

Supplementary subject and methods

Patients

Samples were collected from 3 Institutions: the Department of Experimental, Diagnostic and the Specialty Medicine of University of Bologna (Italy, n=121) in Italy, the Research Center for Molecular Medicine (CeMM) of the Austrian Academy of Sciences (n=161, of which 98 were previously published^{1,2}, genomic data not released) for samples from Austria, Czech Republic, and Serbia, and the University of Michigan (US, n=114, all previously published³, GSE23452). Samples with acute promyelocytic leukemia were excluded.

REDCap was hosted at Istituto Seràgnoli (Department of Experimental Diagnostic and Specialty Medicine, University of Bologna, Italy).

Cells and DNA isolation

Pre-treatment bone marrow and/or peripheral blood cells were processed by Ficoll-Hypaque. DNA was extracted using AllPrep DNA/RNA MiniKit (QIAGEN) in accordance with manufacturer's instructions. DNA quality and quantity were assessed using the NanoDrop Spectrophotometer (NanoDrop Technologies).

Samples from CeMM and University of Michigan were processed as previously described¹⁻³.

SNP Array protocol

DNA from AML samples was processed through the following steps: digestion, ligation, amplification and purification, fragmentation, labeling, hybridization, washing, staining and finally scanned to obtain the CEL files, according to manufacturer's instruction. The fragmentation protocol of the group of samples performed at University of Bologna has been adjusted by increasing the volume of 10x Fragmentation Buffer and Fragmentation Reagent in order to improve the efficiency of DNA fragmented. We used an adjusted fragmentation protocol, with a bigger volume of 10x Fragmentation Buffer and Fragmentation Reagent in order to obtain a better quality of DNA fragmented, under Affymetrix Technical Support's suggestion.

Detection of chromothripsis

Chromothripsis occurred in 1 or 2 chromosomes per patient characterized by a cluster of breakpoints, regularity of oscillating CN states (2-3 CN states, e.g. from heterozygous deletions to amplifications or from heterozygous deletions to amplifications in more than one copy or to homozygous deletions) within 10 subsequent rearrangements interspersed in diploid regions and a high and variable number of breakpoints.

Chromothripsis events were detected as described in methods, then confirmed by visual inspection using both Nexus CN Software v. 8.0 (BioDiscovery) and R package "Rawcopy" 36, in order to verify the presence of a CN-LOH in the B-allele frequency (BAF), to assess technical quality (hybridization level and quality of the physical array) and to obtain an extensive overview of chromosomal aberrations.

SNP microarray analysis

The threshold of CN gain and loss were set at 0.15 and -0.15, respectively. Copy Neutral LOH or UPD (Uniparental Disomy) was defined as a region displaying LOH without a CN loss. Chromosomal CNA of at least 1 kb and with a minimum of 8 probes per segment were considered and sex chromosomes were excluded from genomic analysis for the lack of paired normal controls for all cases.

Microarray statistical analyses

CEL file reports were exported from Nexus CN (BioDiscovery) v. 8.0 and went through statistical analysis using R v3.3.2⁴ and Bioconductor v3.4 (BiocInstaller 1.24.0) with following packages: "org.Hs.eg.db" v3.4.0⁵, "reactome.db" v1.58.0⁶, "clusterProfiler" v3.2.11⁷, "ReactomePA" v1.18.1⁸

All p-values were adjusted for multiple testing with Benjamini-Hochberg method⁹. Fisher's exact test was used to compare frequencies in genes' event between two groups. Genes which are not reported in Atlas of Genetics and Cytogenetics in Oncology and Haematology¹⁰ and which did not contain at least one event at single gene's level with an adjusted p-value lower than 10^{-4} in the Fisher exact test comparison were filtered out. For testing at a pathway level, genes were annotated in the Reactome database¹¹. Firstly, pathway enrichment analysis was performed at patient level by means of an over-representation test (based on hypergeometric distribution)⁷. Then, the adjusted p-values obtained for a certain pathway across all patients were used as predictor variable in a logistic regression model fitted against the case/control classification as dependent variable (0=case, 1=control, ctrl); p-values from all the performed logistic regression tests were in turn adjusted for multiple testing. The significance level was set at 10^{-4} (adj-p < 10^{-4} ; CI 99.9999%).

Non-coding genes (LINC- and -IT genes), or genes considered not detectable by the limits of SNP array because involving RNA transcripts (microRNAs, small nucleolar RNAs, antisense RNAs, small cajal body-specific RNAs) or with uncertain function (LOC genes) and highly recurrent as a CN Variant (olfactory receptors, mucins, keratins, ryanodine receptors, cub and sushi multiple domain proteins, neurexins, contactins) were filtered out from the list of CNA.

Molecular analyses

The mutational status of exons 5-9 of *TP53*, exon 12 of *NPM1*, exons 13-15 and 20 of *FLT3* was determined using specific primers for qualitative PCR. Mutated samples were then confirmed by Sanger sequencing, using the same PCR primers. *CEBPA*, *IDH1*, *IDH2* and *TP53* mutations were directly investigated by PCR and Sanger sequencing. Presence of *FLT3*-tyrosine kinase domain (TKD) mutations was determined by digestion with specific restriction enzymes (Promega), for the detection of *FLT3*-internal tandem duplication (ITD), insertions in *NPM1* and *DNMT3A* mutations, a denaturing high performance liquid chromatography screening (D-HPLC; Transgenomic) was done, followed by direct Sequencing of D-HPLC-positive samples. The expression of WT1 was quantified with a Real Time PCR assay (Ipsogen WT1 ProfileQuant Kit) using *ABL* as control gene (sensitivity 10^{-4}).

In the group of samples from CEMM and University of Michigan, the mutational status of the same genes described above plus *RUNX1*, *CBL*, *NRAS* were assessed as previously described¹⁻³.

Chromosome Banding Analysis

CBA was performed on bone marrow cells after short-term cultures (24 and/or 48 hours). Briefly, the cells were treated with colchicine and hypotonic solution and the pellet was fixed and washed in methanol/acetic acid (3:1). The cells were then re-suspended in fixative and dropped on slides. Karyotypes were examined after G banding technique and described according to International System for Human Cytogenetic Nomenclature (ISCN 2016)¹². Complex karyotype was defined as three or more chromosomal abnormalities in the same clone.

FISH

To characterize the chromosomes involved in chromothripsis, we used whole chromosome painting probes specific for chromosomes 5 and 12, (Kreatech, LeicaBiosystem, Wetzlar, Germany), LSI *MYC* Dual Color, Break Apart Rearrangement Probe (Vysis, Abbott Molecular, IL, USA), LSI *MLL* Dual Color, Break Apart Rearrangement Probe (Vysis) *CEP17* Spectrum Green/LSI *TP53* Spectrum Orange (Vysis), *EVI1* (MECOM) Tricolor Breakapart Probe (Cytocell, Cambridge, UK). The slides were counterstained with DAPI and analyzed using fluorescent microscopes equipped with FITC/TRITC/AQUA/DAPI filter sets and the Genikon imaging system software (Nikon Instruments, Tokyo, Japan).

Clinical Statistical analysis

Clinical data collection included age at diagnosis, de novo/secondary AML, White Blood Cells (WBC) count at diagnosis, therapy information [induction therapy (collected in macro categories: ‘chemotherapy’ – ‘hypomethylating agents’ – ‘best supportive therapy’ for the high grade of variability, overall and in each institution), response to induction, consolidation courses], Hematopoietic Stem Cell Transplant (HSCT), death or last follow-up date, cytogenetic data at diagnosis, molecular data (mutations of *TP53*, *FLT3*, *NPM1*, *IDH1*, *IDH2*, *DNMT3A*, *CEBPA*, *RUNX1*, *CBL*, *nRAS*, *WT1* expression) at diagnosis.

Logistic regression models were applied to identify putative risk factors. Survival analyses were carried out with Kaplan-Meier method¹³ and significance was assessed by Log Rank test¹⁴. Regression models were built based on Cox Hazard Ratio¹⁵. Analyses and graphs were obtained with IBM SPSS Statistics 20®.

References for supplementary methods

- 1 Klampfl T, Milosevic JD, Puda A, Schanegger A, Bagienski K, Berg T *et al.* Complex Patterns of Chromosome 11 Aberrations in Myeloid Malignancies Target CBL, MLL, DDB1 and LMO2. *PLoS One* 2013; **8**: e77819.
- 2 Milosevic JD, Puda A, Malcovati L, Berg T, Hofbauer M, Stukalov A *et al.* Clinical significance of genetic aberrations in secondary acute myeloid leukemia. *Am J Hematol* 2012; **87**: 1010–6.
- 3 Parkin B, Erba H, Ouillet P, Roulston D, Purkayastha A, Karp J *et al.* Acquired genomic copy number aberrations and survival in adult acute myelogenous leukemia. *Blood* 2010; **116**: 4958–4967.
- 4 R Core Team. R: A language and environment for statistical computing. 2016.
- 5 Marc Carlson. org.Hs.eg.db: Genome wide annotation for Human. 2016.
- 6 Willem Ligtenberg. reactome.db: A set of annotation maps for reactome. 2016.
- 7 Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi A J Integr Biol* 2012; **16**: 284–287.
- 8 Yu G, He Q-Y, Patel H, Brookes S, Chandler H, Palermo R *et al.* ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol BioSyst* 2016; **12**: 477–479.
- 9 Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR. *J R Stat Soc Ser B* 1995; **57**: 289–300.
- 10 Huret J-L, Ahmad M, Arsaban M, Bernheim A, Cigna J, Desangles F *et al.* Atlas of Genetics and Cytogenetics in Oncology and Haematology in 2013. *Nucleic Acids Res* 2013; **41**: D920–D924.
- 11 Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* 2016; **44**: D481-7.
- 12 *ISCN 2016 : an international system for human cytogenomic nomenclature*. KARGER, 2016.
- 13 Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* 1958; **53**: 457–481.
- 14 Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother reports* 1966; **50**: 163–70.
- 15 D. R. Cox. Regression Models and Life-Tables. *J R Stat Soc* 1972; **34**: 187–220.

Supplementary tables

Table S1: Missing value for every parameter considered

	Valid	Missing
chromothripsis	395	0
sex	395	0
pathology	395	0
secondary	372	23
de novo	372	23
WBC > 100.000/mm³	152	243
WBC > 30.000/mm³	152	243
karyotype	352	43
ELN risk	352	43
induction_therapy	308	87
mylotarg in induction	274	121
induction courses	113	282
response to induction	289	106
allogenic HSCT	283	112
TP53 loss	395	0
MAPD	395	0
TP53 mutation status	324	71
FLT3 mutation status	298	97
NPM1 mutation status	286	109
IDH1 mutation status	121	274
IDH2 mutation status	135	260
DNMT3A mutation status	38	357
CEBPA mutation status	106	289
RUNX1 mutation status	87	308
CBL mutation status	91	304
NRAS mutation status	95	300

Table S2: Differences in baseline characteristics and therapy in the population enriched for *TP53* alteration (loss and/or mutation). (Only patients with available data are included)

	Number of patients with chromothripsis	Number of patients without chromothripsis	χ^2 p
HSCT-received	14/18	23/30	ns
Treated with anthracycline based chemotherapy at diagnosis	10/19	23/38	ns
Response to induction	3/10	7/21	ns
De novo AML	15/20	26/39	ns
WBC at diagnosis (>30.000/mm ³)	0/12	10/14	.067
<i>NPM1</i> mutation	0/20	1/34	ns
<i>FLT3</i> mutation	0/20	3/34	ns

Table S3: best 1% scoring REACTOME pathways enriched per amplifications in one or more copy in chromothripsis patients compared with non-chromothripsis patients (sorted by score)

Pathway Super Category	Pathway Name	Altered genes/genes in the pathway	Q-VAL
Metabolism	Glycosaminoglycan metabolism	109/116	7.98E-06
Cell Cycle	E2F mediated regulation of DNA replication	26/30	2.75E-05
Disease	Constitutive Signaling by Aberrant PI3K in Cancer	58/59	2.75E-05
DNA Repair	DNA Repair	138/141	2.75E-05
Hemostasis	Platelet activation, signaling and aggregation	182/190	2.75E-05
Hemostasis	Cell surface interactions at the vascular wall	98/101	2.75E-05
Hemostasis	Tie2 Signaling	18/18	2.75E-05
Immune System	Signaling by Interleukins	99/107	2.75E-05
Immune System	Interleukin-2 signaling	38/41	2.75E-05
Immune System	Signaling by the B Cell Receptor (BCR)	184/190	2.75E-05
Immune System	Fc epsilon receptor (FCERI) signaling	162/171	2.75E-05
Immune System	CD209 (DC-SIGN) signaling	20/21	2.75E-05
Metabolism	Glycerophospholipid biosynthesis	87/90	2.75E-05
Metabolism	Phospholipid metabolism	133/139	2.75E-05
Signal Transduction	SOS-mediated signalling	14/14	2.75E-05
Signal Transduction	Signaling by EGFR	164/168	2.75E-05
Signal Transduction	Signalling to ERKs	37/37	2.75E-05
Signal Transduction	Signaling by FGFR	148/151	2.75E-05
Signal Transduction	Signaling by ERBB4	141/143	2.75E-05
Signal Transduction	Signaling by Leptin	21/21	2.75E-05
Signal Transduction	Signaling by FGFR1	148/151	2.75E-05
Signal Transduction	Signaling by FGFR2	148/151	2.75E-05
Signal Transduction	Signaling by FGFR3	148/151	2.75E-05
Signal Transduction	Signaling by FGFR4	148/151	2.75E-05
Signal Transduction	RHO GTPases Activate Formins	94/102	2.75E-05
Transmembrane transport of small molecules	Ion channel transport	159/177	2.75E-05

Table S4: best 1% scoring REACTOME pathways enriched per heterozygous and homozygous deletions in chromothripsis patients compared with non chromothripsis patients (sorted by score)

Pathway Super Category	Pathway Name	Altered genes/genes in the pathway	Q-VAL
Immune System	CTLA4 inhibitory signaling	9/11	6.95E-13
Immune System	CLEC7A (Dectin-1) induces NFAT activation	10/12	6.95E-13
Metabolism	Synthesis of PIPs at the late endosome membrane	9/10	6.95E-13
Fanconi Anemia pathway	Fanconi Anemia pathway	19/24	1.30E-12
Metabolism	alpha-linolenic (omega3) and linoleic (omega6) acid metabolism	8/13	1.30E-12
Metabolism	alpha-linolenic acid (ALA) metabolism	8/13	1.30E-12
Metabolism of proteins	Calnexin/calreticulin cycle	10/11	1.42E-12
Cell Cycle	G0 and Early G1	19/21	1.47E-12
Disease	Diseases of metabolism	30/35	1.47E-12
Extracellular matrix organization	Laminin interactions	24/30	1.47E-12
Immune System	Growth hormone receptor signaling	21/23	1.47E-12
Metabolism	Heme biosynthesis	07/11	1.47E-12
Metabolism	Synthesis of bile acids and bile salts via 24-hydroxycholesterol	09/10	1.47E-12
Metabolism of proteins	Synthesis of glycosylphosphatidylinositol (GPI)	13/17	1.47E-12
Metabolism of proteins	N-glycan trimming in the ER and Calnexin/Calreticulin cycle	12/13	1.47E-12
Signal Transduction	Pre-NOTCH Transcription and Translation	14/19	1.47E-12
Signal Transduction	The canonical retinoid cycle in rods (twilight vision)	13/16	1.47E-12
Transmembrane transport of small molecules	Metal ion SLC transporters	21/25	1.47E-12
Transmembrane transport of small molecules	Ion transport by P-type ATPases	24/39	1.47E-12

Supplementary Figures

Figure S1: Clinical and biological characteristics in patients with and without chromothripsis

Panel (A) WBC at diagnosis in patients with (1) and without (0) chromothripsis, $p = .040$; panel (B) *FLT3* mutational status in patients with (1) and without (0) chromothripsis: 1 = wild-type *FLT3*, 2 = *FLT3* ITD mutation; 3 = *FLT3* TKD mutation; panel (C) age at diagnosis in patients with (1) and without (0) chromothripsis; panel (D) ELN risk in patients with (1) and without (0) chromothripsis: 1 = LR, 2 = INT-1; 3 = INT-2; 4 = HR.

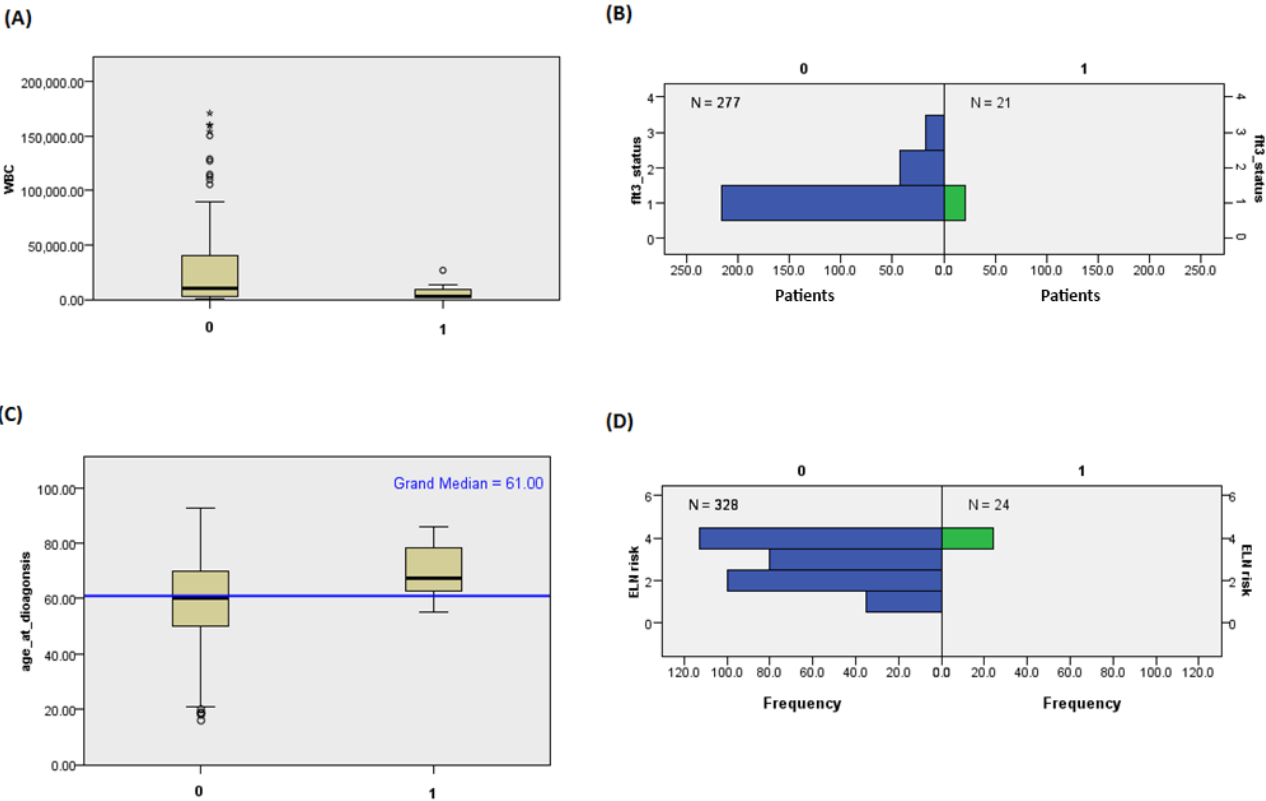


Figure S2: *TP53* altered status and chromothripsis impact on survival and clinical data

OS in patients with (green line) and without chromothripsis (blue line): panel (A) population enriched for *TP53* alteration, $p=ns$; panel (B) population enriched for *TP53* loss, $p=.049$; panel (C) population enriched for *TP53* mutation, $p=ns$. Panel (D) distribution of WBC at diagnosis in the population enriched for *TP53* alteration with (green bars) or without (blue bars) chromothripsis, median test $p=ns$; panel (E) distribution of age at diagnosis in the population enriched for *TP53* alteration with (green bars) or without (blue bars) chromothripsis, median test $p=ns$.

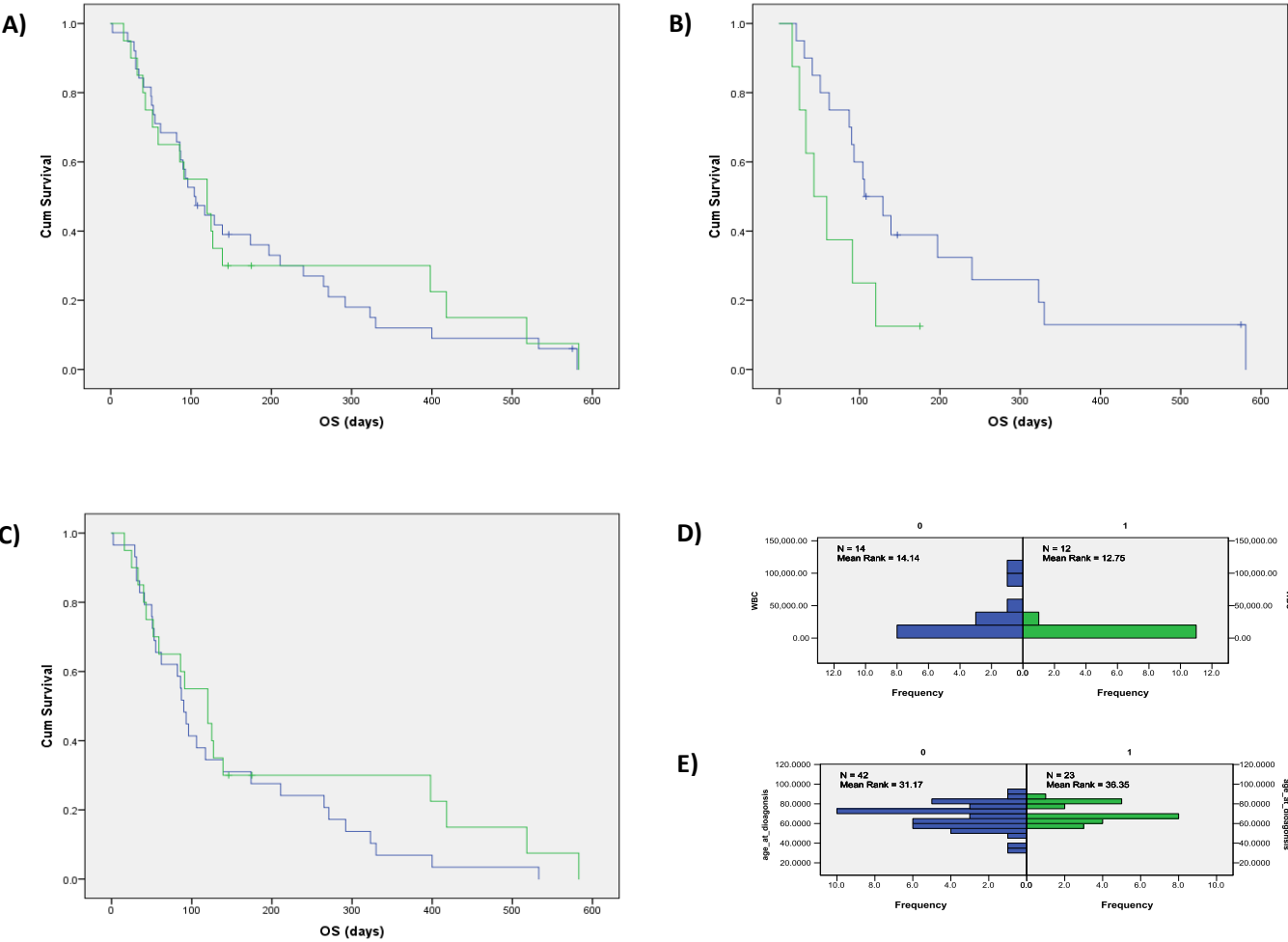


Figure S3: Representation of all chromosomes affected by chromothripsis in our cohort of adult AML patients.

Samples' IDs are reported on the top of each image representing a chromosome affected by chromothripsis: in some patients chromothripsis occurred in 2 different chromosomes. The figure is plotted with R package "Rawcopy" using R 3.3.2.

