biomol*

Detect Quantif 2018;156–12.

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.
12. 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;1111.
13. 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.

Table S1: Target transcripts of the ArcherDX FusionPlex[®] Heme v1 panel.

Gene	Accession	Exons	Direction
<i>ABL1</i>	NM_005157	1-4	5'
<i>ALK</i>	NM_004304	20	5'. 3'
<i>BCR</i>	NM_004327	1-3, 12-20	5'
<i>CBFB</i>	NM_022845	1-6	3'
<i>FGFR1</i>	NM_023110	3-10	5'
<i>JAK2</i>	NM_004972	6-20	5'
<i>JAK2</i>	NM_004972	9-12	3'
<i>KMT2A</i>	NM_005933	4-15	3'
<i>KMT2A</i>	NM_005933	1-3	5'
<i>MECOM</i>	NM_004991	1-4	5'
<i>MKL1</i>	NM_020831	1-6	5'
<i>NOTCH1</i>	NM_017617	24-29	5'
<i>NUP214</i>	NM_005085	17-19	5'
<i>PDGFRA</i>	NM_006206	11-14	5'
<i>PDGFRB</i>	NM_002609	8-14	5'
<i>PICALM</i>	NM_007166	16-19	3'
<i>RARA</i>	NM_000964	2-5	5'
<i>RBM15</i>	NM_022768	1, 2	3'
<i>RUNX1</i>	NM_001754	1-4	5'
<i>RUNX1T1</i>	NM_001198679	1-4	5'
<i>TAL1</i>	NM_003189	2-4	5'
<i>TCF3</i>	NM_003200	11-17	3'

Table S2: Target transcripts of the ArcherDX FusionPlex[®] Heme v2 panel.

Gene	Accession	Exons	Direction	Gene	Accession	Exons	Direction
<i>ABL1</i>	NM_005157	1-5	5'	<i>MLLT10</i>	NM_004641	2-18	5'
<i>ABL2</i>	NM_005158	2- 8	3'	<i>MLLT10</i>	NM_004641	7-10	3'
<i>ALK</i>	NM_004304	2, 4, 6, 10, 16-23	5'	<i>MLLT4</i>	NM_001040000	2	5'
<i>BCL11B</i>	NM_138576	3, 4	5'	<i>MYC</i>	NM_002467	1, 2	5'
<i>BCL11B</i>	NM_138576	2, 3	3'	<i>MYH11</i>	NM_002474	7-16	5'
<i>BCL2</i>	NM_000633	3	3'	<i>NF1</i>	NM_000267	14	3'
<i>BCL2</i>	NM_000633	2	5'	<i>NF1</i>	NM_000267	36	5'
<i>BCL6</i>	NM_001706	2, 3	5'	<i>NFKB2</i>	NM_002502	14-21	3'
<i>BCR</i>	NM_004327	1-3,8,12-16	3'	<i>NOTCH1</i>	NM_017617	24-29	5'
<i>BIRC3</i>	NM_001165	4-7	3'	<i>NOTCH1</i>	NM_017617	24	3'
<i>CBFB</i>	NM_022845	4, 5	3'	<i>NTRK3</i>	NM_002530	15	5'
<i>CCND1</i>	NM_053056	5	3'	<i>NTRK3</i>	NM_001007156	15	5'
<i>CCND3</i>	NM_001760	2	5'	<i>NUP214</i>	NM_005085	17-19	5'
<i>CDK6</i>	NM_001259	1- 4	3'	<i>NUP98</i>	NM_016320	8-17	3'
<i>CHD1</i>	NM_001270	1, 2	5'	<i>NUP98</i>	NM_016320	12, 13	5'
<i>CHIC2</i>	NM_012110	1- 3	3'	<i>P2RY8</i>	NM_178129	1	3'
<i>CIITA</i>	NM_000246	1, 2	3'	<i>PAG1</i>	NM_018440	2	5'
<i>CREBBP</i>	NM_004380	2-6	5'	<i>PAX5</i>	NM_016734	1, 4-8	3'
<i>CRLF2</i>	NM_022148	1	5'	<i>PAX5</i>	NM_016734	6 8	5'
<i>CSF1R</i>	NM_005211	9-14	5'	<i>PDCD1LG2</i>	NM_025239	1-3	5'
<i>DEK</i>	NM_003472	2, 3	3'	<i>PDCD1LG2</i>	NM_025239	5, 6	3'
<i>DUSP22</i>	NM_020185	1, 2	3'	<i>PDGFRA</i>	NM_006206	9-14	5'
<i>EBF1</i>	NM_024007	10-15	3'	<i>PDGFRB</i>	NM_002609	8-14	5'
<i>EIF4A1</i>	NM_001416	2, 3	5'	<i>PICALM</i>	NM_007166	16-19	3'
<i>EPOR</i>	NM_000121	7, 8	3'	<i>PML</i>	NM_002675	2	5'
<i>ERG</i>	NM_004449	7-11	5'	<i>PML</i>	NM_002675	2-7	3'
<i>ETV6</i>	NM_001987	1-6	3'	<i>PRDM16</i>	NM_022114	1-4	5'
<i>ETV6</i>	NM_001987	2-6	5'	<i>PTK2B</i>	NM_173176	2-8	5'
<i>FGFR1</i>	NM_023110	12, 17	3'	<i>RARA</i>	NM_000964	2-5	5'
<i>FGFR1</i>	NM_023110	2-11, 17	5'	<i>RBM15</i>	NM_022768	1	3'
<i>GLIS2</i>	NM_032575	2, 3	5'	<i>ROS1</i>	NM_002944	33-36	5'
<i>IKZF1</i>	NM_006060	1-3	3'	<i>RUNX1</i>	NM_001754	5-9	5'
<i>IKZF1</i>	NM_006060	7, 8	5'	<i>RUNX1</i>	NM_001754	2-8	3'
<i>IKZF2</i>	NM_016260	3, 4	3'	<i>RUNX1T1</i>	NM_001198679	2, 3	5'
<i>IKZF3</i>	NM_012481	2-7	3'	<i>SEMA6A</i>	NM_020796	1, 2	3'
<i>JAK2</i>	NM_004972	6-20	5'	<i>SETD2</i>	NM_014159	1-12	3'
<i>JAK2</i>	NM_004972	9-12	3'	<i>STIL</i>	NM_003035	1, 2	3'
<i>KAT6A</i>	NM_006766	13-16	3'	<i>TAL1</i>	NM_003189	2, 4	5'
<i>KLF2</i>	NM_016270	2, 3	5'	<i>TAL1</i>	NM_001290404	2, 3	5'
<i>KMT2A</i>	NM_005933	2, 3	5'	<i>TCF3</i>	NM_003200	11-18	3'
<i>KMT2A</i>	NM_005933	4-35	3'	<i>TFG</i>	NM_006070	2-4	3'
<i>MALT1</i>	NM_006785	9	3'	<i>TP63</i>	NM_003722	3-5	5'
<i>MECOM</i>	NM_004991	1-4	5'	<i>TYK2</i>	NM_003331	16, 18	5'
<i>MKL1</i>	NM_020831	4-6	5'	<i>ZCCHC7</i>	NM_032226	1, 2	3'
<i>MLF1</i>	NM_022443	2-4	5'	<i>ZCCHC7</i>	NM_032226	2-4	5'

Table S3: Sequencing targets of the Illumina TruSight® Myeloid panel.
Indicated regions refer to the hg19 reference genome.

Gene	Target Exons	Gene	Target Exons	Gene	Target Exons	Gene	Target Exons
<i>ABL1</i>	4-6	<i>DNMT3A</i>	Full	<i>KDM6A</i>	Full	<i>RAD21</i>	full
<i>ASXL1</i>	12	<i>ETV6</i>	Full	<i>KIT</i>	2, 8-11, 13, 17	<i>RUNX1</i>	full
<i>ATRX</i>	8-10, 17-31	<i>EZH2</i>	Full	<i>KRAS</i>	2, 3	<i>SETBP1</i>	4 (partial)
<i>BCOR</i>	Full	<i>FBXW7</i>	9-11	<i>MLL</i>	5-8	<i>SF3B1</i>	13-16
<i>BCORL1</i>	Full	<i>FLT3</i>	14, 15, 20	<i>MPL</i>	10	<i>SMC1A</i>	2, 11, 16, 17
<i>BRAF</i>	15	<i>GATA1</i>	2	<i>MYD88</i>	3-5	<i>SMC3</i>	10, 13, 19, 23, 25, 28
<i>CALR</i>	9	<i>GATA2</i>	2-6	<i>NOTCH1</i>	26-28, 34	<i>SRSF2</i>	1
<i>CBL</i>	8, 9	<i>GNAS</i>	8, 9	<i>NPM1</i>	12	<i>STAG2</i>	Full
<i>CBLB</i>	9, 10	<i>HRAS</i>	2, 3	<i>NRAS</i>	2, 3	<i>TET2</i>	3-11
<i>CBLC</i>	9, 10	<i>IDH1</i>	4	<i>PDGFRA</i>	12, 14, 18	<i>TP53</i>	2-11
<i>CDKN2A</i>	Full	<i>IDH2</i>	4	<i>PHF6</i>	Full	<i>U2AF1</i>	2, 6
<i>CEBPA</i>	Full	<i>IKZF1</i>	Full	<i>PTEN</i>	5, 7	<i>WT1</i>	7, 9
<i>CSF3R</i>	14-17	<i>JAK2</i>	12, 14	<i>PTPN11</i>	3 + 13	<i>ZRSR2</i>	full
<i>CUX1</i>	full	<i>JAK3</i>	13				

Table S4: Sequencing targets of the QIASeq™ Targeted DNA Human Myeloid Neoplasms panel.

<i>ABL1</i>	<i>CALR</i>	<i>DNMT1</i>	<i>HNRNPK</i>	<i>KLHL6</i>	<i>NPM1</i>	<i>PTEN</i>	<i>STAG2</i>
<i>ADA</i>	<i>CARD11</i>	<i>DNMT3A</i>	<i>HRAS</i>	<i>KMT2A</i>	<i>NRAS</i>	<i>PTPN11</i>	<i>STAT3</i>
<i>ANKRD26</i>	<i>CBL</i>	<i>EED</i>	<i>IDH1</i>	<i>KMT2C</i>	<i>NSD1</i>	<i>RAD21</i>	<i>STXBP2</i>
<i>ASXL1</i>	<i>CBLB</i>	<i>EGFR</i>	<i>IDH2</i>	<i>KRAS</i>	<i>NTRK3</i>	<i>RB1</i>	<i>SUZ12</i>
<i>ASXL2</i>	<i>CBLC</i>	<i>ELANE</i>	<i>IKZF1</i>	<i>LRRC4</i>	<i>OR13H1</i>	<i>RELN</i>	<i>TAL1</i>
<i>ATM</i>	<i>CDKN2A</i>	<i>EP300</i>	<i>IKZF3</i>	<i>LUC7L2</i>	<i>OR8B12</i>	<i>RUNX1</i>	<i>TERC</i>
<i>BCL6</i>	<i>CEBPA</i>	<i>ETV6</i>	<i>IL7R</i>	<i>MAP2K1</i>	<i>P2RY2</i>	<i>SETBP1</i>	<i>TERT</i>
<i>BCOR</i>	<i>CHEK2</i>	<i>EZH2</i>	<i>JAK1</i>	<i>MLH1</i>	<i>PAX5</i>	<i>SF1</i>	<i>TET2</i>
<i>BCORL1</i>	<i>CREBBP</i>	<i>FAM154B</i>	<i>JAK2</i>	<i>MPL</i>	<i>PCDHB1</i>	<i>SF3A1</i>	<i>TNFRSF13B</i>
<i>BCR</i>	<i>CRLF2</i>	<i>FAM47A</i>	<i>JAK3</i>	<i>MSH2</i>	<i>PDGFRA</i>	<i>SF3B1</i>	<i>TP53</i>
<i>BIRC3</i>	<i>CSF1R</i>	<i>FAS</i>	<i>KAT6A</i>	<i>MSH6</i>	<i>PHF6</i>	<i>SH2B3</i>	<i>TPMT</i>
<i>BLM</i>	<i>CSF3R</i>	<i>FBXW7</i>	<i>KCNA4</i>	<i>MYC</i>	<i>PML</i>	<i>SH2D1A</i>	<i>TUBA3C</i>
<i>BRAF</i>	<i>CTCF</i>	<i>FLRT2</i>	<i>KCNK13</i>	<i>MYD88</i>	<i>PMS2</i>	<i>SMARCB1</i>	<i>U2AF1</i>
<i>BRCA1</i>	<i>CUX1</i>	<i>FLT3</i>	<i>KDM6A</i>	<i>NBN</i>	<i>PRAMEF2</i>	<i>SMC1A</i>	<i>U2AF2</i>
<i>BRCA2</i>	<i>DAXX</i>	<i>GATA1</i>	<i>KDR</i>	<i>NF1</i>	<i>PRF1</i>	<i>SMC3</i>	<i>WAS</i>
<i>BRINP3</i>	<i>DDX41</i>	<i>GATA2</i>	<i>KIT</i>	<i>NOTCH1</i>	<i>PRPF40B</i>	<i>SRP72</i>	<i>WRN</i>
<i>C17orf97</i>	<i>DNM2</i>	<i>GJB3</i>	<i>KLHDC8B</i>	<i>NPAT</i>	<i>PRPF8</i>	<i>SRSF2</i>	<i>WT1</i>

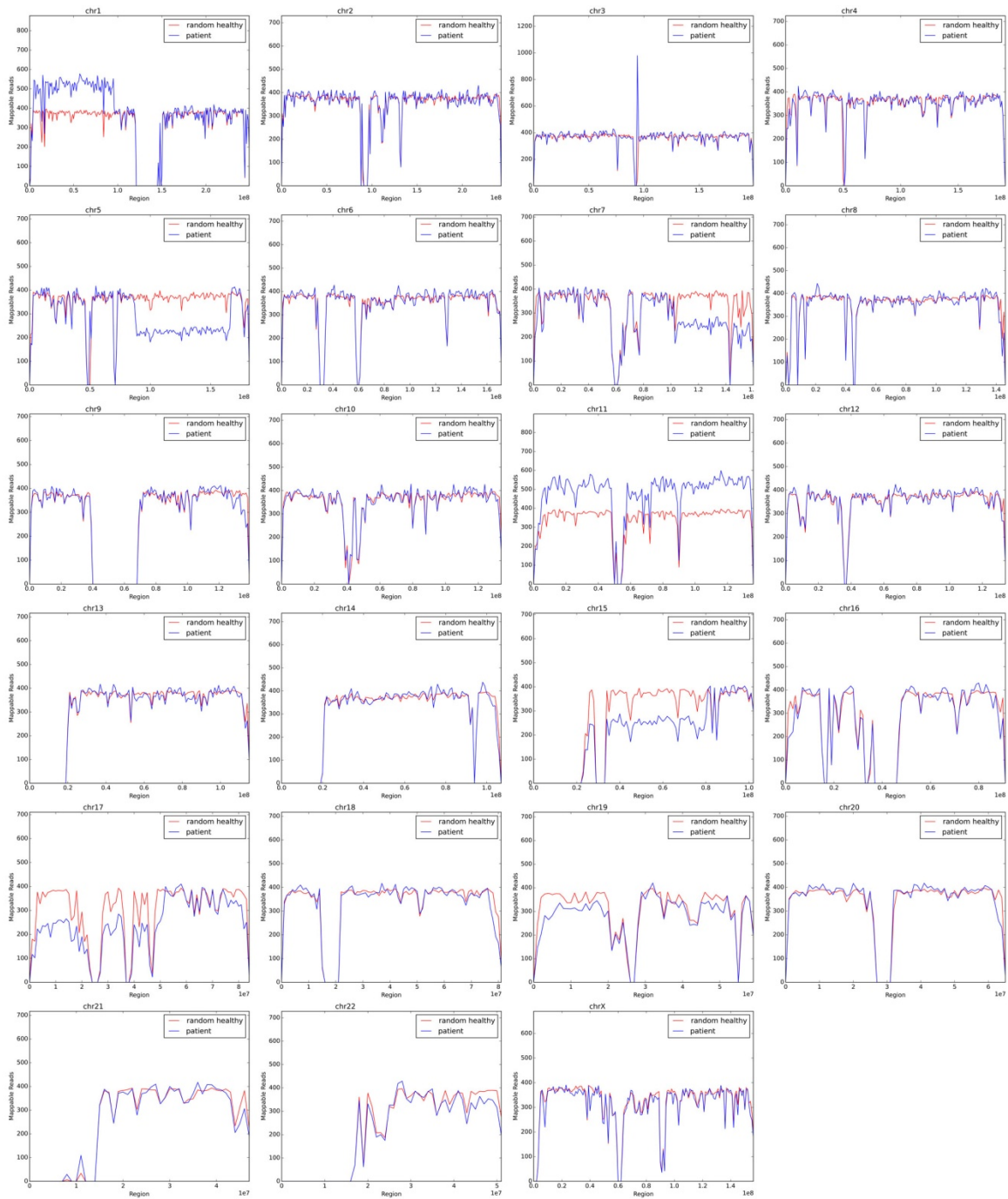


Fig. S1: Graphical representation of the CNV karyotype of patient sample AML-1.

Read numbers in 1 Mb windows were normalized to 1×10^6 total reads for the patient sample.

Random healthy: Mean read numbers for *in silico*-generated normal female karyotypes (N=2,007).

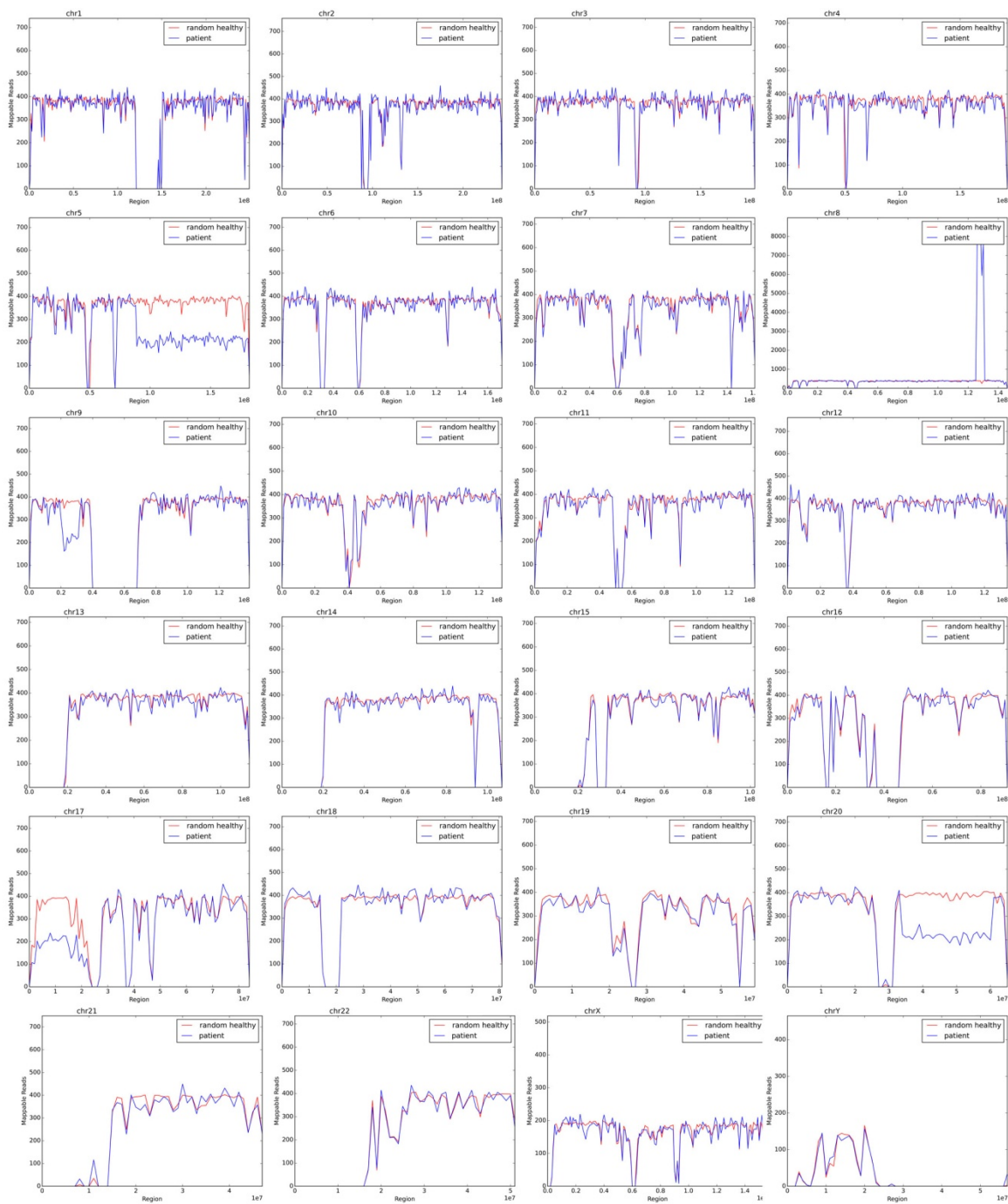


Fig. S2: Graphical representation of the CNV karyotype of patient sample AML-2.

Read numbers in 1 Mb windows were normalized to 1×10^6 total reads for the patient sample. Random healthy: Mean read numbers for *in silico*-generated normal male karyotypes (N=1,398).

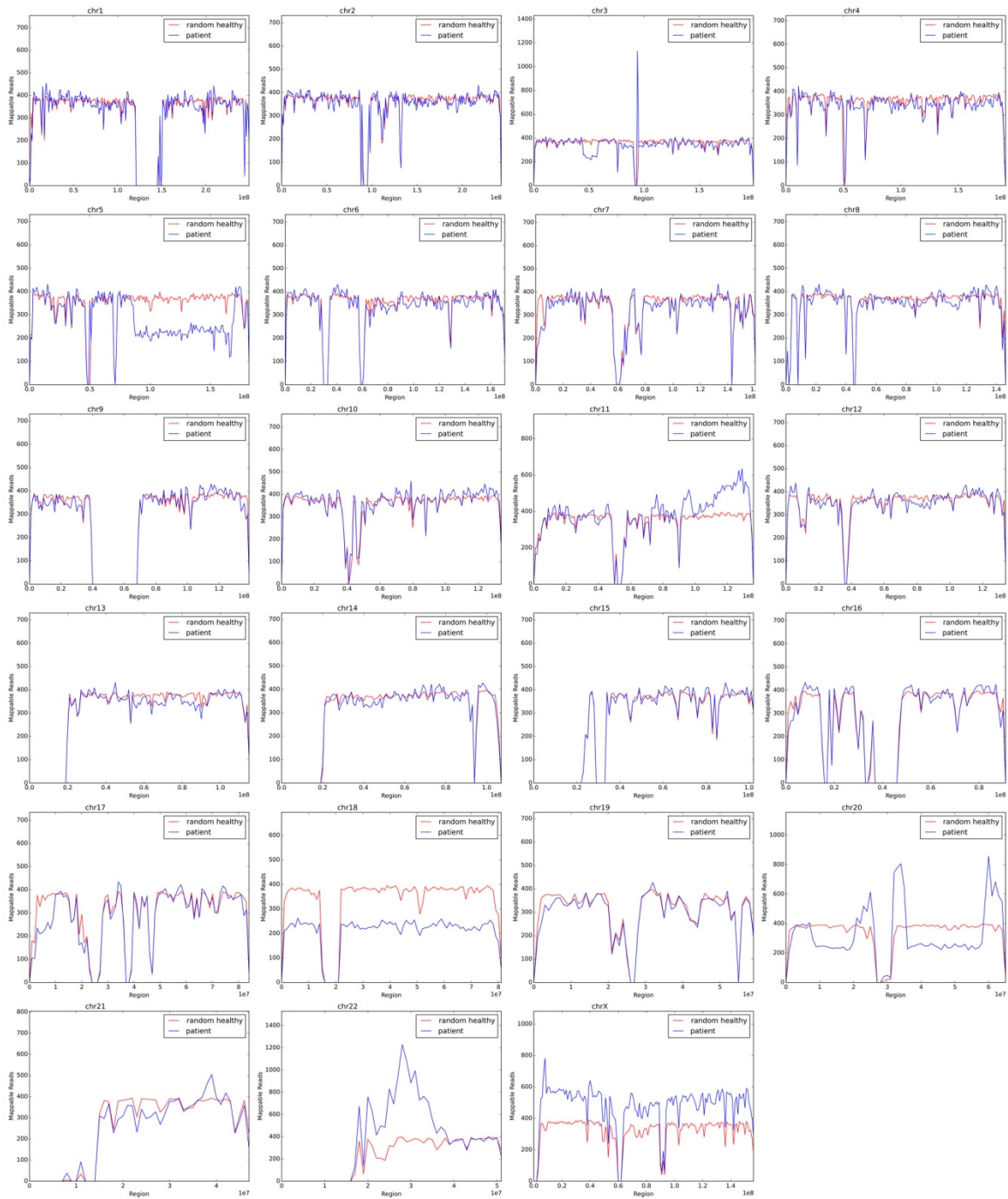


Fig. S3: Graphical representation of the CNV karyotype of patient sample AML-3.

Read numbers in 1 Mb windows were normalized to 1×10^6 total reads for the patient sample.

Random healthy: Mean read numbers for *in silico*-generated normal female karyotypes (N=2,007).

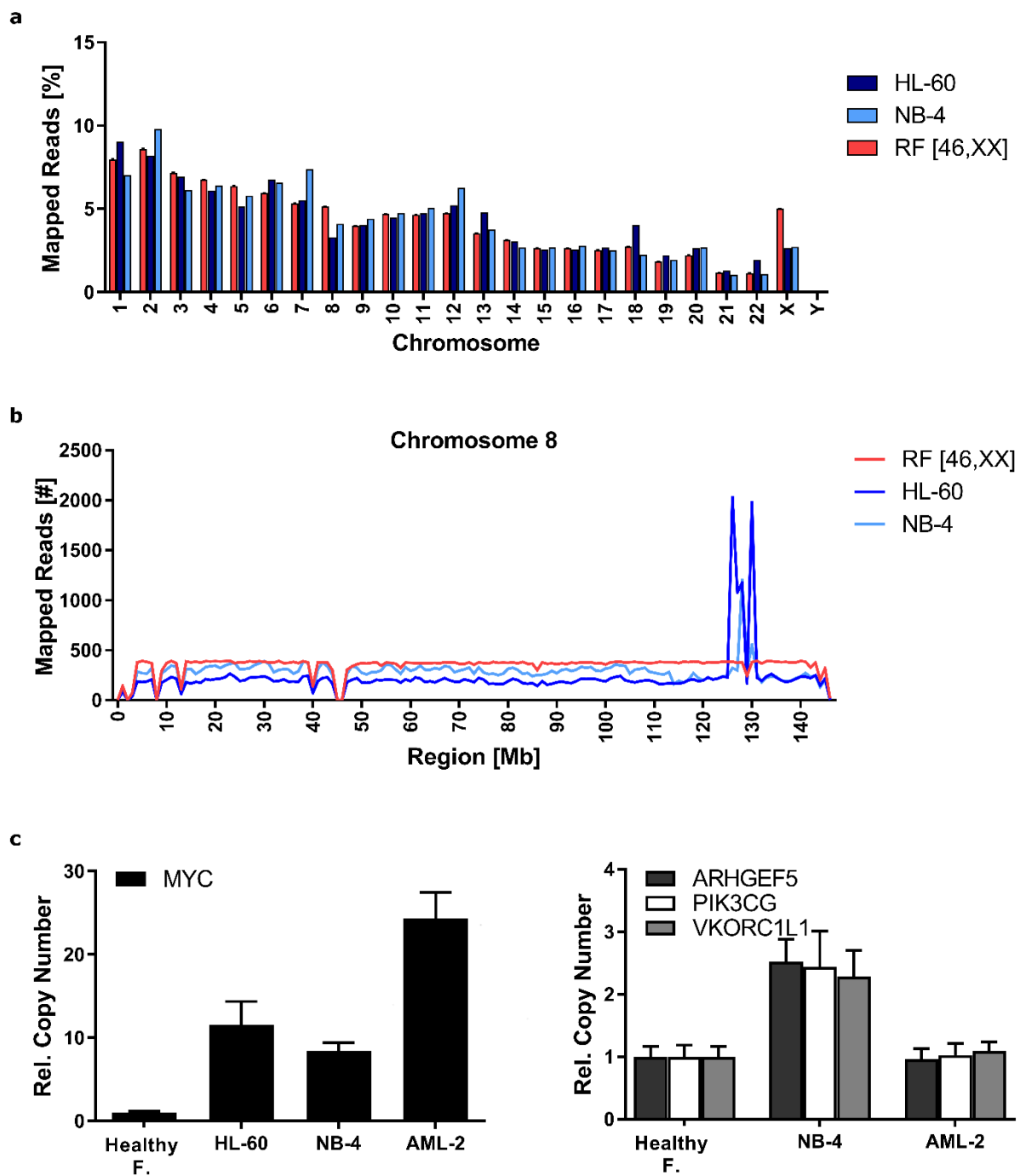


Fig. S4: Detection of double minutes and marker chromosomes by CNV karyotyping.

Whole genome libraries from HL-60 and NB-4 cells were sequenced and analyzed by CAI[N]. (a) Read frequencies on whole chromosomes. (b) Region plots for chromosome 8. Read numbers in 1 Mb windows were normalized to 1×10^6 total reads. RF: random female (N=2,819). Error bars represent SD. See also figures S5-S6 and tables S5-S6. (c) Left: qPCR validation of the *MYC* copy number gain in HL-60, NB-4 cells and AML-2 compared to a

healthy female control (Healthy F1, cf. Fig. 2). Right: qPCR validation of the copy number of 7q in NB-4 and AML-2. Error bars represent SD from triplicate assays.

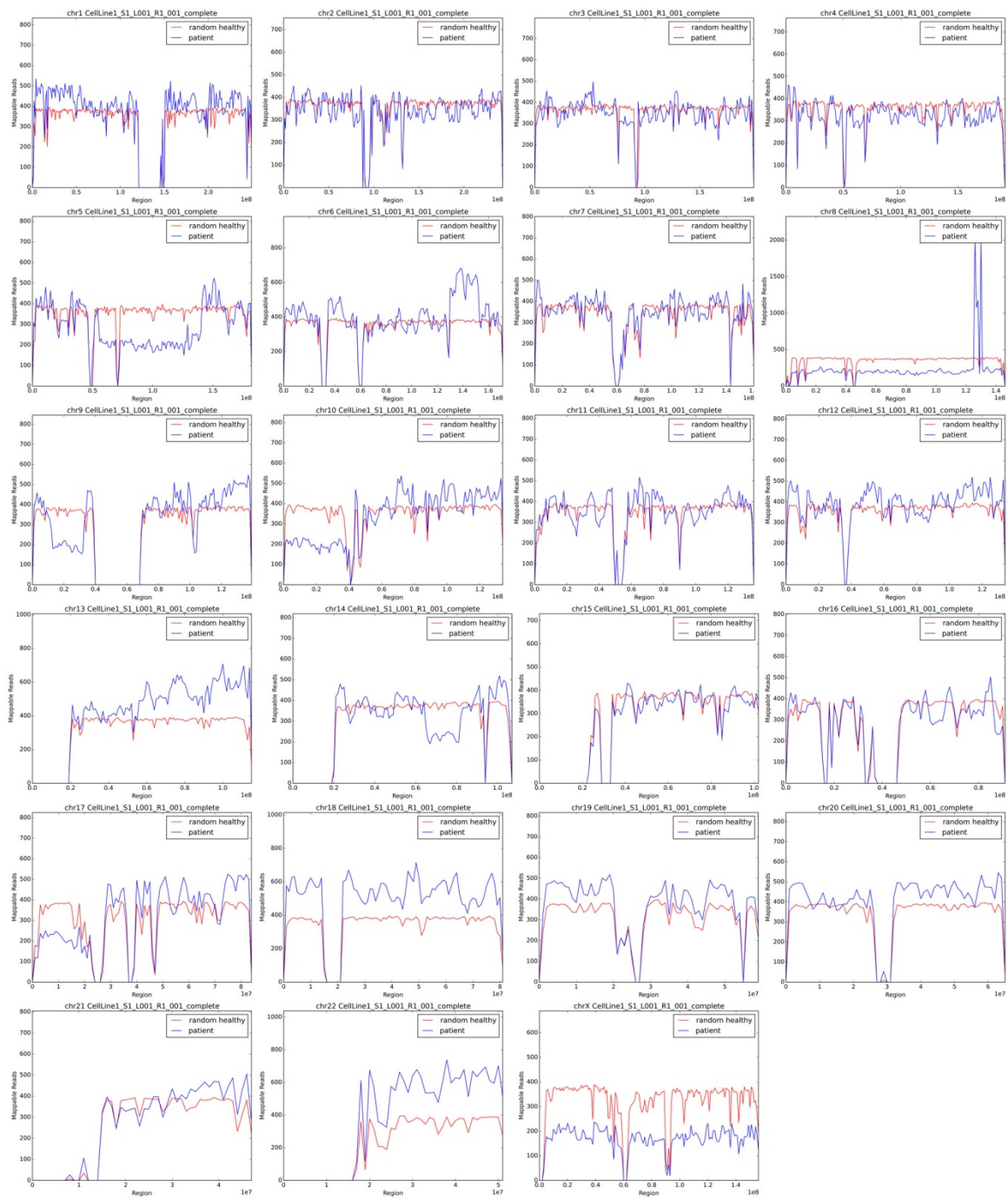


Fig. S5: Graphical representation of the CNV karyotype of HL-60 cells.

Read numbers in 1 Mb windows were normalized to 1×10^6 total reads for the cell line („patient“) sample. Random healthy: Mean read numbers for *in silico*-generated normal female karyotypes (N=2,007).

Table S5: Karyotypes of HL-60 cells.

Original	DSMZ*	CGH	FISH	NGS (CAI[N]) (raw output)	NGS (CAI[N]) (reviewed)
45; -X, -5, -8, -16, +M1, +M2, +M3	82-88 <4n>XX, -X, -X, -8, -8, -16, -17, -17, +18, +22, +2mar, ins(1;8) (p?31;q24hsr)x2, der(5)t(5;17) (q11;q11)x2, add(6)(q27)x2, der(9)del(9)(p13) t(9;14)(q?22;q?22) x2, der(14)t(9;14) (q?22;q?22)x2, der(16)t(16;17)(q22;q22)x1- 2, add(18)(q21) - sideline with: -2, -5, -15, del(11)(q23.1q23.2)	46, X, -X, dic(5;17) (q11;p11), hsr(8)(q24), del(9)(p21p23), del(9)(q32), del(10)(p12p15), del(14)(q11) add(16)(q?), der(16)t(5;16) (q31;q24), +18 +der(?)t(?;14) (?;q24)	44,X,-X, -5, dic(5;17)(q11;p11), del(7)(p?), der(7)t(5;7)(q11;q?31) t(5;16)(q11;q?), add(8)(q?), der(9)del(9)(p2?) t(9;14)(q2?;q2?), del(10)(p?), ins(11;8)(q13?;?), der(14) t(14;15)(q1?;q?), -15, der(16) t(5;16)(q?;q?22~24), der(16) t(7;16)(?;q?22~24), +18	add(1)(p36.33p32.1) add(1)(p22.2p21.2) del(1)(p21.1p21.1) add(1)(p13.3p12) add(1)(q21.1q24.2) add(1)(q25.3q25.3) add(1)(q32.1q44) del(2)(p24.1p24.1) add(2)(p23.3p23.3) del(2)(p16.3p16.1) del(2)(q22.1q22.3) del(2)(q24.1q24.3) del(2)(q32.1q32.3) del(3)(p26.3p26.1) add(3)(p25.3p25.3) add(3)(p21.31p21.1) del(3)(p13p11.1) add(3)(q21.2q22.1) add(4)(p16.1p16.1) del(4)(p15.33p15.33) del(4)(p15.1p12) del(4)(q12q13.3) del(4)(q22.1q23) del(4)(q28.1q28.3) del(4)(q31.23q32.3) del(4)(q34.1q34.3) del(5)(p14.3p14.1) add(5)(p13.3p13.3) add(5)(q11.1q11.1) del(5)(q11.2q31.2) add(5)(q31.3q33.2) add(6)(p25.3p22.2) add(6)(p21.32p21.1) add(6)(q15q15) del(6)(q16.1q16.1) add(6)(q21q21) add(6)(q23.2q25.1) add(6)(q25.3q27) add(7)(p22.3p22.1) add(7)(p13p12.2) add(7)(q11.23q11.23) del(7)(q21.11q21.11) add(7)(q32.1q32.1) add(7)(q34q36.3) del(8)(p23.2q24.12) add(8)(q24.21q24.21) del(8)(q24.22q24.3) add(9)(p24.3p24.1) del(9)(p23p21.2) add(9)(p21.1p13.1) add(9)(q21.13q21.31) add(9)(q22.1q31.1) add(9)(q31.3q34.3) del(10)(p15.1p11.21) add(10)(q11.1q11.23) add(10)(q21.2q26.3) add(11)(p15.1p15.1) del(11)(p14.3p14.3) add(11)(p13p13) del(11)(p12p12) add(11)(p11.2p11.2) add(11)(q12.1q14.1) del(11)(q14.3q22.1) add(12)(p13.33p13.1) add(12)(p12.1p11.21) add(12)(q13.11q14.1) add(12)(q23.3q24.33) add(13)(q12.13q13.1) add(13)(q14.11q34) add(14)(q11.2q13.1) del(14)(q21.2q21.2) add(14)(q21.3q21.3) del(14)(q23.2q31.1) add(14)(q31.3q32.33) del(15)(q11.2q13.1) add(15)(q15.1q15.1) add(16)(q24.1q24.1) del(16)(q24.2q24.2) del(17)(p13.3p11.2) add(17)(q11.1q25.3) add(18)(p11.32p11.21) add(18)(q11.1q23) add(19)(p13.3p12) add(19)(q13.11q13.43) add(20)(p13p11.23) add(20)(p11.21q13.33) add(21)(q22.11q22.3) add(22)(q11.1q13.33) del(X)(p22.3q28)	add(1)(p36.33p32.1) del(5)(q11.2q31.2) add(5)(q31.3q33.2) add(6)(q25.3q27) -8, add(8)(q24.21), del(9)(p23p21.2) add(9)(q22.1q34.3) del(10)(p15.1p11.21) add(10)(q21.2q26.3) +13 del(14)(q23.2q31.1) -17p +17q +18 +19 +20 +22 -X

*www.dsmz.de/catalogues/details/culture/ACC-3.html, accessed 09.04.2017

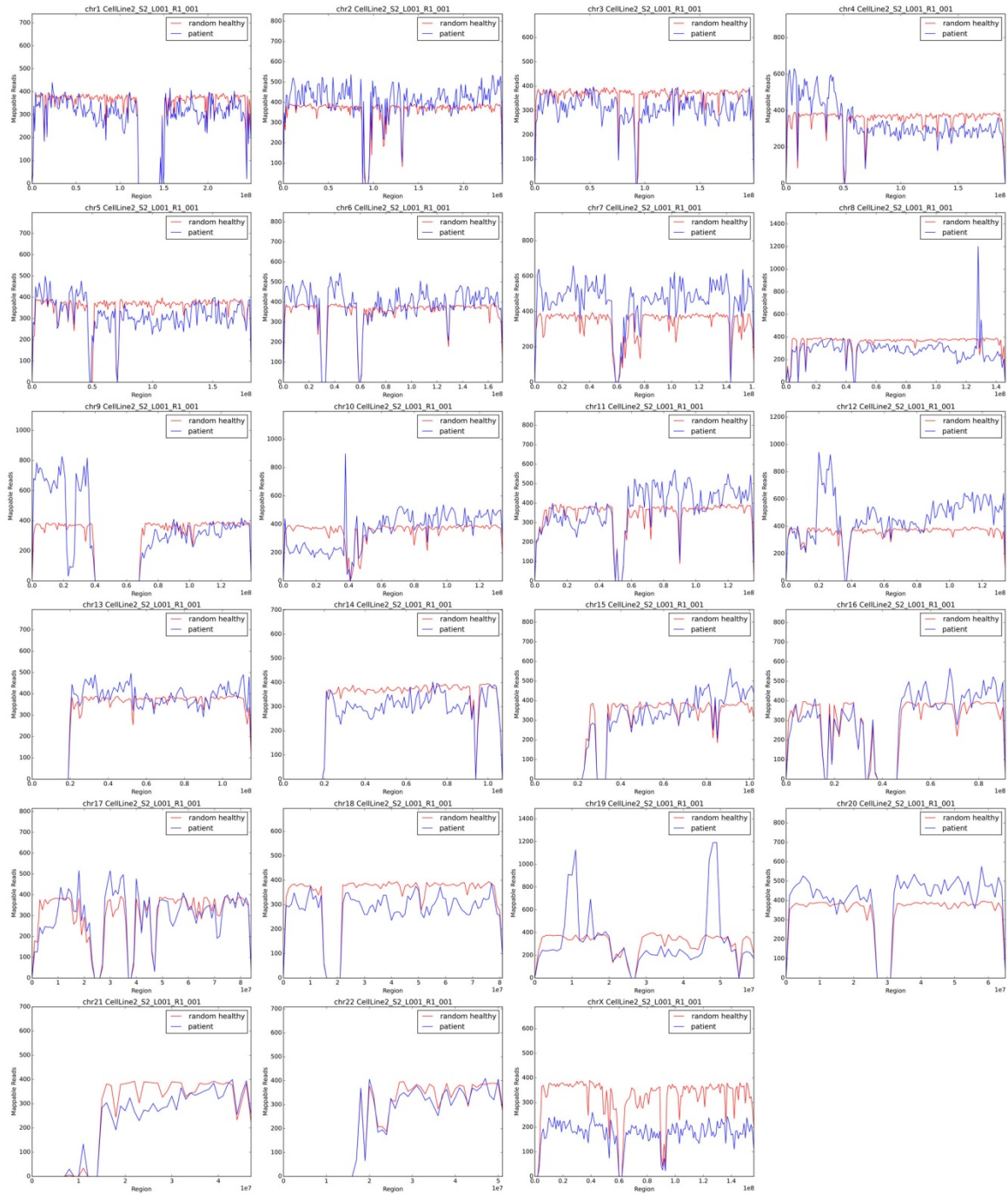


Fig. S6: Graphical representation of the CNV karyotype of NB-4 cells.

Read numbers in 1 Mb windows were normalized to 1×10^6 total reads for the cell line („patient“) sample. Random healthy: Mean read numbers for *in silico*-generated normal female karyotypes (N=2,007).

Table S6: Karyotypes of NB-4 cells.

Original	DSMZ*	CGH	FISH#	NGS (CAI[N]) (raw output)	NGS (CAI[N]) (reviewed)
80-87, XXX, -X, -3, -8, -9, -10, -12, -12, -14, -14, -18, -19, -19, der(12) t(12;?) (p12;?), t(15;17) (q22;q11- 12), t(15;17), der(19) t(19;?) +variable markers	78(71-81) <3n>XX, -X, +2, +6, +7, +7, +11, +12, +13, +14, +17, -19, +20, +4mar, der(8)t(8;?)(q24;?), der(11)t(11;?) (?->::11p15- >11q22.1::11q13- >22.1;), der(12)t(12;?) (p11;?), 14p+, t(15;17) (q22;q11-12.1), der(19)t(10;19) (q21.1;p13.3)x2	add(2)(p23- q33), del(3)(p14), add(4p), del(4) (q13.1-qter), add(5p), +7 del(8) (q21-q24.21), add(8)(q24.21), del(8) (q24.1-qter), add(9p21.1- pter), del(9)(p21.1), del(9)(qcen- q31), del(10p), add(10) (q21-q24), del(10)(q25- q26), add(11) (q12.2-q21), add(12)(p12), add(12) (q12-qter), add(15) (q22.2-qter), del(17)(p13), -18 del(19p13.2), add(19)(q13.33) del(21q-22.2)	75-78,XX, -X, +2, ins(3;3)(q26;q24q26), +i(4)(p10), +5, +del(5)(q14q35), der(5)t(5;16)(q11;q12), der(5)t(3;5)(?;q3?), +6, +7, +del(7)(q11q35-36), +der(7)t(7;17)(q11;?), der(8)t(8;9)(q23;q22), del(9)(q11q34), +i(9)(p10)x2, der(11,20), del(12)(q11), der(12)t(12;12) (p12;q15), +13, der(13;13)(q10;q10), del(13)(q13q21)del(13) (q21), der(14)t(14;19) (p11;?p11), t(15;17)(q22;q21), +16, +der(17)t(15;17) (q22;q21), der(18)t(11;18;20), +r(18), -19, der(19)t(10;19) (q21;q13)x2, +20, +22	del(1)(p33p12) del(1)(q23.1q41) add(2)(p25.1p11.2) add(2)(q11.1q22.1) add(2)(q24.3q31.3) add(2)(q32.2q37.3) del(3)(p26.3p26.1) del(3)(p24.3p22.3) del(3)(p21.1p11.1) del(3)(q11.2q13.33) del(3)(q23q26.33) add(4)(p16.3p11) del(4)(q13.1q21.23) del(4)(q22.1q35.2) add(5)(p13.3p12) del(5)(q11.2q34) add(6)(p25.3p22.2) add(6)(p21.3p21.1) add(6)(q23.2q24.2) add(6)(q25.1q27) add(7)(p22.3p11.2) add(7)(q11.1q36.3) del(8)(p23.2p21.2) del(8)(p12q24.13) add(8)(q24.21q24.21) del(8)(q24.22q24.3) add(9)(p24.3p21.3) del(9)(p21.2p21.2) add(9)(p21.1p13.3) del(9)(p13.2p13.2) del(9)(q21.11q22.2) del(9)(q22.32q31.3) del(10)(p15.1p11.1) add(10)(q11.1q26.3) del(11)(p15.4p14.1) add(11)(q12.1q25) del(12)(p13.1p13.1) add(12)(p12.3p11.21) add(12)(q12q24.33) add(13)(q12.11q13.1) add(13)(q14.11q14.3) add(13)(q32.1q33.1) del(14)(q12q24.1) del(14)(q31.1q32.11) del(15)(q11.2q13.1) del(15)(q21.1q21.3) add(15)(q24.1q26.3) del(16)(p13.3p13.12) add(16)(p12.2p12.2) add(16)(q12.1q24.3) del(17)(p13.2p13.2) add(17)(p13.1q21.2) del(18)(p11.32p11.23) del(18)(q11.1q23) del(19)(p13.3p13.3) add(19)(p13.2p13.2) del(19)(q11q13.31) add(19)(q13.32q13.33) del(19)(q13.41q13.43) add(20)(p13p11.23) add(20)(q11.21q13.33) del(21)(q21.1q22.11) del(X)(p22.33q28)	-1, +2, -3, add(4)(p16.3p11), del(4)(q13.1q35.2), del(5)(q11.2q34), +7, -8, add(8)(q24.21), add(9)(p24.3p21.3), del(9)(p21.2p21.2), add(9)(p21.1p13.3), del(9)(q21.11q31.3), del(10)(p15.1p11.1), del(10)(q11.1q26.3), add(11)(q12.1q25), +12, add(13)(q12.11q14.3), -14, del(17)(p13.2p13.2) -18, der(19) del(p13.3) add(p13.2) del(q11q13.31) add(q13.32q13.33) del(q13.41q13.43), +20, -21, -X

*www.dsmz.de/catalogues/details/culture/ACC-207.html, accessed 09.04.2017

Composite karyotype from 3 clones

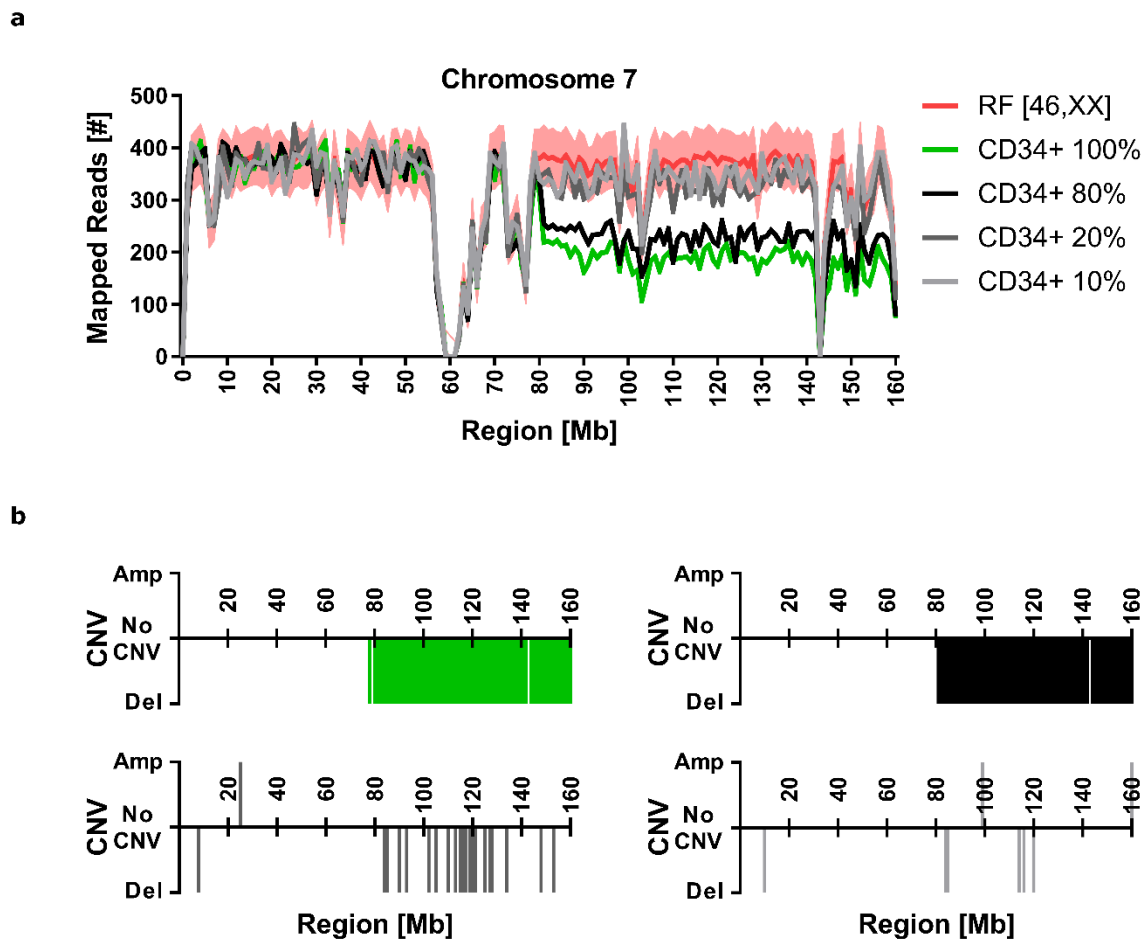
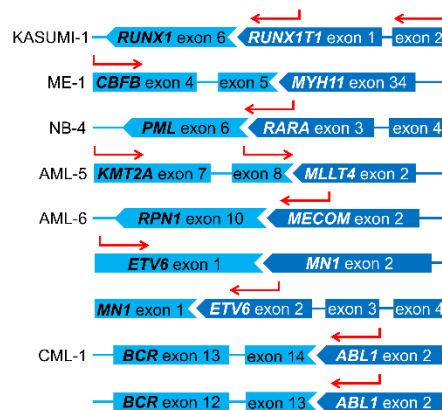


Fig. S7: Sensitivity of CNV karyotyping by NGS and CAI[N] analysis.

CD34-positive blasts were enriched from the peripheral blood of patient AML-4 with known loss of 7q by magnetic bead separation. DNA from the CD34-positive fraction was diluted in healthy donor DNA (Healthy F1, cf. Fig. 2) to yield samples with 100, 80, 20 and 10% blast DNA content. The dilution series was subjected to CNV karyotyping by low coverage whole genome sequencing and CAI[N] analysis. **(a)** Region plots for chromosome 7. The range $\pm 3 \cdot \text{SD}$ 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).

a



b

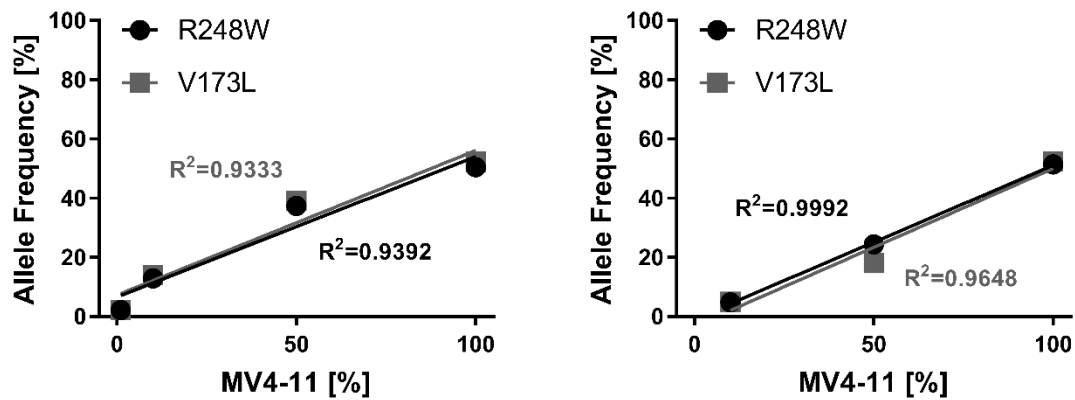


Fig. S8: Fusion and SNV detection in AML cell lines using amplicon panels.

(a) Fusion analysis using the FusionPlex[®] Heme panel (v1 or v2, cf, Table S7) in AML cell lines and patients. Graphical representations of the detected fusions were redrawn from the Archer[®] Analysis software (not to scale). (b) Sensitivity of SNV detection. Dilutions of MV4-11 DNA in DNA from a healthy donor (Healthy F1, cf. Fig. 2) were sequenced using the TruSight[®] Myeloid panel with BaseSpace[®] analysis (left) or the QIAseq[™] Human Myeloid Neoplasms panel (Qiagen) with smCounter (right).

Table S7: Fusion analysis using the FusionPlex® Heme panel.

Cell lines and patient samples AML-5, -14, -21, CML-1, and HES-1 were analyzed using the FusionPlex® Heme v1 panel with Archer Analysis software v4.1 with standard or, respectively, advanced settings (CML-1), samples AML-6, 16, -22-26 and ALL-1 were analyzed using the FusionPlex® Heme v2 panel with Archer Analysis software v5.0 with advanced settings. See Table 1 for more detailed patient and sample information. If several fusions involving the same fusion partners were detected, only the top hit is indicated (exception: BCR-ABL1 in CML-1).

Sample	Reference Karyotype	Fusions (NGS)	Supporting Evidence		Read Statistics		On Target (%)
			Reads (#/%)	Start Sites	Total*	Unique	
KASUMI-1 ¹	45<2n>X,-Y,-9,-13,-16,+3mar,t(8:21)(q22;q22),der(9)t(9:?)p22;?) ,der(15)t(9;15)(?q11;?p11)	RUNX1-RUNX1T1	273 / 41.6	92	1,965,129	114,172	98.95
ME-1 ¹	48(45-48)<2n>XY,+8,+20,der(2)t(2:11)(q31;q13), der(9)t(9:18)(q12.2;q11),der(16)t(11;16)(q13;q22)inv(16)(p13q22),der(17)t(2;17)(p25;p12)	CBFB-MYH11	185 / 25.7	93	1,878,486	153,166	99.09
NB-4 ¹	78(71-81)<3n>XX,-X,+2,+6,+7,+7,+11,+12,+13,+14,+17,-19,+20,+4mar,der(8)t(8;?)q24;?),der(11)t(11;?)q22.1;11p15->11q22.1::11q13->22.1;),der(12)t(12;?)p11;?),14p+,t(15;17)(q22;q11-12.1),der(19)t(10;19)(q21.1;p13.3)x2	PML-RARA	64 / 44.4	27	1,793,403	78,902	98.41
HL-60 ¹	82-88<4n>XX,-X,-X,-8,-8,-16,-17,-17,+18,+22,+2mar,ins(1;8)(p?31;q24hsr)x2,der(5)t(5;17)(q11;q11)x2,add(6)(q27)x2,der(9)del(9)(p13)t(9;14)(q?22;q?22)x2,der(14)t(9;14)(q?22;q?22)x2, der(16)t(16;17)(q22;q22)x1-2,add(18)(q21)	none found	n.a	n.a	1,348,786	92,217	98.8
CL-Mix	KASUMI-1 + ME-1 + NB4 + HL-60	RUNX1-RUNX1T1 CBFB-MYH11 PML-RARA	466 / 67.3 60 / 5.0 14 / 8.5	164 45 14	2,066,340	150,899	98.95
CL-Mix 1:25	KASUMI-1 + ME-1 + NB4 + HL-60 + 46,XX	RUNX1-RUNX1T1	29 / 100	19	1,485,550	100,989	99.07
AML-5	46,XX,t(6:11)(q27;q23) [19]; 46,XX [1]	KMT2A-MLL2	551 / 74.2	201	1,692,760	137,378	99.08
AML-6 ²	46,XX,inv(3)(q21q26); 46,idem,t(12:22)(p12;q12)	RPN1-MECOM ETV6-MN1	46 / 16.7 11 / 1.6	26 10	661,172	353,037	89.9
AML-14	n.a (presumably 46,XY)	none found	n.a	n.a	1,308,041	285,625	34.71
AML-16 ³	46,XX,t(8:21)(q22;q22) [6], 45,X,-X,t(8:21)(q22;q22) [12],46,XX [2]	RUNX1-RUNX1T1	1522 / 39.8	186	1,544,161	348,642	91.0
AML-21	46,XX,del(7q),t(8:21)(q22;q22)	RUNX1-RUNX1T1	147 / 20.7	53	2,229,896	160,953	99.24
AML-22	48-49,XY,+19,+der(21)X1-2[cp4]; 46,XY[7]	No known fusions	n.a	n.a	2,006,916	266,201	93.3
AML-23	46,XY [6];46,XY, inv(16)(p13q22) [18]	CBFB-MYH11	183 / 36.0	101	1,679,072	207,025	89.1
AML-24	45,XY, -7 [7], 46, XY, -7 +21 [9], 46, XY [5]	none found	n.a	n.a	2,018,647	185,625	87.7
AML-25	46,XX [20]	none found	n.a	n.a	1,726,674	194,012	86.9
AML-26	nuc ish 5p15(hTERTx3),5q31 (CDC25C,EGR1x3), 8cen(D8Z2x4), 17p13(TP53x3), 17q11(NF1x3)	No known fusions	n.a	n.a	2,882,054	322,859	91.3
CML-1	46,XX, t(9:22)(q34;q11) [20]	BCR-ABL1 (b3a2) BCR-ABL1 (b2a2)	23 / 29.9 11 / 14.3	7 7	2,499,730	109,532	98.15
HES-1	46, XY [15]	none found	n.a	n.a	1,508,621	176,621	44.37
ALL-1	47,XX, t(1:9)(q22;q34),+17 [3],	RCSD1-ABL1	908 / 44.9	211	1,087,906	196,031	89.0

46,XX,t(1:9)(q22;q34),der(6) t(6;17)(p23;q21) [2]							
--	--	--	--	--	--	--	--

* Total fragments after trimming adapters

¹Reference karyotypes for cell lines from www.dsmz.de/catalogue; accessed 14.01.2018 (KASUMI-1, ME-1) and 09.04.2017 (NB4, HL-60): www.dsmz.de/catalogues/details/culture/ACC-220.html, www.dsmz.de/catalogues/details/culture/ACC-537.html, www.dsmz.de/catalogues/details/culture/ACC-207.html, www.dsmz.de/catalogues/details/culture/ACC-3.html

² Sample AML-6: The sample was taken after third line treatment, which led to a partial remission. Only presumed persisting clones are indicated; cf. also Table 1.

³ For detection of the *RUNX1-RUNX1T1*-fusion transcript the following additional changes were made in the advanced analysis settings: DETERMINISTIC_NORMALIZATION: Off (default On), MIN_READS_FOR_VALID_FUSION: 3 (default 5), MIN_UNIQUE_START_SITES_FOR_VALID_FUSION: 2 (default 3).

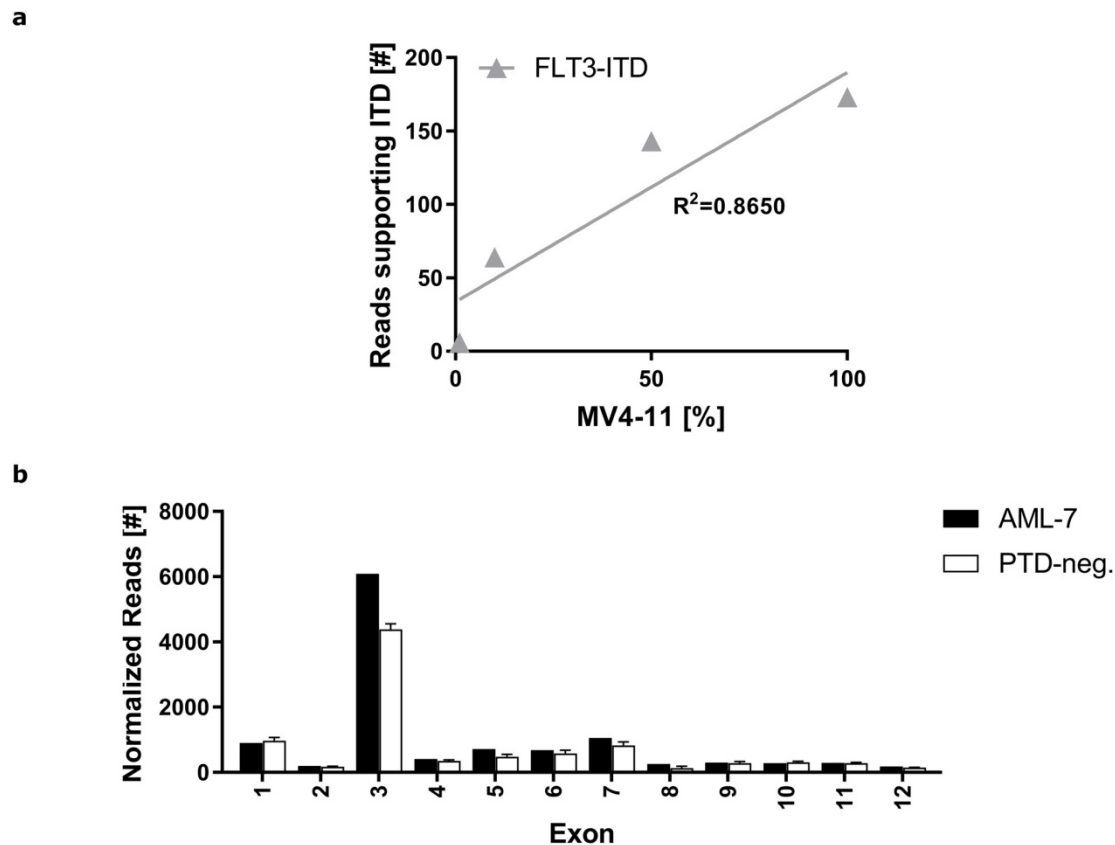


Fig. S9: Detection of *FLT3*-ITDs and *KMT2A*-PTDs in AML samples by targeted amplicon resequencing.

(a) Sensitivity of ITD detection. The MV4-11 dilution series sequenced using the TruSight[®] Myeloid panel (cf. Fig. S8c) was analyzed by ITD-seek. The ITD was detectable down to a dilution of 10%. The analytical sensitivity of ITD detection using the QIAseq[™] Human Myeloid Neoplasms panel with smCounter was >10% (not shown). (b) Detection of *KMT2A*-PTDs using the QIAseq[™] Targeted DNA Human Myeloid Neoplasms panel and PTDi. Normalized read numbers ($n_{\text{exon}}/n_{\text{library}}$) for exon 1-12 are shown, error bars for the controls represent SD.

Table S8: Mutation analysis using amplicon panels in AML cell lines.

Cell lines were analyzed using the TruSight® Myeloid panel.

Cell Line	Gene	Variant (Protein)	VAF (%)	Confirmation	Total Aligned Reads (R1/R2)	Overall Aligned Reads (%)	Mean Amplicon Coverage
HL-60	NRAS	Q61L	52.9	COSMIC	3,064,340 / 3,032,894	94.86	10,689.6
NB-4	KRAS	A18D	73.7	COSMIC	3,091,605 / 3,041,957	96.13	10,743
	TP53	R248Q	99.0				
MV4-11	TP53	V173L	52.2	Sanger Sequencing	3,185,629 / 3,177,326	95.94	11,186.6
	TP53	R248W	50.4				
ME-1	NRAS	Q61H	53.4	Sanger Sequencing	3,138,570 / 3,073,974	94.79	10,894.2
SKNO-1	KIT	N822K	99.1	Sanger Sequencing	3,730,770 / 3,668,524	95.39	12,980.9
	TP53	R248Q	99.2				
Variant not passing filters (coverage < 100 reads)							
HL-60	CDKN2A	R80*	100	COSMIC			

Table S9: Mutation analysis using amplicon panels in AML patient samples.

Patient samples AML-5, -14, -16 and -20-25 were analyzed using the TruSight® Myeloid panel, samples AML-7 and AML-8 were analyzed using the QiaSeq™ Myeloid Neoplasms panel. See. Table 1 for more detailed patient and sample information.

Pat. ID	Mutations Ref. Lab. (VAF %)	Mutations NGS (VAF%)	Total Aligned Reads (R1/R2)	Overall Aligned Reads (%)	Mean Amplicon Coverage
AML-5	negative for NPM1 mut, FLT3-ITD,FLT3-TKD,KMT2A-PTD, CEBPA mut.	negative for NPM1 mut, FLT3-ITD,FLT3-TKD,KMT2A-PTD, CEBPA mut	4,106,754 / 4,074,562	95.5	14,380.5
AML-12	negative for NPM1 mut, FLT3-ITD,FLT3-TKD,KMT2A-PTD	negative for NPM1 mut, FLT3-ITD,FLT3-TKD,KMT2A-PTD	4,149,802 / 4,111,370	96.21	14,520.4
AML-14	negative (ASXL1 S191L reported at initial diagnosis)	ASXL1 S191L not included in target region	3,412,323 / 3,412,576	94.77	12,015.7
AML-16	negative for NPM1 mut, FLT3-ITD, FLT3-TKD, KMT2A-PTD	negative for NPM1 mut, FLT3-ITD,FLT3-TKD,KMT2A-PTD	10,732,647 / 10,621,780	97.69	37,595.8
AML-21	negative for FLT3-ITD	negative	1,288,554 / 1,307,140	84.08	4,560.8
AML-22	n.a	negative	1,576,513 / 1,591,410	78.89	4,146.8
AML-23	n.a	negative	965,806 / 980,437	58.37	3,416.5
AML-24	negative for NPM1 mut, FLT3-ITD	negative	1,172,349 / 1,186,060	62.81	4,146.8
AML-25	negative for NPM1 mut, FLT3-ITD	negative	2,036,154 / 2,046,639	72.95	7,177.2
HES-1	negative	negative	4,854,912 / 4,868,471	97.41	17,118.2
Pat. ID	Mutations Ref. Lab. (VAF %)	Mutations NGS (VAF%)	Total Read Fragments	Reads with Primer On Target (%)	Mean Molecular Tag Depth
AML-7	KMT2A-PTD (Ratio 120,597), IDH1 R132C (28%), SRSF2 P95L (24%)	positive (cf. Fig. S9) IDH1 R132C (25,53%), SRSF2 P95L (27,91%),	1,955,685	96.14	329.45
AML-8	IDH1 R132G (37%) NPM1 W288C fs*12 (21%)	IDH1 R132G (45,34%); NPM1 W288fs (40,7%)	1,521,345	95.58	270.86

Table S10: Cytogenetic risk classification by chromosome analysis and NGS: samples without risk-defining translocations.

See Table 1 for conventional and NGS karyotypes.

Sample	Numerical Aberrations				Karyotype Class		Cytogenetic Risk	
	Ref.	NGS	Concordant	Discordant	Ref.	NGS	Ref.	NGS
AML-1	6	7	6	1	complex	complex	adverse	adverse
AML-2	6	5	4	2	complex	complex	adverse	adverse
AML-3	4	8	4	4	complex	complex	adverse	adverse
AML-4	2/1	1	1 / 1	1 / 0	n.a (aberrant with derivative chromosome and -7q)	aberrant with -7q	intermediate	intermediate
AML-7	0	0	yes	n/a	aberrant / numerically normal with balanced translocation	numerically normal	intermediate	intermediate
AML-8	0	0	yes	n/a	normal	numerically normal	intermediate	intermediate
AML-9	0	0	yes	n/a	normal	numerically normal	intermediate	intermediate
AML-12	1	2	1	1	n.a (normal)	aberrant / non complex with -7q	intermediate	intermediate
AML-14	n/a	0	likely (100% chimerism)	n/a	n.a	numerically normal	No evidence of disease	
AML-15	n/a	0	likely	n/a	n.a (normal)	numerically normal	intermediate	intermediate
AML-17	0	0	yes	n/a	normal	numerically normal	intermediate	intermediate
AML-18	0	0	yes	n/a	n.a (normal)	numerically normal	intermediate	intermediate
AML-20	n/a	0	likely	n/a	n.a (normal)	numerically normal	intermediate	intermediate
AML-22	2	2	2	0	aberrant with two chromosomal changes	aberrant with two chromosome gains	intermediate	intermediate
AML-24	2	2	2	0	aberrant with -7 and chromosome gain	aberrant with -7 and chromosome gain	adverse	adverse
AML-25	0	0	0	yes	normal	numerically normal	intermediate	intermediate

Table S11: Cytogenetic risk classification by chromosome analysis and NGS: samples with risk-defining translocations.

See Table 1 for conventional and NGS karyotypes.

Sample	Numerical Aberrations				Karyotype Class		Cytogenetic Risk	
	Ref.	NGS	Concordant	Discordant	Ref.	NGS	Ref.	NGS
AML-5	0	0	yes	n/a	aberrant / numerically normal with t(v;11)	numerically normal	adverse	adverse*
AML-6	0	0	yes	n/a	aberrant / numerically normal with inv(3) and balanced translocation	numerically normal	adverse	adverse*
AML-10	0	0	yes	n/a	aberrant / numerically normal with t(v;11)	numerically normal	adverse	n/a**
AML-11	1	1	1	0	aberrant with t(8;21) and chromosome gain	aberrant with chromosome gain	favorable	n/a**
AML-13	0	0	yes	n/a	aberrant / numerically normal with inv(16)	numerically normal	favorable	n/a**
AML-16	1	1	1	0	aberrant with t(8;21) and chromosome loss	aberrant with chromosome loss	favorable	favorable*
AML-19	0	0	yes	n/a	aberrant with t(15;17) and balanced translocation	numerically normal	favorable	n/a**
AML-21	1	1	1	0	aberrant with t(8;21) and -7q	aberrant with -7q	favourable	favourable*
AML-23	0	0	yes	n/a	aberrant / numerically normal with inv(16)	numerically normal	favourable	favourable*
AML-28	0	1	0	1	aberrant / numerically normal with t(9;11)	aberrant with chromosome gain	intermediate	n/a**

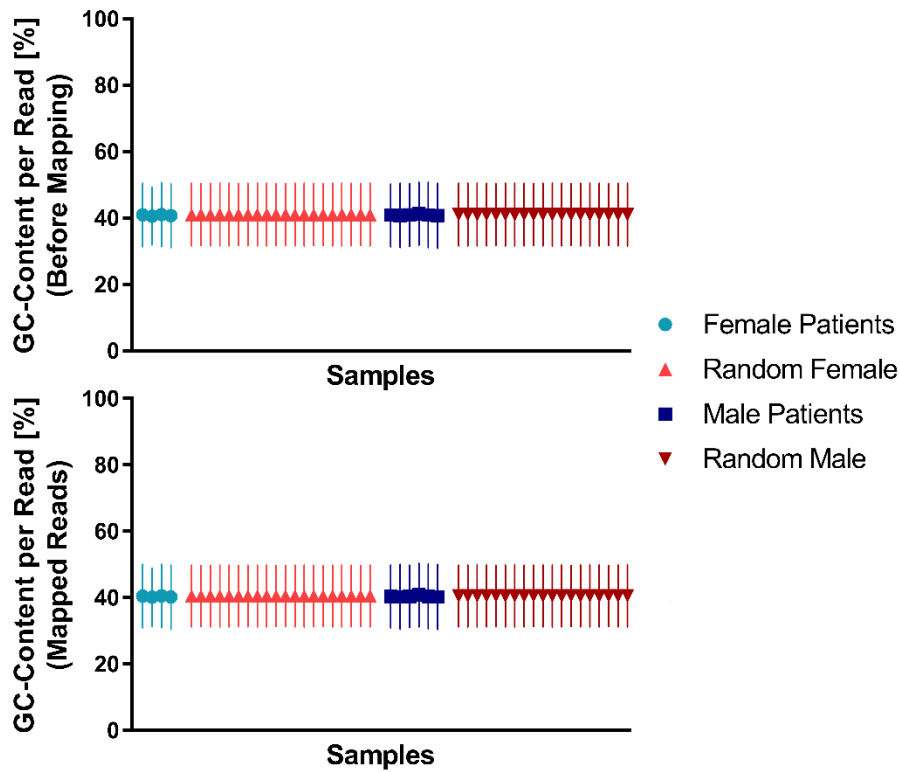
* Risk group considering numerical karyotyping and fusion analysis.
**Fusion analysis not performed.

Table S12: Normality tests of mapped read frequencies in 1 Mb windows.

D'Agostino-Pearson and Shapiro-Wilk normality tests were performed using the SciPy statistical functions module. Windows with $\text{Mean}_{\text{window}} \leq 0.1 \cdot \text{Mean}_{\text{all windows on chromosome}}$ were excluded from the analyses. $N_{\text{Random Female}}=2,819$, $N_{\text{Random Male}}=2,605$.

Chr.	Random Female										Random Male									
	Total #	D'Agostino-Pearson				Shapiro-Wilk				Total #	D'Agostino-Pearson				Shapiro-Wilk					
		Passed		Not Passed		Passed		Not Passed			Passed		Not Passed		Passed		Not Passed			
		#	%	#	%	#	%	#	%		#	%	#	%	#	%	#	%		
1	222	218	98.2	4	1.8	218	98.2	4	1.8	222	219	98.6	3	1.4	218	98.2	4	1.8		
2	235	233	99.1	2	0.9	233	99.1	2	0.9	235	232	98.7	3	1.3	232	98.7	3	1.3		
3	196	192	98.0	4	2.0	191	97.4	5	2.6	196	194	99.0	2	1.0	196	100.0	0	0.0		
4	188	184	97.9	4	2.1	183	97.3	5	2.7	188	184	97.9	4	2.1	183	97.3	5	2.7		
5	178	175	98.3	3	1.7	177	99.4	1	0.6	178	176	98.9	2	1.1	176	98.9	2	1.1		
6	164	159	97.0	5	3.0	159	97.0	5	3.0	164	159	97.0	5	3.0	159	97.0	5	3.0		
7	156	153	98.1	3	1.9	154	98.7	2	1.3	156	154	98.7	2	1.3	155	99.4	1	0.6		
8	141	141	100.0	0	0.0	141	100.0	0	0.0	141	139	98.6	2	1.4	137	97.2	4	2.8		
9	110	107	97.3	3	2.7	106	96.4	4	3.6	110	108	98.2	2	1.8	109	99.1	1	0.9		
10	133	130	97.7	3	2.3	131	98.5	2	1.5	133	132	99.2	1	0.8	132	99.2	1	0.8		
11	131	128	97.7	3	2.3	127	96.9	4	3.1	131	130	99.2	1	0.8	129	98.5	2	1.5		
12	132	130	98.5	2	1.5	130	98.5	2	1.5	132	130	98.5	2	1.5	131	99.2	1	0.8		
13	96	96	100.0	0	0.0	96	100.0	0	0.0	96	96	100.0	0	0.0	96	100.0	0	0.0		
14	86	86	100.0	0	0.0	85	98.8	1	1.2	86	83	96.5	3	3.5	84	97.7	2	2.3		
15	75	73	97.3	2	2.7	74	98.7	1	1.3	75	74	98.7	1	1.3	72	96.0	3	4.0		
16	76	74	97.4	2	2.6	74	97.4	2	2.6	76	76	100.0	0	0.0	76	100.0	0	0.0		
17	79	76	96.2	3	3.8	76	96.2	3	3.8	79	77	97.5	2	2.5	77	97.5	2	2.5		
18	75	73	97.3	2	2.7	73	97.3	2	2.7	75	75	100.0	0	0.0	75	100.0	0	0.0		
19	56	55	98.2	1	1.8	55	98.2	1	1.8	56	56	100.0	0	0.0	56	100.0	0	0.0		
20	60	59	98.3	1	1.7	58	96.7	2	3.3	60	58	96.7	2	3.3	58	96.7	2	3.3		
21	34	32	94.1	2	5.9	32	94.1	2	5.9	34	33	97.1	1	2.9	32	94.1	2	5.9		
22	35	34	97.1	1	2.9	33	94.3	2	5.7	35	35	100.0	0	0.0	33	94.3	2	5.7		
X	151	149	98.7	2	1.3	148	98.0	3	2.0	151	142	94.0	9	6.0	139	92.1	12	7.9		
Y										21	15	71.4	6	28.6	15	71.4	6	28.6		
Total	2,809	2,757	98.1	52	1.9	2,754	98.0	55	2.0	2,830	2,777	98.1	53	1.9	2,770	97.9	60	2.1		

a



b

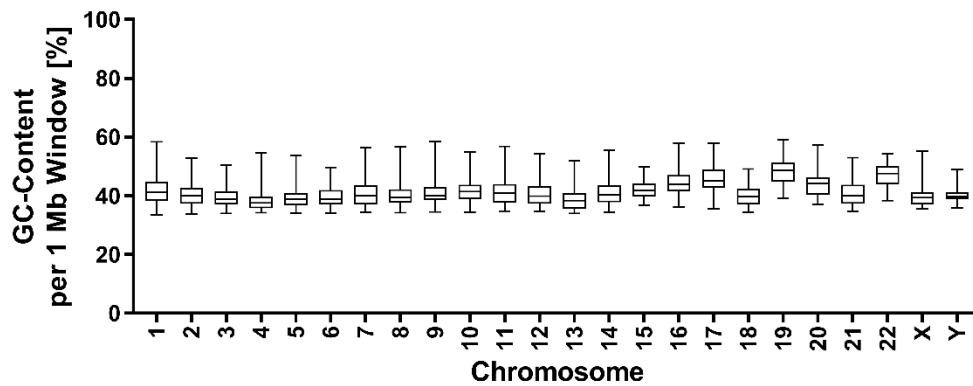


Fig. S10: GC-content of *in-silico* reference reads and experimental sequencing reads.

(a) GC-content of 150 bp reads before mapping and after alignment to the reference genome for random normal female and male karyotypes (N=20 each) and patient samples (female: AML-6, -10, -11b, -17; male: AML-2, -9, -15, -19, -20, HES-1; cf. Table 1). Analysis of reference samples started with approx. 1×10^6 (1,080,784 – 1,089,871) random reads, analysis of experimental samples included all high quality reads from fastq.files (1,290,831 – 4,993,971). The GC-content for each read was calculated relative to all identifiable bases, i.e

excluding n's. Mean and SD are indicated. (b) GC-content in 1 Mb windows of the reference genome. The GC-content in each window was calculated relative to all known bases, windows containing unknown bases (n's) only were excluded. Whiskers indicate the range of data points (minimum to maximum).