Supplemental Information

Genomic Heterogeneity in Core-Binding Factor Acute Myeloid Leukemia and its Clinical Implication

Nikolaus Jahn,[1] Tobias Terzer,[2] Eric Sträng,[3] Anna Dolnik,[3] Sibylle Cocciardi,[1] Ekaterina Panina,[1] Andrea Corbacioglu,[1] Julia Herzig,[1] Daniela Weber,[1] Anika Schrade,[1] Katharina Götze,[4] Thomas Schröder,[5] Michael Lübbert,[6] D. Wellnitz,[7] Elisabeth Koller,[8] Richard F. Schlenk,[9] Verena I. Gaidzik,[1] Peter Paschka,[1] Frank G. Rücker,[1] Michael Heuser,[10] Felicitas Thol,[10] Arnold Ganser,[10] Axel Benner,[2] Hartmut Döhner,[1] Lars Bullinger,[3,*] Konstanze Döhner[1,*]

*LB and KD shared senior authorship

[1]Department of Internal Medicine III, University Hospital of Ulm, Ulm, Germany; [2]Division of Biostatistics, German Cancer Research Center Heidelberg, Heidelberg, Germany; [3]Department of Hematology, Oncology, Tumor Immunology, Charité University Medicine, Berlin, Germany; [4]Department of Internal Medicine III, University Hospital Klinikum Rechts der Isar, Munich, Germany; [5]Department of Hematology, Oncology, and Clinical Immunology, University of Duesseldorf, Medical Faculty, Duesseldorf, Germany; [6]Department of Internal Medicine I, University Hospital of Freiburg, Freiburg, Germany; [7]Department of Internal Medicine II, University Hospital of Schleswig-Holstein, Kiel, Germany; [8]Department of Internal Medicine III, Hanuschkrankenhaus Wien, Wien, Austria; [9]National Center of Tumor Diseases (NCT), German Cancer Research Center, Heidelberg, Germany; [10]Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany

TABLE OF CONTENTS

SUPPLEMENTAL METHODS

Library Enrichment and Sequencing

The library enrichment was performed using the commercially available SureSelectXT in-solution capture technology from Agilent Technologies (Santa Clara, CA, USA). The library comprised the entire coding region of 230 candidate genes involved in hematological malignancies. For library design Agilent's online tool SureDesign was used. Only exons with a consensus annotation in the RefSeq, Ensembl, CCDS, Gencode, and SNP databases were considered relevant. The UCSC human genome 19 (H. sapiens, hg19, GRCh37, February 2009) served as reference genome for design and determination of genomic coordinates. All probes were tiled with 50% overlap. To enhance uniformity of the capture and thus evenness in coverage the number of copies of 'orphan' and GC-rich baits was increased as the corresponding target regions are methodically underrepresented. No probes were tiled to repetitive and low complexity DNA regions identified and masked by RepeatMasker and NCBI Window Masker plus/minus Duke Uniqueness Track to avoid unspecific probe annealing. Using this approach 34,256 probes with a total size of 1.359 Mbp were generated, covering approximately 96% of the regions of interest. Only very few regions were theoretically missed but practically covered by contiguous probes without any dropout. Assay design and subsequent bioinformatic analysis thus allowed for sensitive detection of single nucleotide variants (SNV) and small insertions/deletions (indels) whereas structural aberrations (copy number alterations and chromosomal rearrangements) as well as large indels were not captured.

Genomic DNA (200ng per sample) extracted from diagnostic bone marrow (81%) or peripheral blood (19%) specimens was used for molecular screening. For samples in which not enough DNA was available, genomic DNA was amplified using REPLI-g Mini Kit (Qiagen, Hilden, Germany). SureSelect library preparation and indexing were performed following the manufacturer's instructions for Illumina paired-end sequencing. Samples were then transferred to a cBot (Illumina, San Diego, CA, USA) to create clonal clusters on a flow cell by bridge amplification (Illumina reagent kit: TruSeq PE

Cluster Kit v3-cBot-HS). Finally, 2x 100 bp paired-end sequencing by synthesis was carried out on a HiSeq2000 (Illumina, San Diego, CA, USA) using Illumina's TruSeq SBS Kit v3-HS reagents.

Variant calling

The sequencing quality of each sample was assessed using the NGS QC toolkit (2.3.3) and, where necessary, adapter and read end trimming were performed using cutadapt (1.8.3) and in-house scripting respectively.

Paired-end reads were then aligned to the hg19 reference using BWA-MEM (0.7.10). Alignments are sorted and indexed by Picard (1 .138) and locally realigned using GATK (3.4.46). For each sample, coverage statistics were calculated using BEDTools (2.24.0) and processed by SAMtools (0.1.19). VarScan2 (2.3.9) was then used for variant calling within the target regions sequenced. All variants were annotated by Annovar (release 22Mar2015) but only non-synonymous mutations affecting exons or splice sites were retained. These were further filtered to remove calls within known regions of segmental duplication, variants annotated in dbSNP (138) but not COSMIC (70) and variants with a minor allele frequency (MAF) above 0.01 in either the 1000 Genomes Project or the Exome Sequencing Project (ESP 6500).

Curation of oncogenic variants

All calls yielded by the computational annotation workflow were subject to further curation. Only variants considered oncogenic were included in the subsequent analyses. The algorithm for mutation reporting is as follows:

a) Removal of all variants that are annotated in SNP databases and occur with a minor allele frequency (MAF) >0.001 in the 1000 Genomes Project, dbSNP150 or the Exome Sequencing Project (ESP 6500).

b) Removal of variants present within regions prone to sequence context specific artifacts, including regions of high depth, enriched for reads of low mapping quality that harbor multiple mismatches

c) Removal of all one bp insertions or deletions present adjacent to regions of more than 5 homopolymer bases (for example insG adjacent to GGGGG) and a variant allele frequency of ≤0.1

d) Removal of all missense variants with a variant allele frequency between ≥0.45 and ≤0.55 or ≥0.9 and 1.0, indicative of polymorphisms, unless they are present with ≥5 counts in COSMIC database (v85) and with ≥1 confirmed somatic.

e) Retention of all frameshift, nonsense or splicing variants with a variant allele frequency ≥0.05

f) Retention of all missense variants with a variant allele frequency between 0.05 and <0.45 or >0.55 and <0.9, indicative of (likely) oncogenic variants.

SUPPLEMENTAL TABLES

Supplemental Table S1: List of all genes targeted by custom sequencing panel

ABCA12	*CHEK2*	*EP300*	*IDH2*	*NEK2*	*PTPN11*	*SMC1A*
ABL1	*CLTCL1*	*EPHA6*	*IKZF2*	*NF1*	*PTPRF*	*SMC3*
ACIN1	*CNNM2*	*ETNK1*	*INPP5D*	*NF2*	*PTPRT*	*SMG1*
ACSS3	*COPRS*	*ETV6*	*IRF1*	*NFE2*	*PXDN*	*SPI1*
ADGRV1	*CREBBP*	*EVI2A*	*IRF4*	*NFE2L1*	*RAB11FIP4*	*SPRED2*
ALK	*CSF1R*	*EVI2B*	*IRF8*	*NFE2L2*	*RAC1*	*SRCAP*
ARHGEF10	*CSF2RB*	*EWSR1*	*JAK1*	*NIPBL*	*RAD21*	*SRSF2*
ARID1A	*CSF3R*	*EZH1*	*JAK2*	*NOTCH1*	*RAD50*	*STAG1*
ARID2	*CSMD1*	*EZH2*	*JAK3*	*NOTCH2*	*RAD51*	*STAG2*
AS3MT	*CSMD2*	*FAM175A*	*JARID2*	*NPM1*	*RASA2*	*STAT3*
ASXL1	*CSNK1A1*	*FAM5C*	*KAT6A*	*NRAS*	*RASA3*	*STAT5A*
ASXL2	*CTCF*	*FAT4*	*KDM5C*	*NRXN1*	*RASEF*	*SUZ12*
ATRX	*CTNNB1*	*FBXW7*	*KDM6A*	*NRXN3*	*RASGRF1*	*SYNE1*
BAP1	*CUX1*	*FGFR2*	*KIF27*	*NSD1*	*RB1*	*TERT*
BCL10	*DCC*	*FLG*	*KIT*	*NT5C2*	*RBBP5*	*TET1*
BCL2	*DDX23*	*FLT3*	*KMT2A*	*NUMA1*	*RBMX*	*TET2*
BCOR	*DDX4*	*FOXP1*	*KMT2D*	*NUP98*	*RHOA*	*TP53*
BCORL1	*DDX41*	*FRMD3*	*KRAS*	*NXF1*	*RMI1*	*TTC39A*
BCR	*DDX54*	*GALNT11*	*LAMA1*	*OBSCN*	*ROBO1*	*U2AF1*
BRAF	*DHX15*	*GALNTL5*	*LAMC3*	*OMG*	*ROBO2*	*U2AF2*
BRCC3	*DHX33*	*GATA1*	*LUC7L2*	*PAX5*	*RPS6KA6*	*UBQLN1*
C6	*DICER1*	*GATA2*	*MAP3K4*	*PDGFB*	*RRAS*	*UBXN11*
C9orf103	*DIS3*	*GFI1*	*MGA*	*PHF6*	*RUNX1*	*WAC*
CALR	*DNAH9*	*GKAP1*	*MLL3*	*PHIP*	*RYR2*	*WHSC1*
CBL	*DNAJB8*	*GNAS*	*MLL5*	*PIK3CA*	*SAMHD1*	*WT1*
CCND2	*DND1*	*GNB1*	*MN1*	*PLEKHH1*	*SETBP1*	*YLPM1*
CDHR1	*DNM2*	*H3F3A*	*MPL*	*PLEKHS1*	*SETD2*	*ZBTB33*
CDK4	*DNMT3A*	*H3F3B*	*MYC*	*PPM1D*	*SETDB1*	*ZBTB7A*
CDKN1B	*DNMT3B*	*HCN1*	*MYH9*	*PRKAG2*	*SF1*	*ZMYM3*
CDKN2A	*DYNC1H1*	*HIPK2*	*MYLK2*	*PRPF40A*	*SF3A1*	*ZNF318*
CDKN2B	*EED*	*HNRNPK*	*MYO1F*	*PRPF40B*	*SF3B1*	*ZNF687*
CDKN2C	*EEFSEC*	*HRAS*	*NCOA7*	*PRPF8*	*SH2B3*	*ZRSR2*
CEBPA	*EGFR*	*IDH1*	*NDE1*	*PTEN*	*SMARCB1*	

Supplemental Table S2 Incidence and comparison of recurrent mutations in each CBF cohort

	t(8;21) n=190 patients		**inv(16)** n=160 patients		**Total cohort** n=350 patients		
	n	**%**	**n**	**%**	**n**	**%**	***p*-value**
NRAS	31	16	62	39	93	27	**< 0.0001**
KIT	48	25	42	26	90	26	0.9024
FLT3	25	13	33	21	58	17	0.0827
***FLT3*-ITD**	9	5	4	3	13	4	0.0597
ZBTB7A	36	19	1	1	37	11	**< 0.0001**
ASXL1	27	14	2	1	29	8	**< 0.0001**
KRAS	3	2	26	16	29	8	**< 0.0001**
ASXL2	27	14	0	0	27	8	**< 0.0001**
TET2	20	11	5	3	25	7	**0.011**
WT1	8	4	16	10	24	7	**0.0358**
RAD21	20	11	0	0	20	6	**< 0.0001**
CCND2	19	10	1	1	20	6	**< 0.0001**
KDM6A	16	8	3	2	19	5	**0.0081**
SMC1A	17	9	0	0	17	5	**< 0.0001**
SRCAP	8	4	8	5	16	5	0.8001
EZH2	12	6	2	1	14	4	**0.0251**
SYNE1	10	5	4	3	14	4	0.2742
MGA	10	5	3	2	13	4	0.154
NF1	3	2	10	6	13	4	**0.0248**
DNMT3A	11	6	1	1	12	3	**0.0076**
DNM2	11	6	0	0	11	3	**0.0012**
DHX15	10	5	1	1	11	3	**0.0136**
CSF3R	9	5	2	1	11	3	0.0719
CBL	9	5	1	1	10	3	**0.0242**
OBSCN	4	2	6	4	10	3	0.522
DNAH9	4	2	6	4	10	3	0.522
BCORL1	1	1	9	6	10	3	**0.0066**
ZNF318	4	2	5	3	9	3	0.7372
SMC3	8	4	0	0	8	2	**0.0088**
BRCC3	8	4	0	0	8	2	**0.0088**
JAK2	7	4	1	1	8	2	0.0752
BCOR	4	2	4	3	8	2	1
STAG2	6	3	1	1	7	2	0.1309
RYR2	5	3	2	1	7	2	0.4606
LAMA1	3	2	4	3	7	2	0.7068
JAK3	6	3	0	0	6	2	0.0333
SETD2	5	3	1	1	6	2	0.2251
CREBBP	5	3	1	1	6	2	0.2251
MPL	4	2	2	1	6	2	0.6918
KMT2D	2	1	4	3	6	2	0.4182
FOXP1	0	0	6	4	6	2	**0.0087**
PHIP	4	2	1	1	5	1	0.3807
PHF6	3	2	2	1	5	1	1
CSMD1	3	2	2	1	5	1	1

KMT2C	2	1	3	2	5	1	0.6635
CSMD2	2	1	3	2	5	1	0.6635
FAT4	2	1	3	2	5	1	0.6635
ALK	2	1	3	2	5	1	0.6635
KDM5C	4	2	0	0	4	1	0.1284
EP300	3	2	1	1	4	1	0.6284
MYO1F	3	2	1	1	4	1	0.6284
SH2B3	0	0	4	3	4	1	**0.0428**
NIPBL	3	2	0	0	3	1	0.2534
GNAS	3	2	0	0	3	1	0.2534
EGFR	3	2	0	0	3	1	0.2534
CLTCL1	3	2	0	0	3	1	0.2534
FBXW7	3	2	0	0	3	1	0.2534
IDH2	3	2	0	0	3	1	0.2534
IDH1	3	2	0	0	3	1	0.2534
ARHGEF10	2	1	1	1	3	1	1
MYH9	2	1	1	1	3	1	1
TP53	2	1	1	1	3	1	1
GPR98	2	1	1	1	3	1	1
RUNX1	2	1	1	1	3	1	1
MYC	2	1	1	1	3	1	1
NRXN3	1	1	2	1	3	1	0.5945
ROBO2	1	1	2	1	3	1	0.5945
HIPK2	1	1	2	1	3	1	0.5945
ETV6	1	1	2	1	3	1	0.5945
NOTCH1	1	1	2	1	3	1	0.5945
CUX1	1	1	2	1	3	1	0.5945
ROBO1	1	1	2	1	3	1	0.5945
PTPN11	0	0	3	2	3	1	0.0946
CEBPA	0	0	3	2	3	1	0.0946
NRXN1	0	0	3	2	3	1	0.0946
PXDN	2	1	0	0	2	1	0.5023
JAK1	2	1	0	0	2	1	0.5023
SUZ12	2	1	0	0	2	1	0.5023
ATRX	2	1	0	0	2	1	0.5023
GATA2	2	1	0	0	2	1	0.5023
BRINP3	2	1	0	0	2	1	0.5023
NUMA1	2	1	0	0	2	1	0.5023
DCC	2	1	0	0	2	1	0.5023
HCN1	1	1	1	1	2	1	1
EPHA6	1	1	1	1	2	1	1
PRPF8	1	1	1	1	2	1	1
PRPF40A	1	1	1	1	2	1	1
NT5C2	1	1	1	1	2	1	1
RPS6KA6	1	1	1	1	2	1	1
KMT2E	1	1	1	1	2	1	1
FGFR2	1	1	1	1	2	1	1
OMG	1	1	1	1	2	1	1
SAMHD1	1	1	1	1	2	1	1

NXF1	1	1	1	1	2	1	1
ACSS3	1	1	1	1	2	1	1
DDX41	1	1	1	1	2	1	1
C6	1	1	1	1	2	1	1
ACIN1	1	1	1	1	2	1	1
PRKAG2	0	0	2	1	2	1	0.2083
CDHR1	0	0	2	1	2	1	0.2083
PPM1D	0	0	2	1	2	1	0.2083
ABCA12	0	0	2	1	2	1	0.2083
SETDB1	0	0	2	1	2	1	0.2083
EEFSEC	0	0	2	1	2	1	0.2083

Supplemental Table S3 Incidence and comparison of recurrent additional chromosomal aberrations in each CBF cohort; abbreviation: NA, not assessed

	t(8;21) n=190 patients		**inv(16)** n=160 patients		**Total cohort** n=350 patients		
	n	**%**	**n**	**%**	**n**	**%**	***p*-value**
del(9q)	26	14	2	1	28	9	**< 0.0001**
del(7q)	5	3	17	12	22	7	**0.0016**
+8	16	9	14	10	30	9	0.8486
+22	0	0	25	17	25	8	**< 0.0001**
-X	31	17	0	0	31	9	**< 0.0001**
-Y	54	30	1	1	55	17	**< 0.0001**
NA	8	-	14	-	22	-	-

Supplemental Table S4 Association with RTK/RAS signaling. Signaling group: *NRAS*, *KIT*, *FLT3*, *KRAS*, *NF1*, *DNM2*, *CSF3R*, *CBL*, *JAK2*, *JAK3*, *PTPN11*.

	RTK.RAS 0 (n=99)	**RTK.RAS 1** (n=234)	**pval**
Age			0.49024 (wilcox.)
Median	50.52	48.72	
Range	18.2 to 79.76	18.05 to 75.68	
NA	0	0	
Hemoglobin			0.37039 (wilcox.)
Median	9.3	9	
Range	4.2 to 15.5	3.5 to 17	
NA	6	7	
Platelets			0.99681 (wilcox.)
Median	33	35	
Range	6 to 314	2 to 382	
NA	6	7	
Wbc			< 0.0001 (wilcox.)
Median	5.11	18.7	
Range	0.7 to 138.85	1.15 to 222.7	
NA	6	7	
Bone Marrow Blasts			0.16987 (wilcox.)

	RTK.RAS 0 (n=99)	**RTK.RAS 1** (n=234)	**pval**
Median	60	68	
Range	5 to 99	13 to 100	
NA	9	32	
Peripheral Blood Blasts			0.00948 (wilcox.)
Median	29	38	
Range	0 to 86	0 to 95	
NA	10	19	
CBF.Type			< 0.0001 (fisher)
inv(16)	26 (26.26 %)	129 (55.13 %)	
t(8;21)	73 (73.74 %)	105 (44.87 %)	

Supplemental Table S5 Association with chromatin. Chromatin group: *ASXL1*, *ASXL2*, *SRCAP*, *KDM6A*, *EZH2*, *SETD2*, *KMT2D*.

	Chromatin 0 (n=239)	**Chromatin 1** (n=94)	**pval**
Age			0.93147 (wilcox.)
Median	48.9	49.11	
Range	18.05 to 75.68	18.2 to 79.76	
NA	0	0	
Hemoglobin			0.50928 (wilcox.)
Median	9.195	8.8	
Range	3.5 to 17	4.7 to 15.5	
NA	9	4	
Platelets			0.60483 (wilcox.)
Median	36.5	32	
Range	2 to 262	5 to 382	
NA	9	4	
Wbc			0.00566 (wilcox.)
Median	14.135	8.15	
Range	1.2 to 222.7	0.7 to 102	
NA	9	4	
Bone Marrow Blasts			0.35453 (wilcox.)
Median	65	65	
Range	5 to 100	13 to 95	
NA	32	9	
Peripheral Blood Blasts			0.23171 (wilcox.)
Median	36	33	
Range	0 to 92	0 to 95	
NA	20	9	
CBF.Type			< 0.0001 (fisher)
inv(16)	139 (58.16 %)	16 (17.02 %)	
t(8;21)	100 (41.84 %)	78 (82.98 %)	

Supplemental Table S6 Association with methylation. Methylation group: *TET2*, *DNMT3A*, *IDH1*, *IDH2*.

	Methylation 0 (n=297)	**Methylation 1** (n=36)	**pval**
Age			0.04561 (wilcox.)
Median	48.53	53.895	
Range	18.05 to 79.76	20.61 to 71.54	
NA	0	0	
Hemoglobin			0.85913 (wilcox.)
Median	9.145	8.95	
Range	3.5 to 17	4.8 to 13.3	
NA	11	2	
Platelets			0.0633 (wilcox.)
Median	36	27	
Range	2 to 382	7 to 200	
NA	11	2	
Wbc			0.58166 (wilcox.)
Median	12.8	10.05	
Range	0.7 to 222.7	1.2 to 117.5	
NA	11	2	
Bone Marrow Blasts			0.76626 (wilcox.)
Median	65	68	
Range	5 to 100	20 to 99	
NA	38	3	
Peripheral Blood Blasts			0.10563 (wilcox.)
Median	34	47.5	
Range	0 to 95	0 to 89	
NA	23	6	
CBF.Type			0.00014 (fisher)
inv(16)	149 (50.17 %)	6 (16.67 %)	
t(8;21)	148 (49.83 %)	30 (83.33 %)	

Supplemental Table S7 Association with transcription. Transcription group: *ZBTB7A*, *WT1*, *BCOR*, *BCORL1*, *FOXP1*.

	Transcription 0 (n =242)	**Transcription 1** (n = 68)	**pval**
Age			0.633 (wilcox.)
Median	49.5	48.4	
Range	18.1 to 79.8	18.07 to 73.1	
NA	0	0	
Hemoglobin			0.769 (wilcox.)
Median	9.1	9.1	
Range	3.8 to 15	3.5 to 14	
NA	15	7	
Platelets			0.703 (wilcox.)
Median	34.5	32.5	
Range	2 to 535	6 to 247	
NA	15	7	
Wbc			0.075 (wilcox.)
Median	13.1	8.6	

	Transcription 0 (n =242)	**Transcription 1** (n = 68)	**pval**
Range	1.1 to 222.7	1.5 to 209.9	
NA	15	7	
Bone Marrow Blasts			0.450 (wilcox.)
Median	65	67	
Range	10 to 100	5 to 97	
NA	42	9	
Peripheral Blood Blasts			0.803 (wilcox.)
Median	37.5	33.0	
Range	0 to 95	0 to 90	
NA	27	12	
CBF.Type			0.428 (fisher)
inv(16)	123 (47.9 %)	32 (42.7 %)	
t(8;21)	134 (52.1 %)	43 (57.3 %)	

Supplemental Table S8 Association with cohesion. Cohesin group: *RAD21*, *SMC1A*, *SMC3*, *STAG2*, *NIPBL*.

	Cohesin 0 (n = 283)	**Cohesin 1 (n = 50)**	**pval**
Age			0.95616 (wilcox.)
Median	48.9	49.38	
Range	18.05 to 79.76	18.2 to 73.2	
NA	0	0	
Hemoglobin			0.01839 (wilcox.)
Median	9.2	8.8	
Range	3.5 to 17	4.2 to 12.4	
NA	10	3	
Platelets			0.00747 (wilcox.)
Median	37	24	
Range	2 to 382	6 to 212	
NA	10	3	
Wbc			0.00438 (wilcox.)
Median	13.3	7.8	
Range	1.15 to 222.7	0.7 to 59.7	
NA	10	3	
Bone Marrow Blasts			0.96795 (wilcox.)
Median	65	69	
Range	5 to 100	13 to 100	
NA	34	7	
Peripheral Blood Blasts			0.06668 (wilcox.)
Median	33.5	39	
Range	0 to 90	0 to 95	
NA	25	4	
CBF.Type			< 0.0001 (fisher)
inv(16)	154 (54.42 %)	1 (2 %)	
t(8;21)	129 (45.58 %)	49 (98 %)	

Supplemental Table S9 Cox model with endpoint relapse-free survival in inv(16) AML cohort

	Exp(coef)	Lower 95% CI	Upper 95% CI	*p*.-value
Age	1.03	1.00	1.05	.056
Female gender	0.72	0.35	1.49	.374
tAML	2.97	1.08	8.16	**.035**
WBC	1.08	1.01	1.16	**.027**
Platelets	1.03	0.95	1.11	.467
Hemoglobin	0.84	0.68	1.05	.129

Supplemental Table S10 Cox model with endpoint relapse-free survival in t(8;21) AML cohort

	Exp(coef)	Lower 95% CI	Upper 95% CI	*p*.-value
Age	1.02	1.00	1.04	.081
Female gender	1.27	0.72	2.24	.407
tAML	0.25	0.03	1.88	.177
WBC	1.08	0.96	1.21	.213
Platelets	1.00	0.96	1.05	.851
Hemoglobin	0.92	0.80	1.06	.235

Supplemental Table S11 Cox model with endpoint overall survival in inv(16) AML cohort

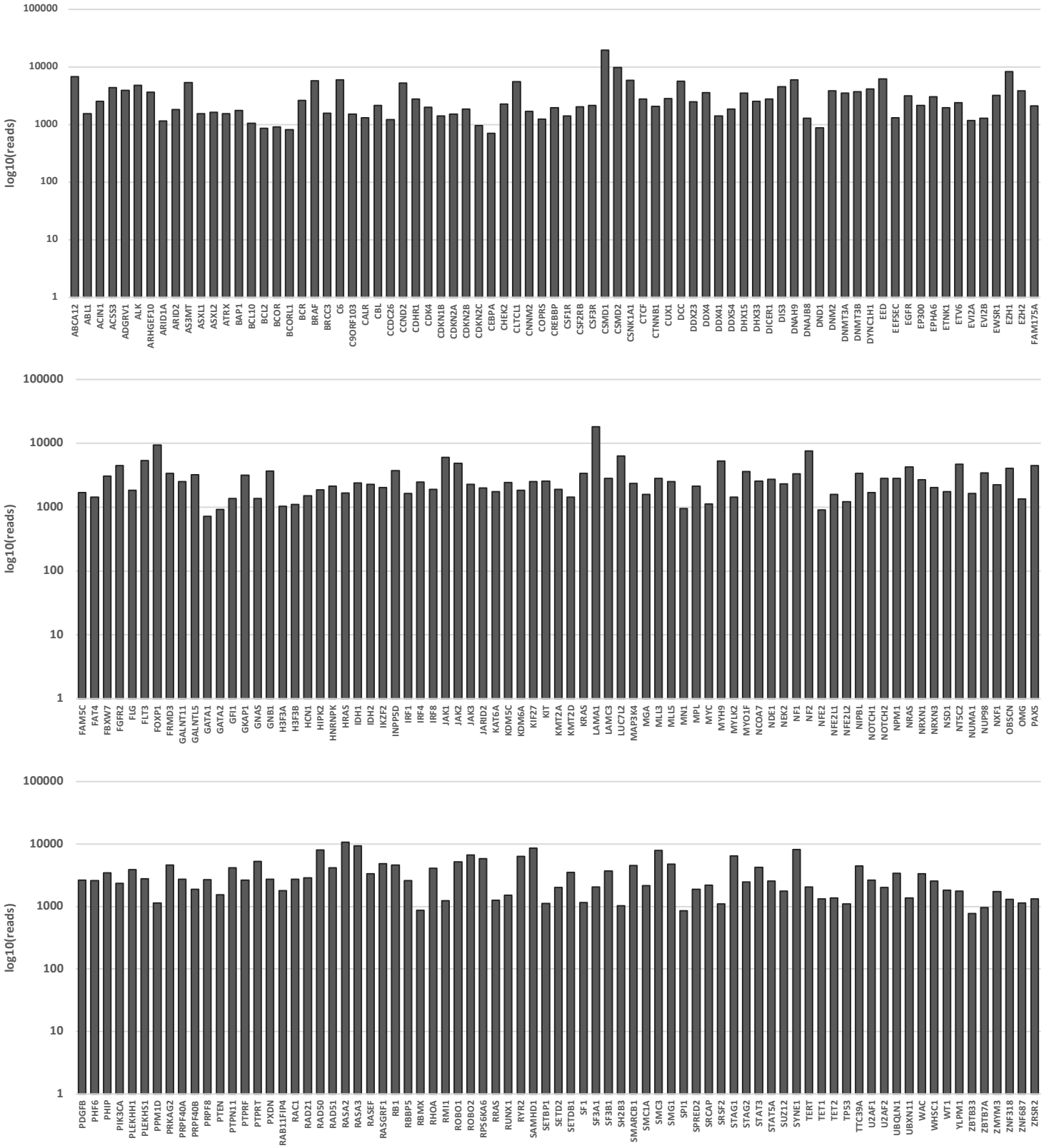
	Exp(coef)	Lower 95% CI	Upper 95% CI	*p*.-value
Age	1.06	1.02	1.09	**.001**
Female gender	0.94	0.40	2.18	.880
tAML	1.72	0.55	5.39	.350
WBC	1.07	0.98	1.18	.145
Platelets	1.03	0.94	1.14	.503
Hemoglobin	0.79	0.61	1.03	.080

Supplemental Table S12 Cox model with endpoint overall survival in t(8;21) AML cohort

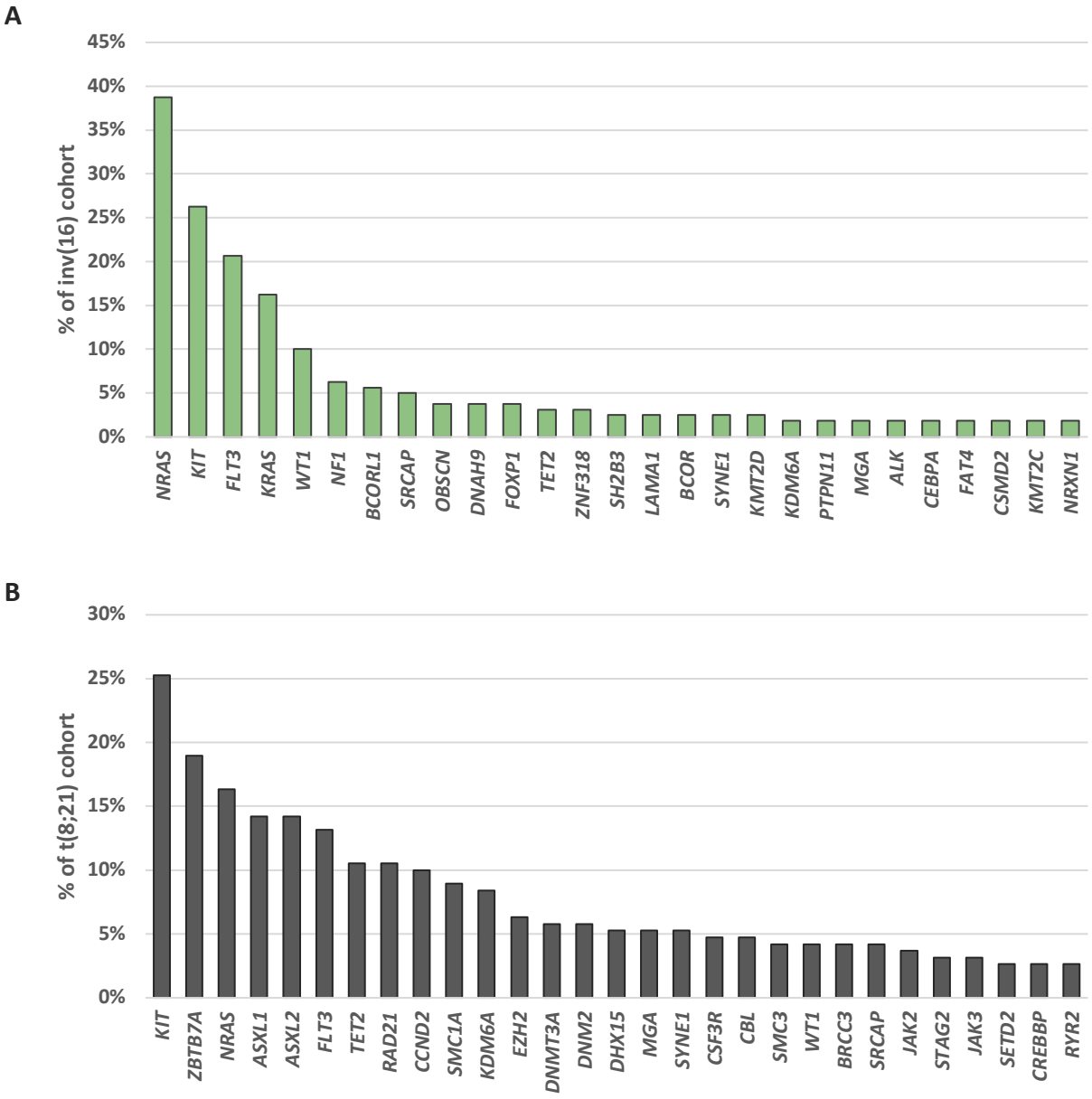
	Exp(coef)	**Lower 95% CI**	**Upper 95% CI**	***p*.-value**
Age	1.03	1.01	1.06	**.009**
Female gender	0.97	0.52	1.78	.914
tAML	0.85	0.24	2.98	.799
WBC	1.06	0.93	1.20	.387
Platelets	1.00	0.95	1.05	.963
Hemoglobin	0.95	0.82	1.10	.497

SUPPLEMENTAL FIGURES

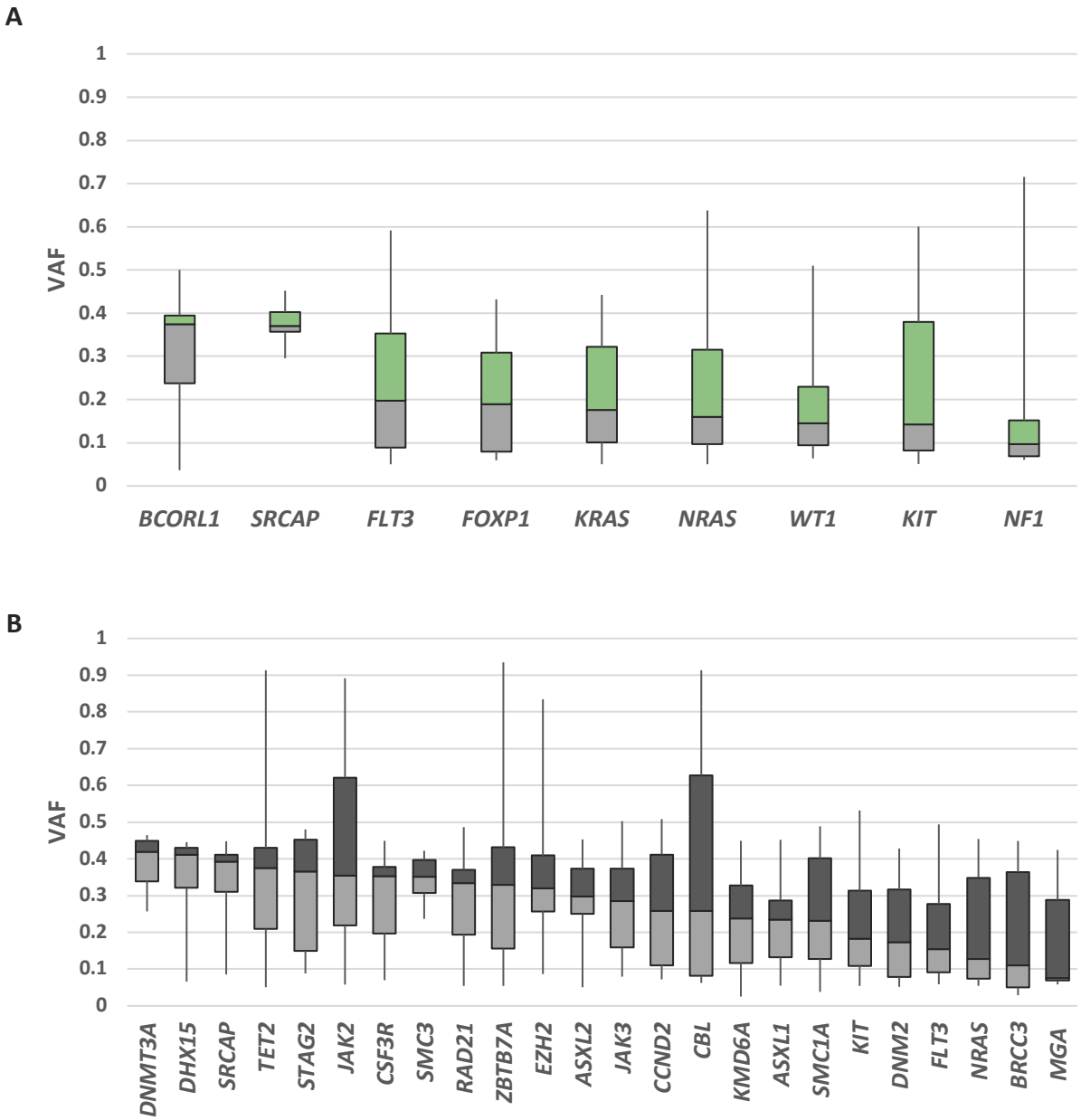
Supplemental Figure S1: Median coverage (sequencing reads) per gene for 230 genes



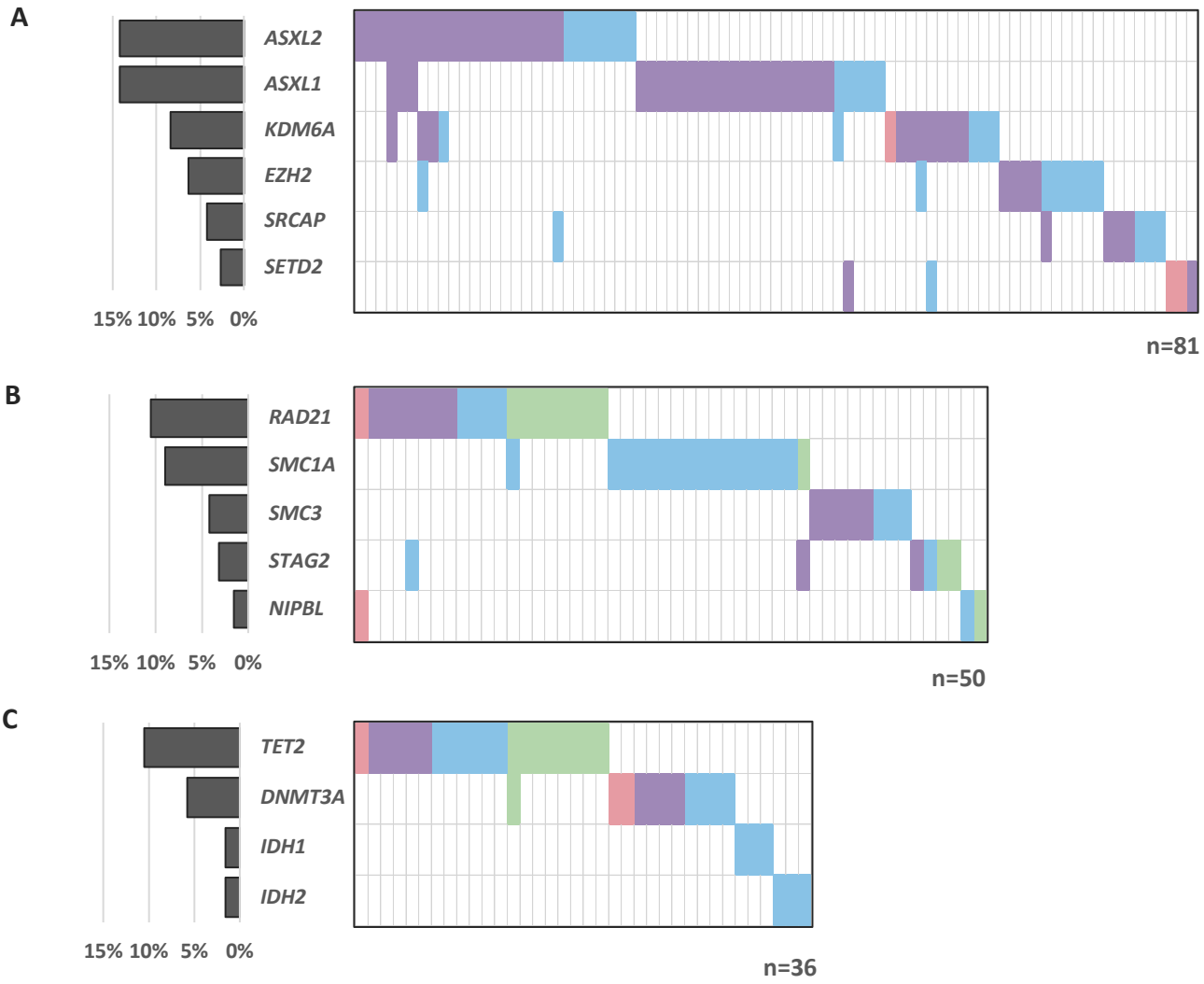
Supplemental Figure S2: Incidence of recurrent mutations in A) inv(16) AML and B) t(8;21) AML cohorts



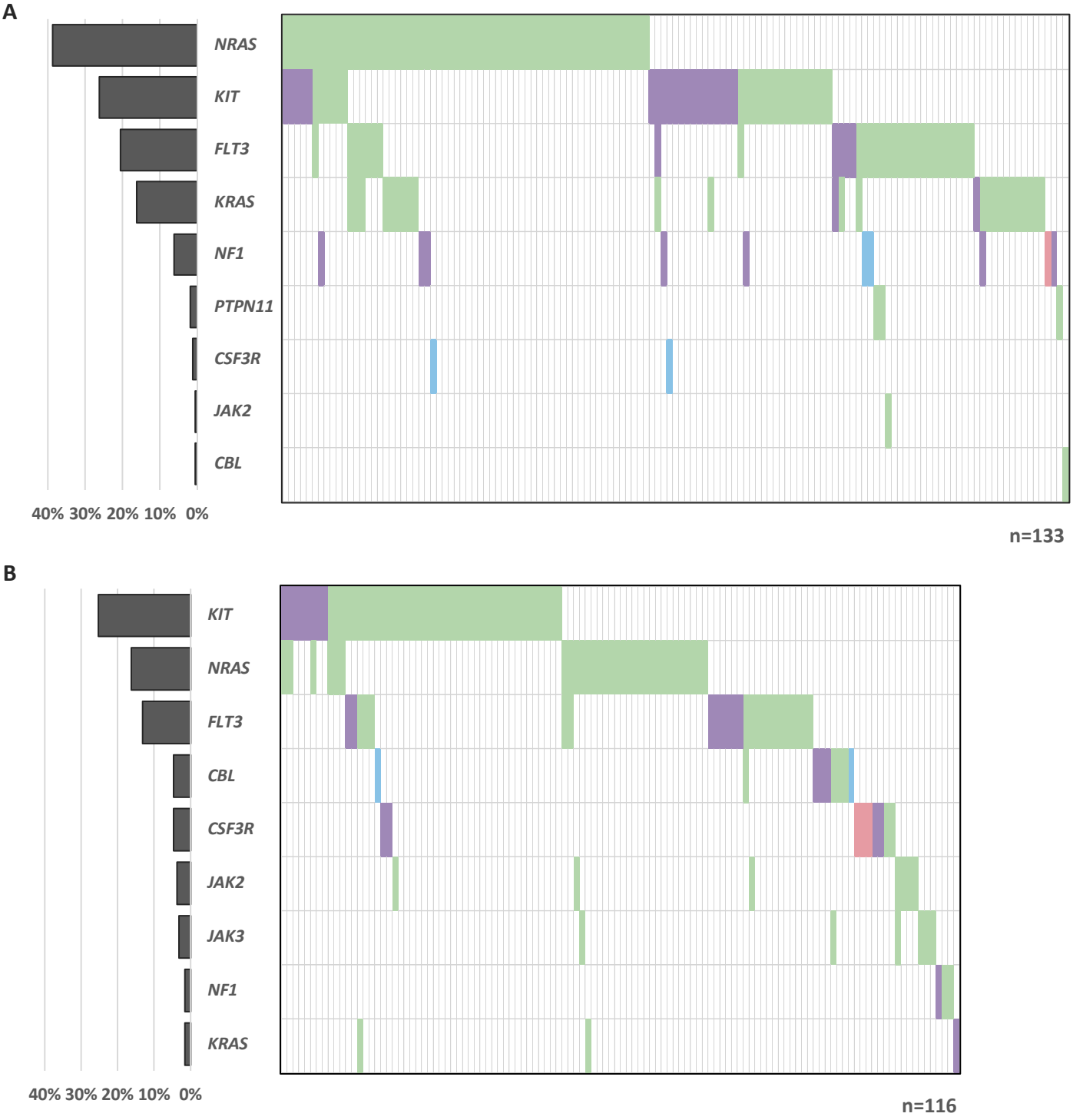
Supplemental Figure S3: Boxplot showing the median, 25%-quantile (gray box), and 75%-quantile (green/black box) of the variant allele fraction (VAF) for recurrently mutated genes in A) inv(16) AML and B) t(8;21) AML cohorts. VAF was corrected for patient specific chromosomal aberrations and sex. For *FLT3* no internal tandem duplication (ITD) mutations were included.



Supplemental Figure S4: Molecular pattern stratified by functional genetic subgroups in t(8;21) AML: A) chromatin modification, B) Cohesin complex and C) methylation. Each box represents a single patient. Wildtype cases are illustrated in white, indels in purple, nonsense mutations in blue, missense mutations in green, and splice site mutations in pink. Gray bars indicate the mutational incidence in the entire t(8;21) AML (n=190) cohort.

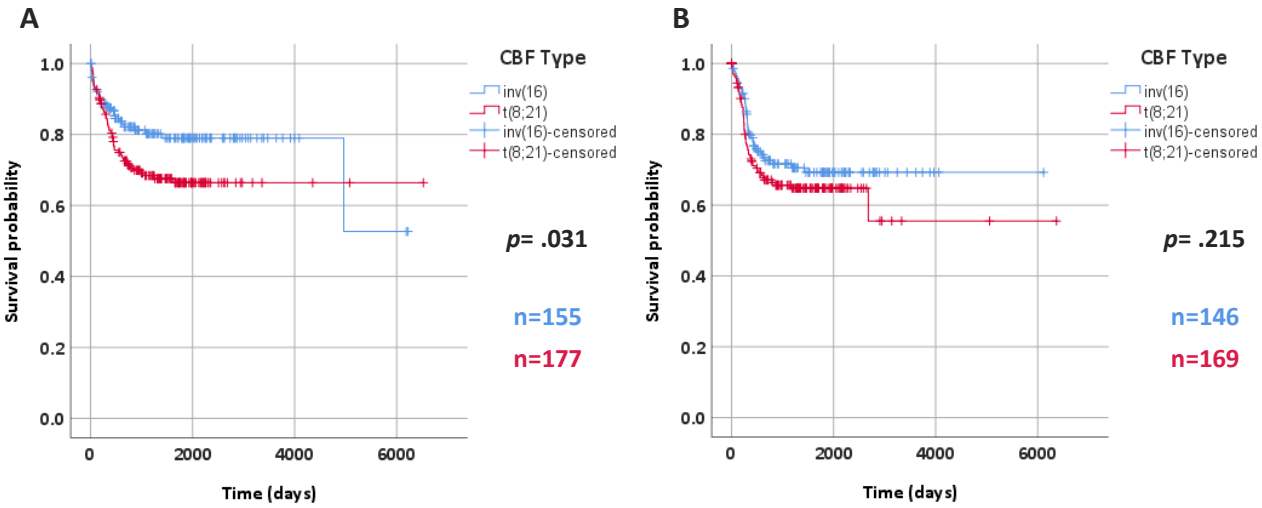


Supplemental Figure S5: Pattern of RTK/RAS-signaling mutations stratified by A) inv(16) AML and B) t(8;21) AML cohorts. Each box represents a single patient. Wildtype cases are illustrated in white, indels in purple, nonsense mutations in blue, missense mutations in green, and splice site mutations in pink. Gray bars indicate the mutational incidence in each cohort.

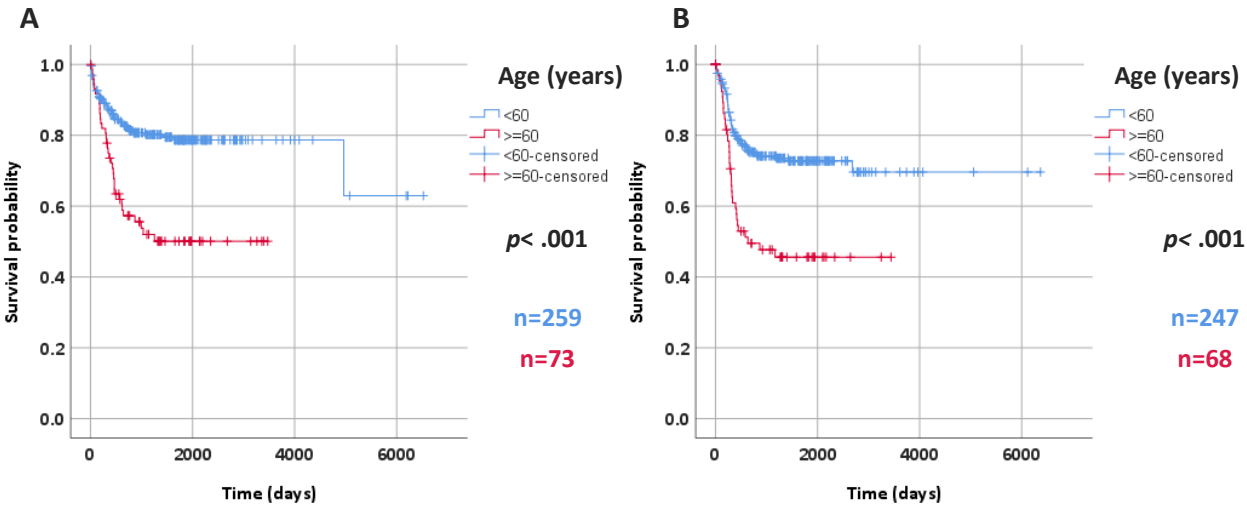


SUPPLEMENTAL RESULTS

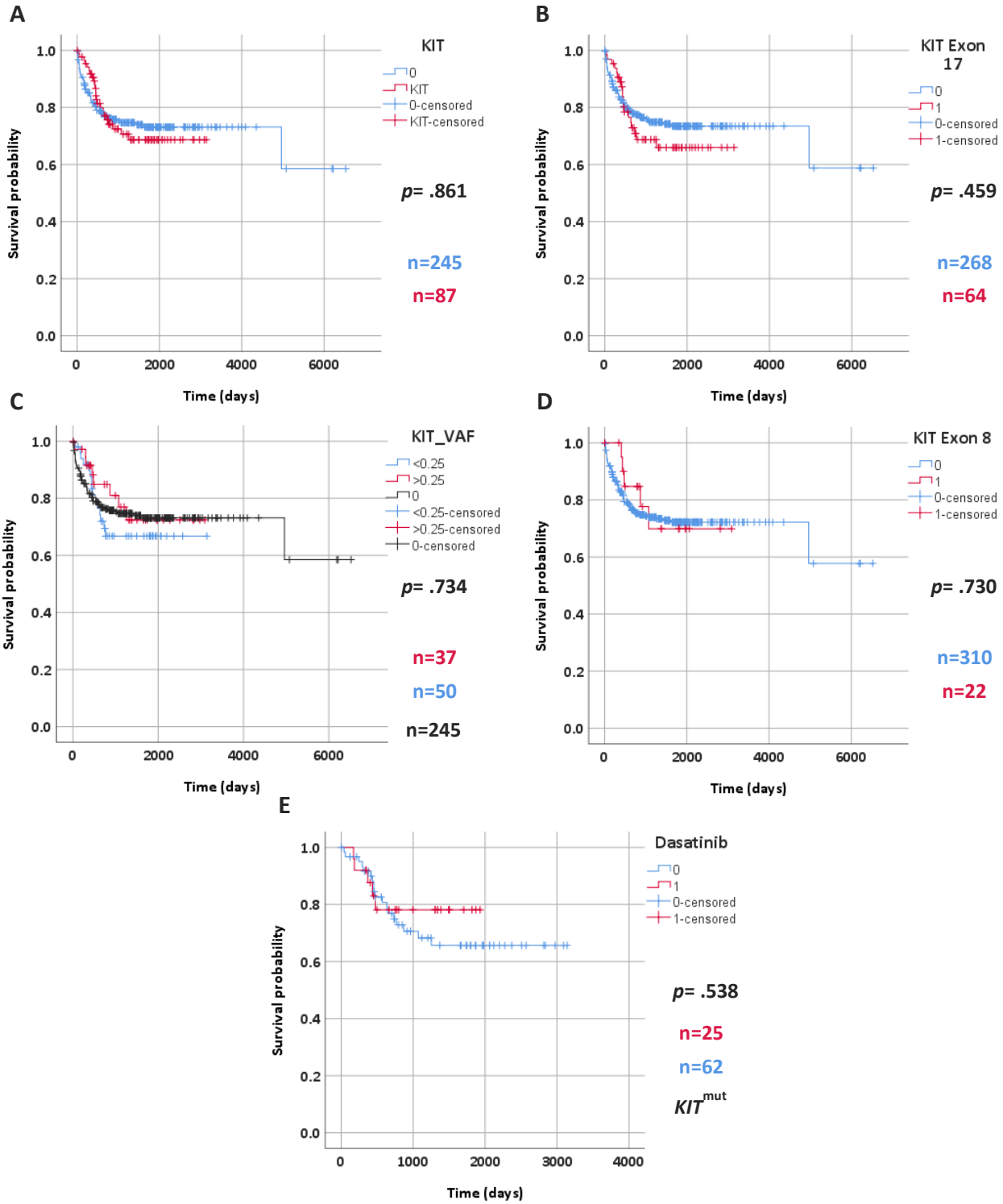
Supplemental Figure S6: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival according to CBF subtype.



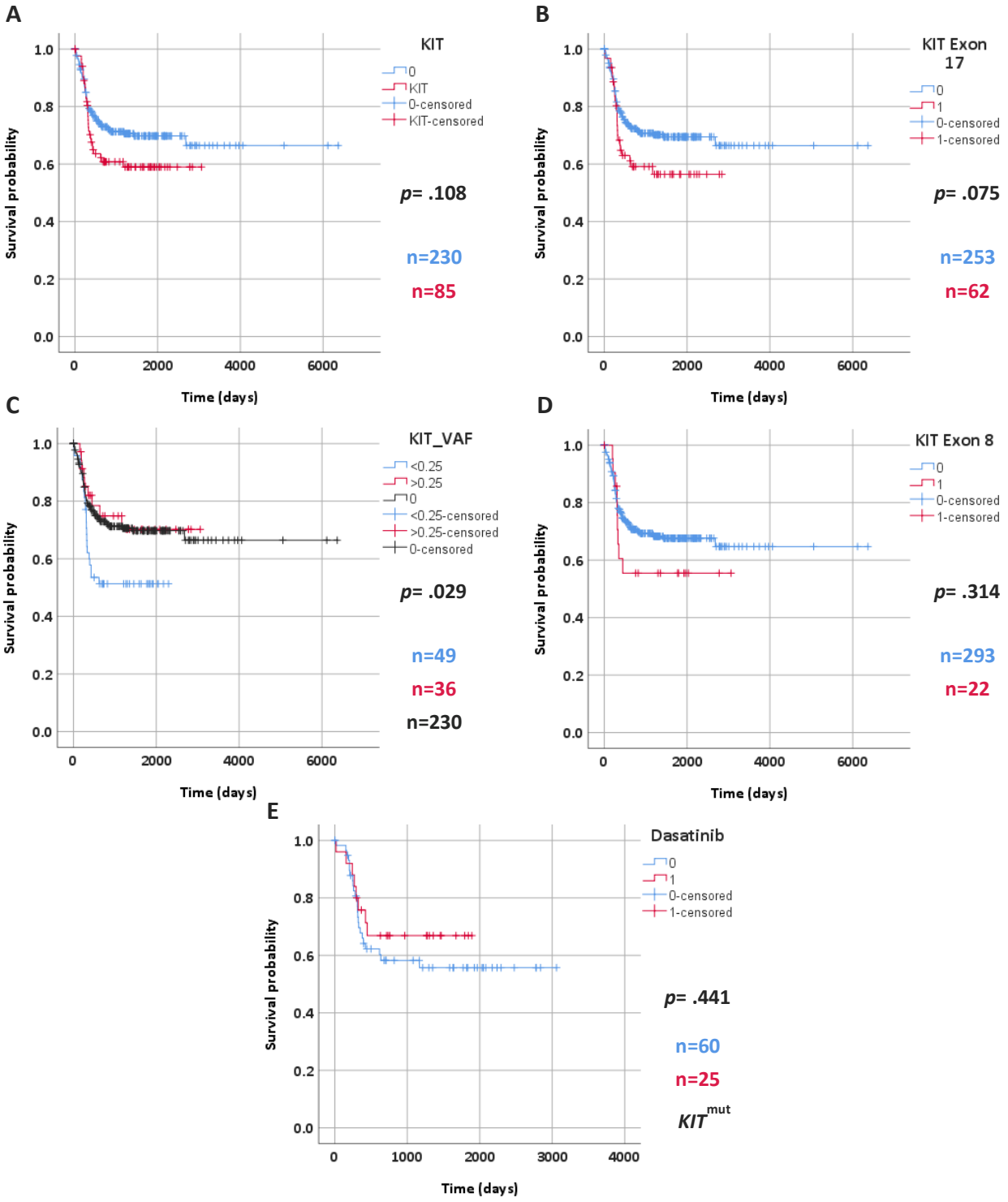
Supplemental Figure S7: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to age.



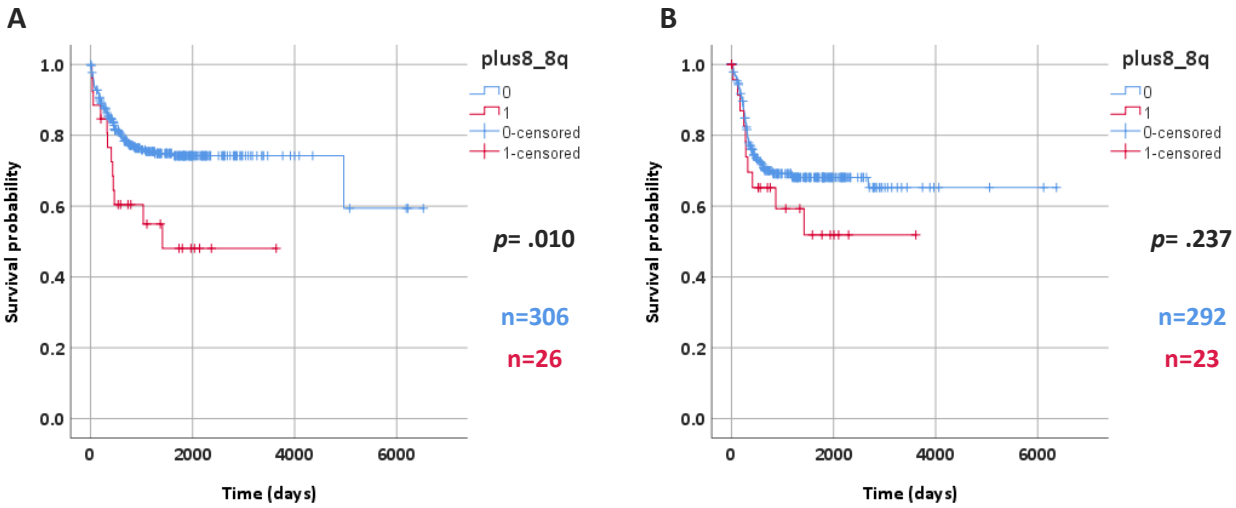
Supplemental Figure S8: Kaplan-Meier estimates for overall survival in CBF cohort according to A) *KIT*, B) *KIT* exon 17, C) *KIT* variant allele fraction (VAF), and D) *KIT* exon 8 mutation status as well as E) in subgroup of *KIT* mutated (*KIT*^{mut}) patients according to treatment with dasatinib.



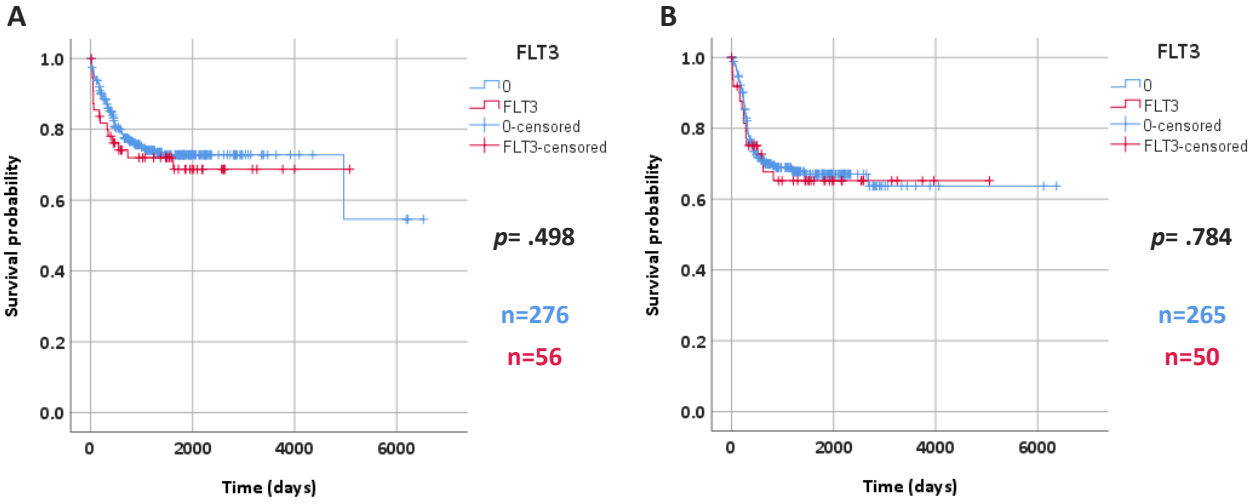
Supplemental Figure S9: Kaplan-Meier estimates for relapse-free survival in CBF cohort according to A) *KIT*, B) *KIT* exon 17, C) *KIT* variant allele fraction (VAF), and D) *KIT* exon 8 mutation status as well as E) in subgroup of *KIT* mutated (*KIT*^{mut}) patients according to treatment with dasatinib.



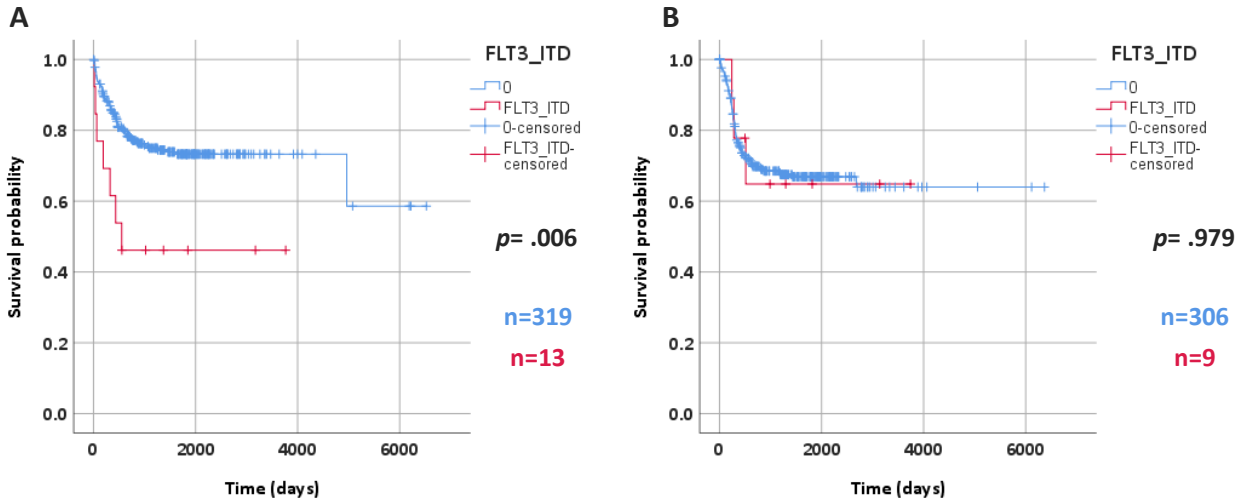
Supplemental Figure S10: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to trisomy 8 presence.



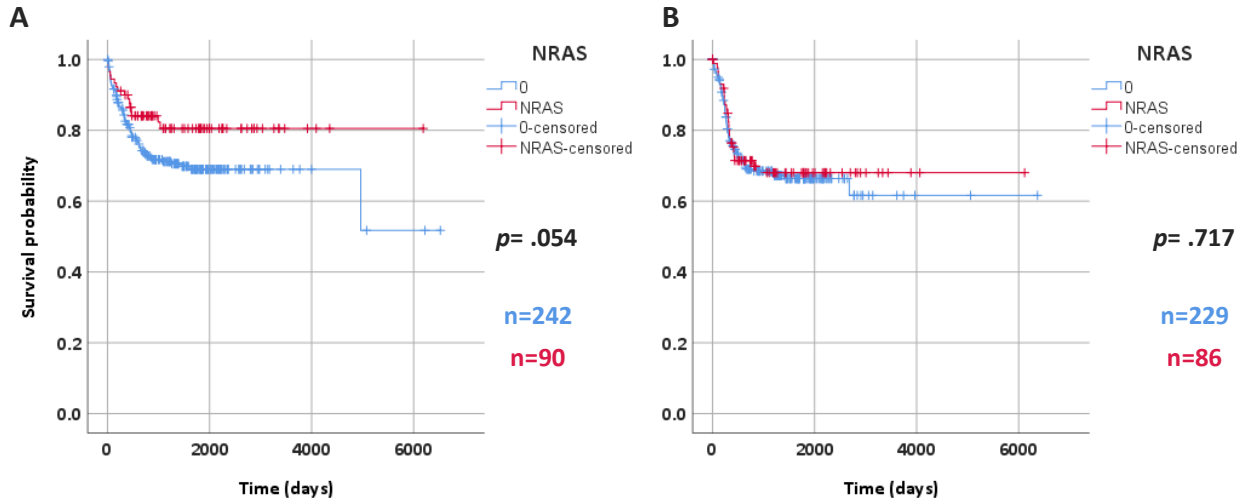
Supplemental Figure S11: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to *FLT3* mutational status. All *FLT3* mutations were included.



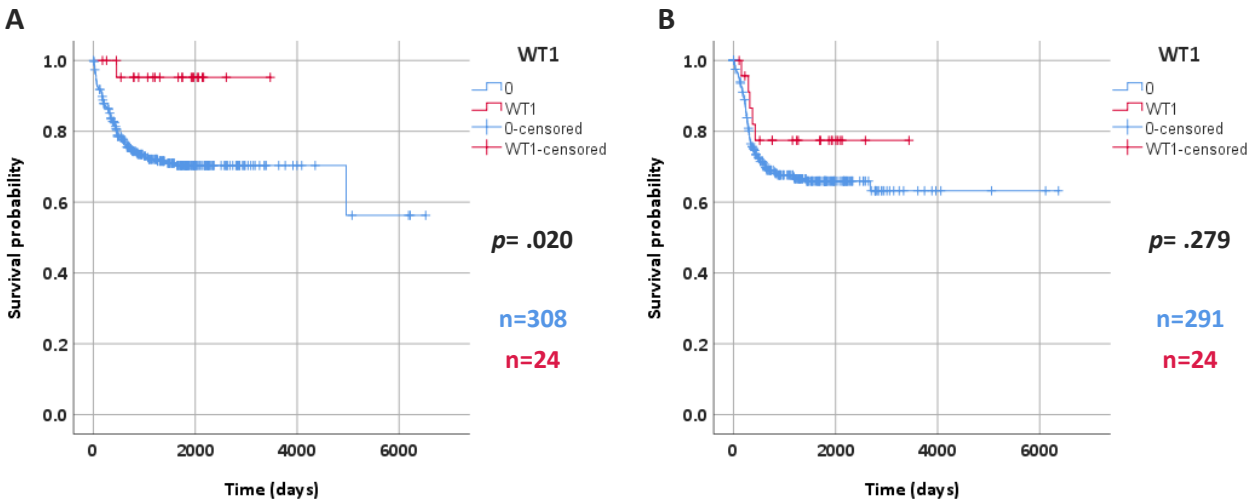
Supplemental Figure S12: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to *FLT3*-ITD status.



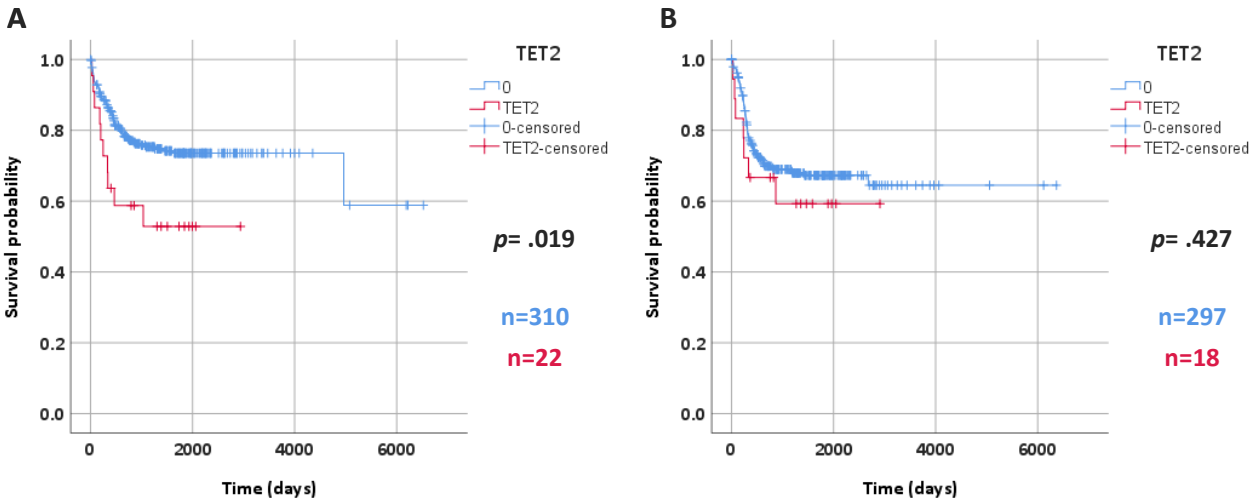
Supplemental Figure S13: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to *NRAS* mutational status.



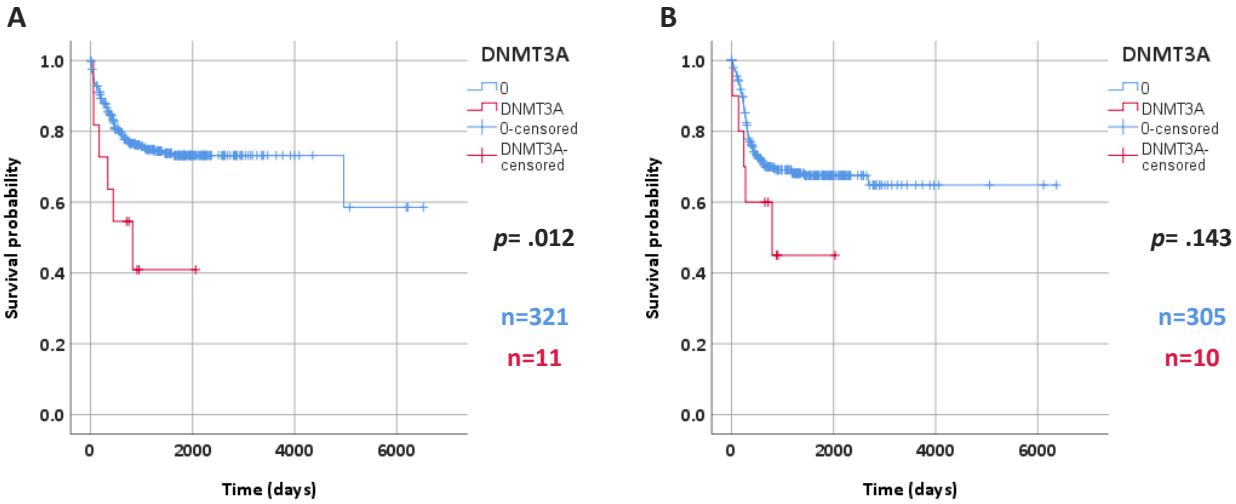
Supplemental Figure S14: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to *WT1* mutational status.



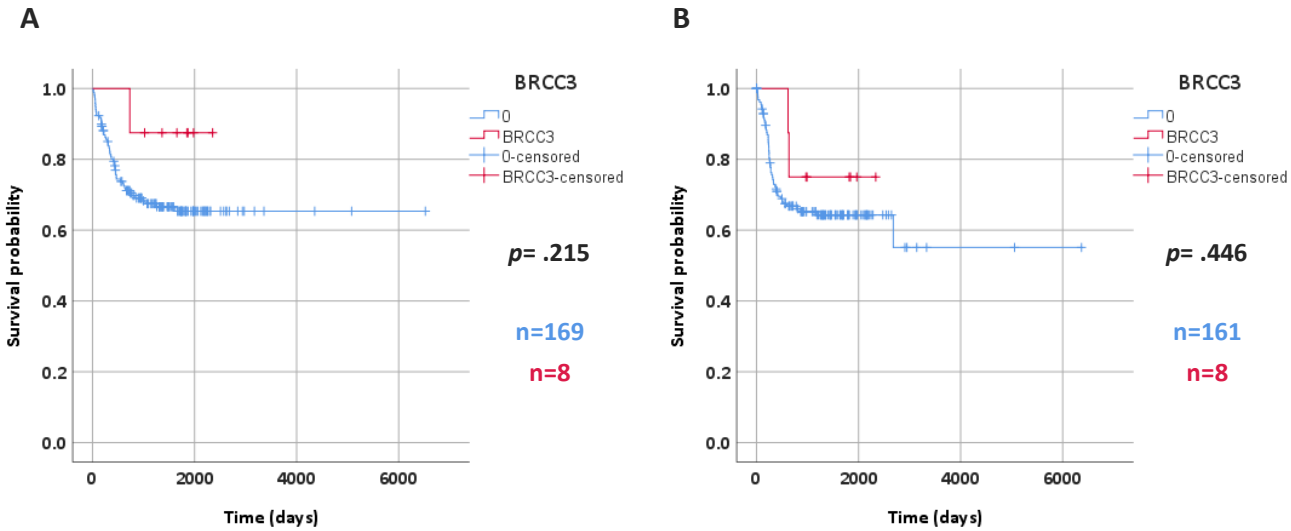
Supplemental Figure S15: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to *TET2* mutational status.



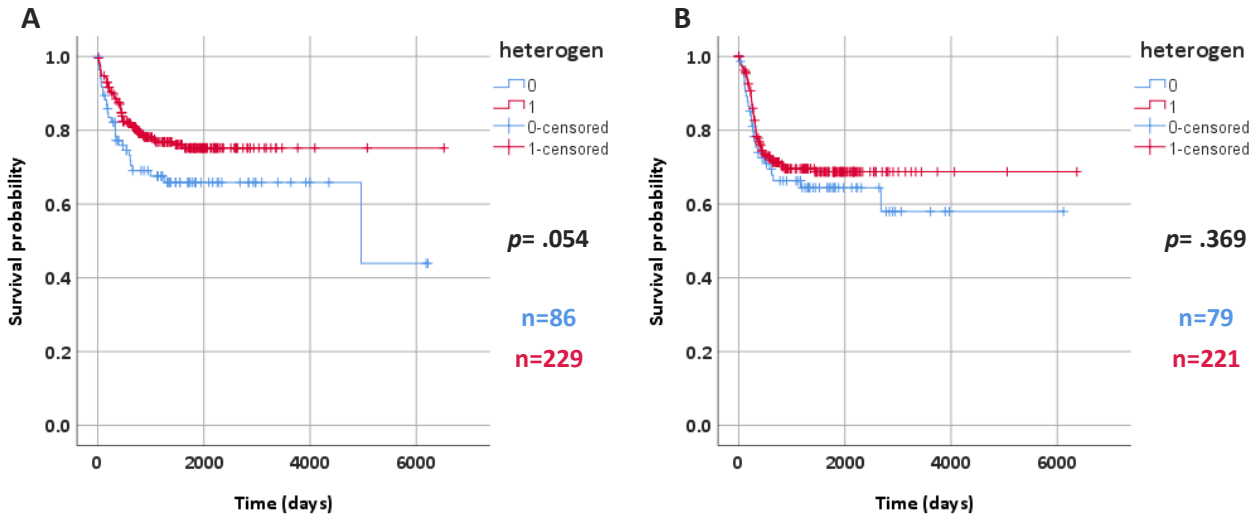
Supplemental Figure S16: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to *DNMT3A* mutational status.



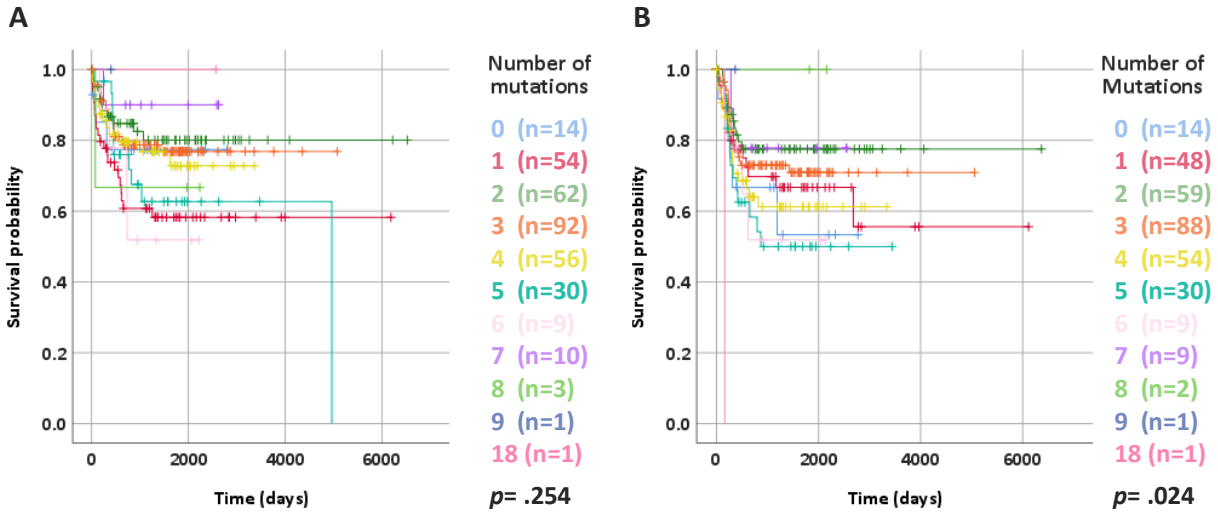
Supplemental Figure S17: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in t(8;21) AML cohort according to *BRCC3* mutations.



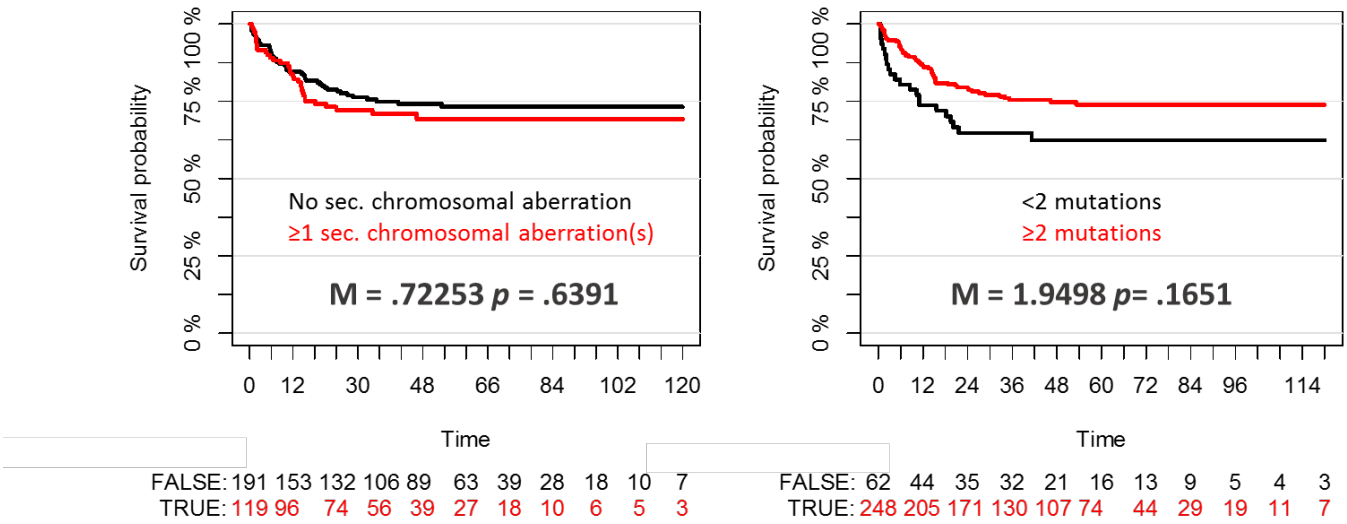
Supplemental Figure S18: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to heterogeneity status. 'Heterogeneous' (red line) is defined as presence of ≥ 2 mutations with significantly different variant allele fractions (VAF) by Pearson Goodness-of-fit test reflecting distinct co-occurring leukemic clones; 'homogenous' (light blue line) as multiple mutations with multiple mutations with similar variant allele fractions or with only a single mutation.

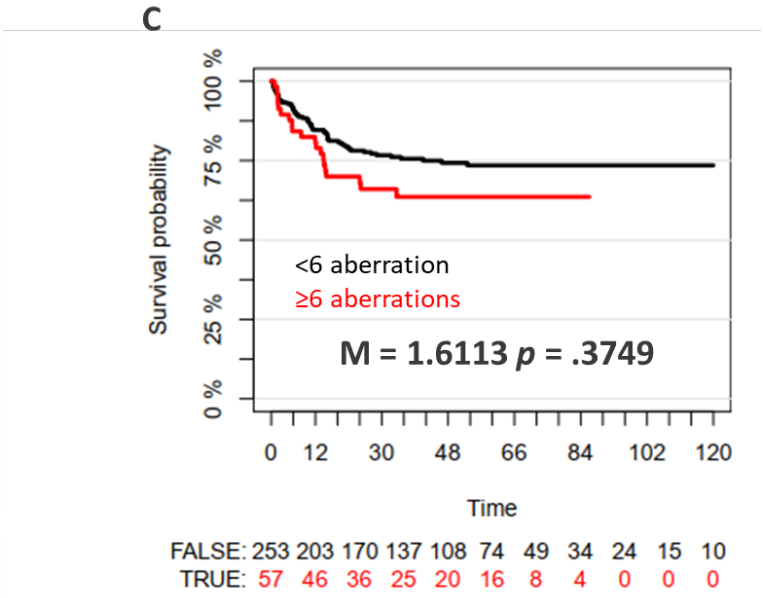


Supplemental Figure S19: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to the number of mutations.

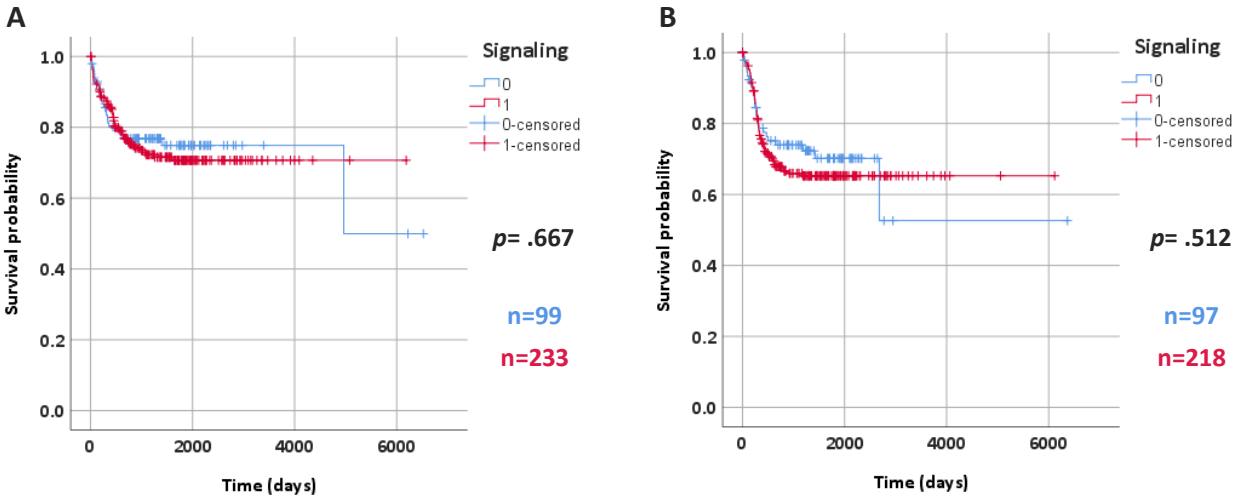


Supplemental Figure S20: Kaplan-Meier estimates for overall survival according to the number of secondary chromosomal aberrations (A), number of mutations (B), and number of total aberrations (C) of total CBF-AML cohort.

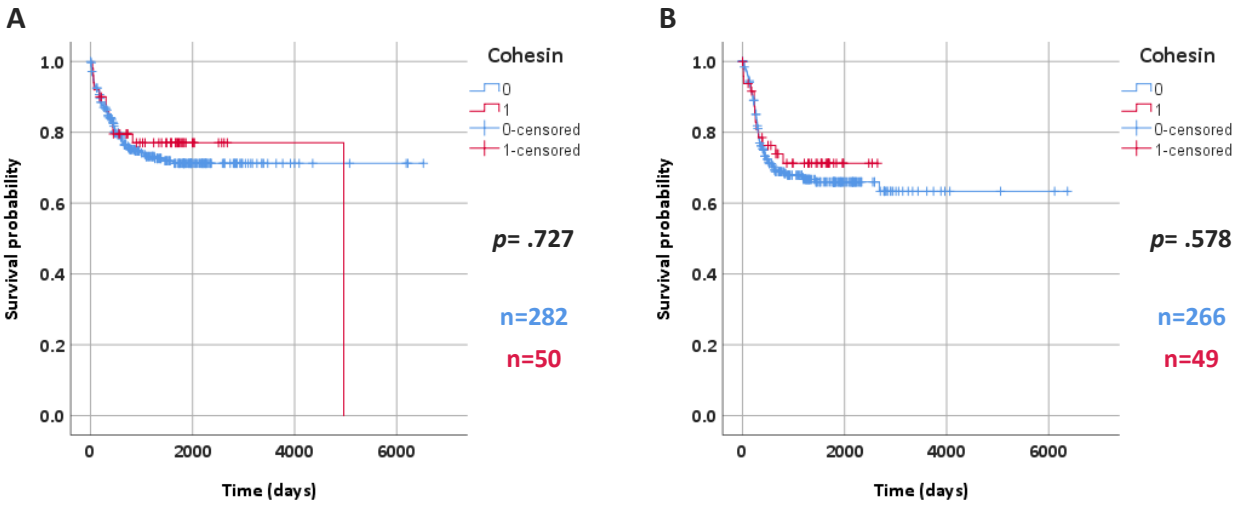




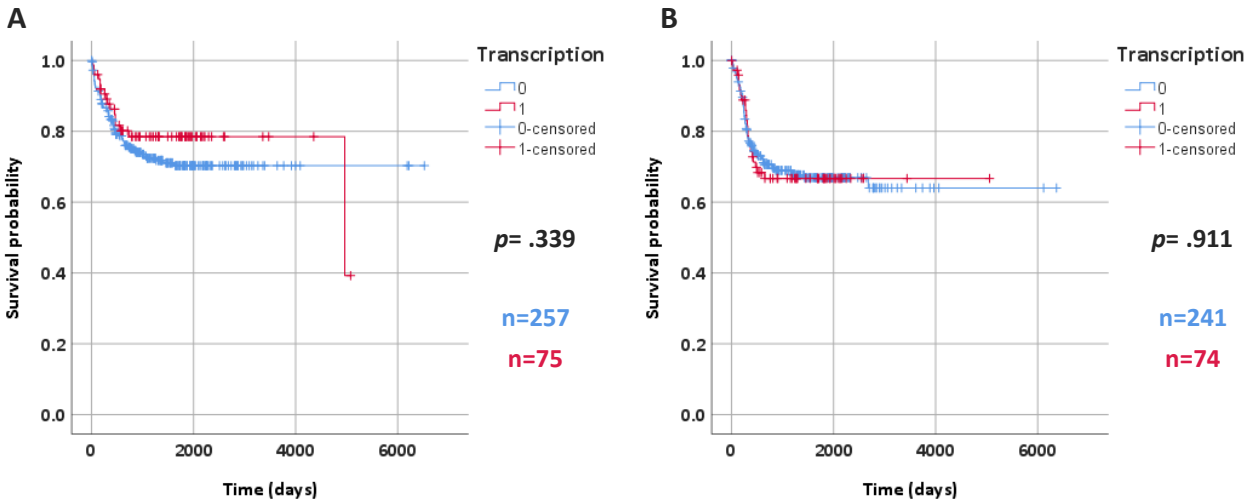
Supplemental Figure S21: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to signaling group mutations. Signaling group: *NRAS*, *KIT*, *FLT3*, *KRAS*, *NF1*, *DNM2*, *CSF3R*, *CBL*, *JAK2*, *JAK3*, *PTPN11*.



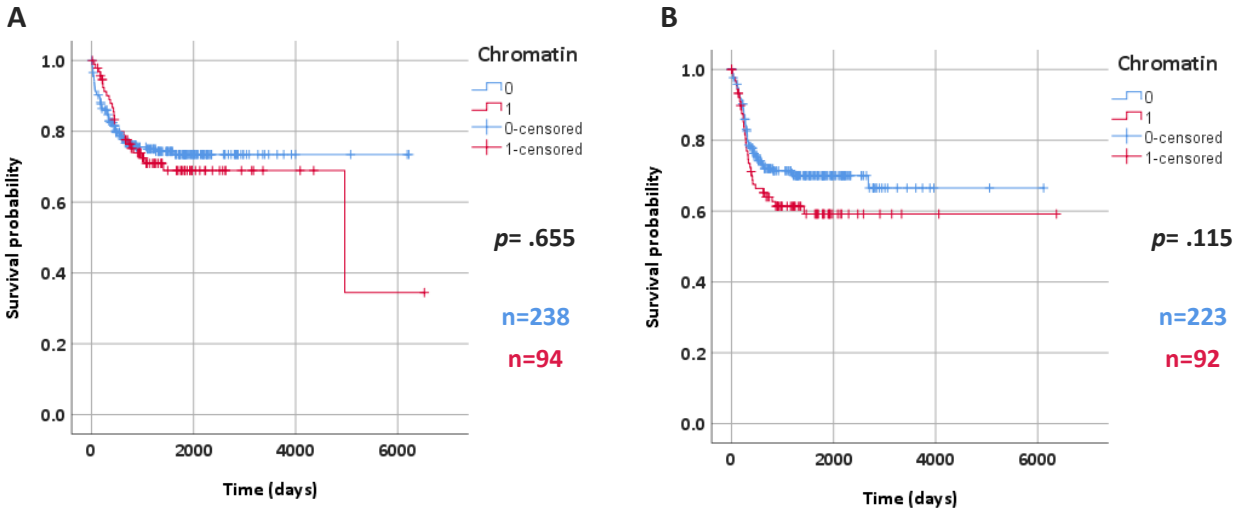
Supplemental Figure S22: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to cohesin group mutations. Cohesin group: *RAD21*, *SMC1A*, *SMC3*, *STAG2*, *NIPBL*.



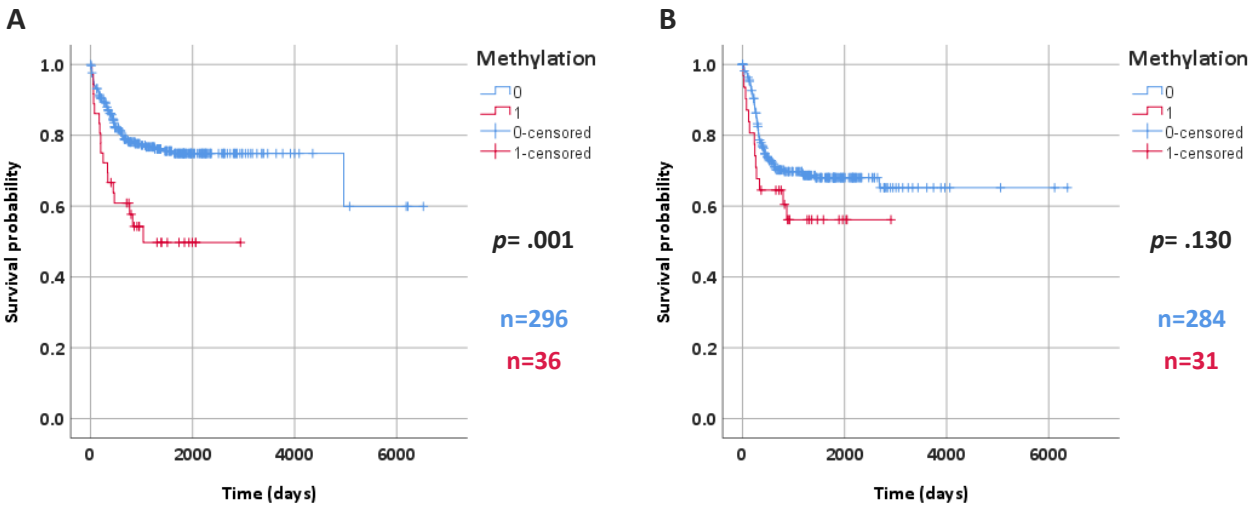
Supplemental Figure S23: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to transcription group mutations. Transcription group: *ZBTB7A*, *WT1*, *BCOR*, *BCORL1*, *FOXP1*.



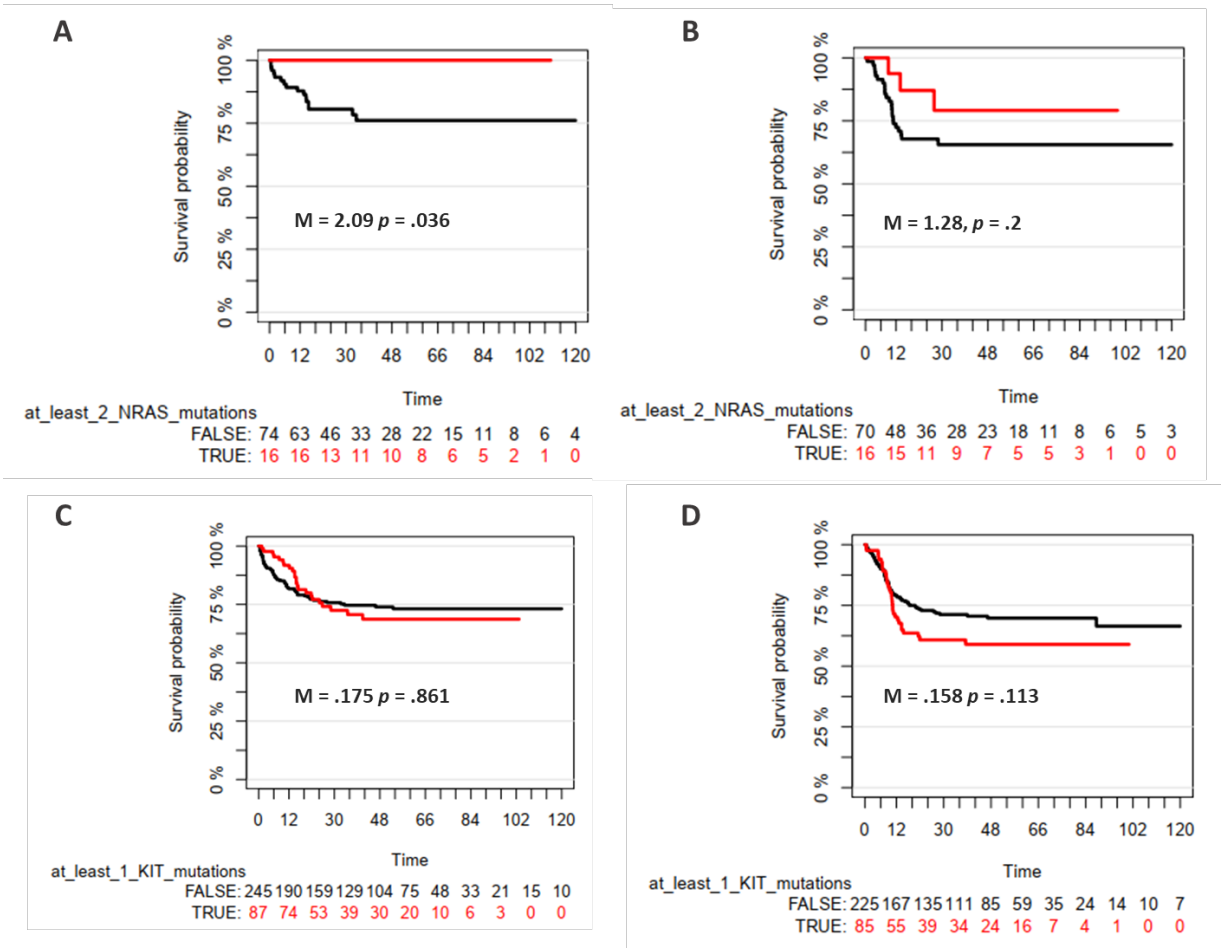
Supplemental Figure S24: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to chromatin group mutations. Chromatin group: *ASXL1*, *ASXL2*, *SRCAP*, *KDM6A*, *EZH2*, *SETD2*, *KMT2D*.



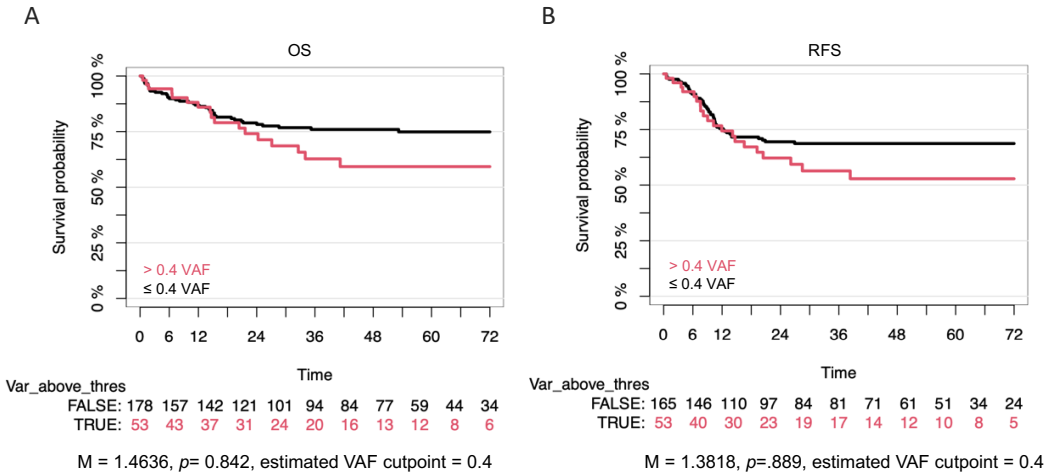
Supplemental Figure S25: Kaplan-Meier estimates for A) overall survival and B) relapse-free survival in CBF cohort according to methylation group mutations. Methylation group: *TET2*, *DNMT3A*, *IDH1*, *IDH2*.



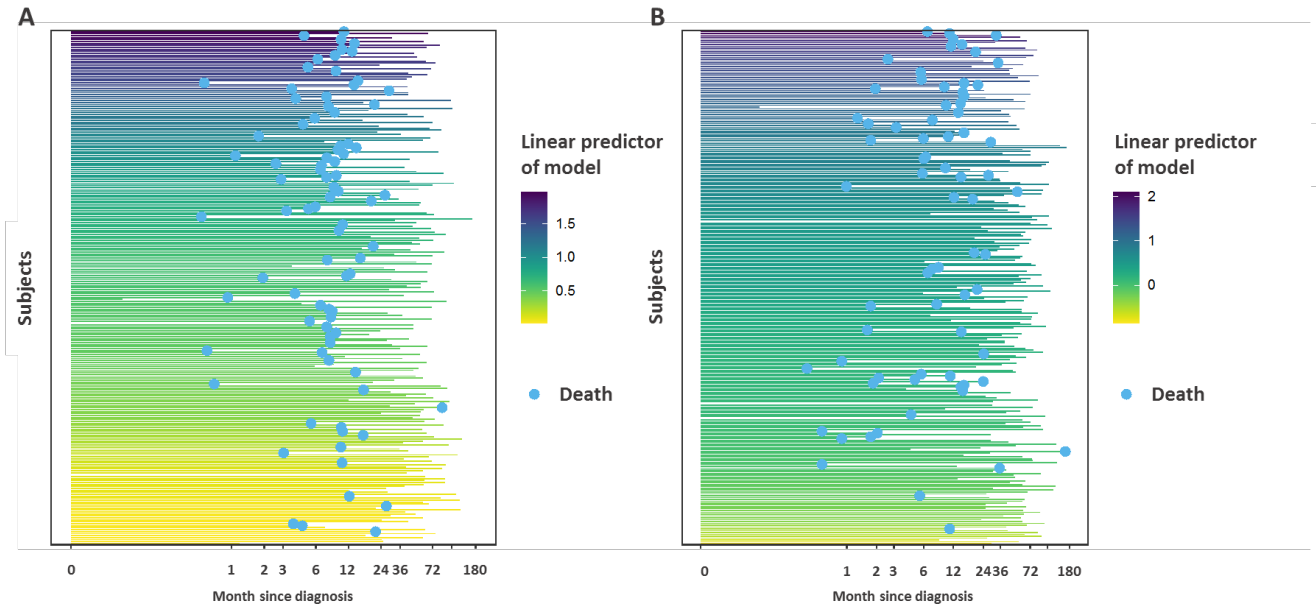
Supplemental Figure S26: Kaplan-Meier estimates for overall (A,C) and relapse-free (B,D) survival according to the number of *NRAS* (A,B) and *KIT* (C,D) mutations of the entire CBF-AML cohort.



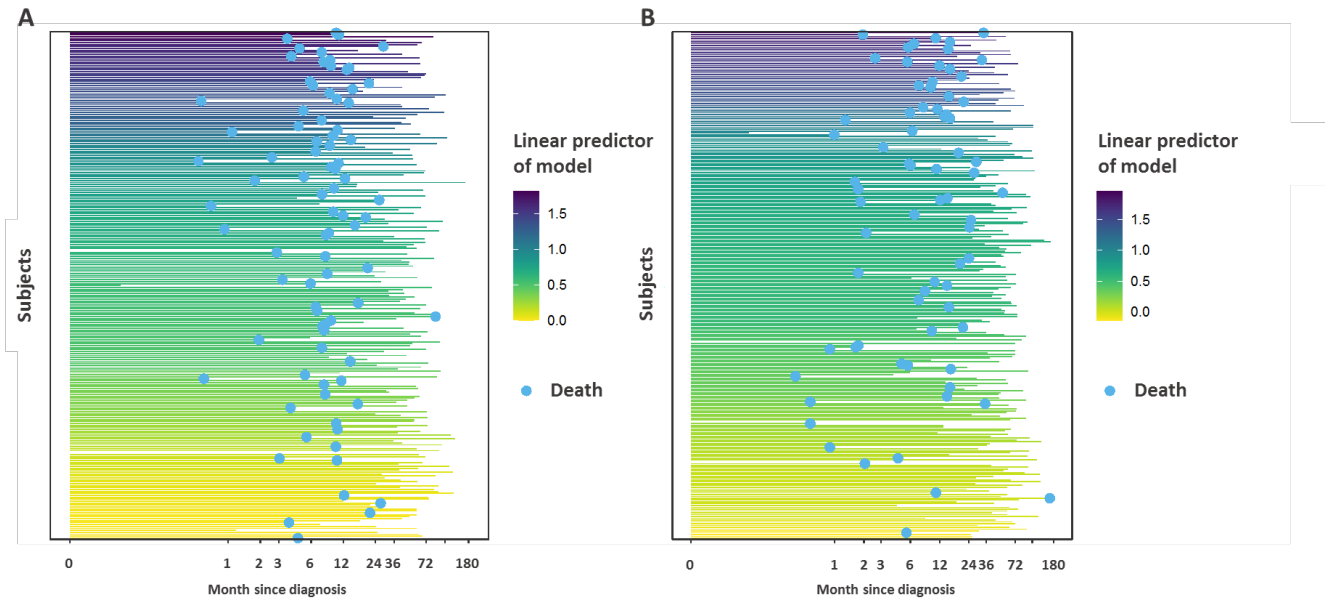
Supplemental Figure S27: Kaplan-Meier estimates for A) overall and B) relapse-free survival according to the variant allele fraction (VAF) in the signaling gene group. Maximally selected LogRank statistics were used for VAF cut-off determination. Kaplan-Meier estimators were derived for the cut point providing the best separation into two groups.



Supplemental Figure S28: Swimmer plots based on 5-fold cross validation of optimized lasso models of relapse-free (A) and overall survival (B) according to clinical features, mutational status and chromosomal aberrations in the entire CBF-AML cohort. A higher linear predictor is associated with a higher risk of death as indicated by clustering of events in the upper half of plots. Each bar represents one patient, the length of the bar the observation time since diagnosis, the color of the bar the corresponding linear predictor of the model and the light blue dots indicate death.



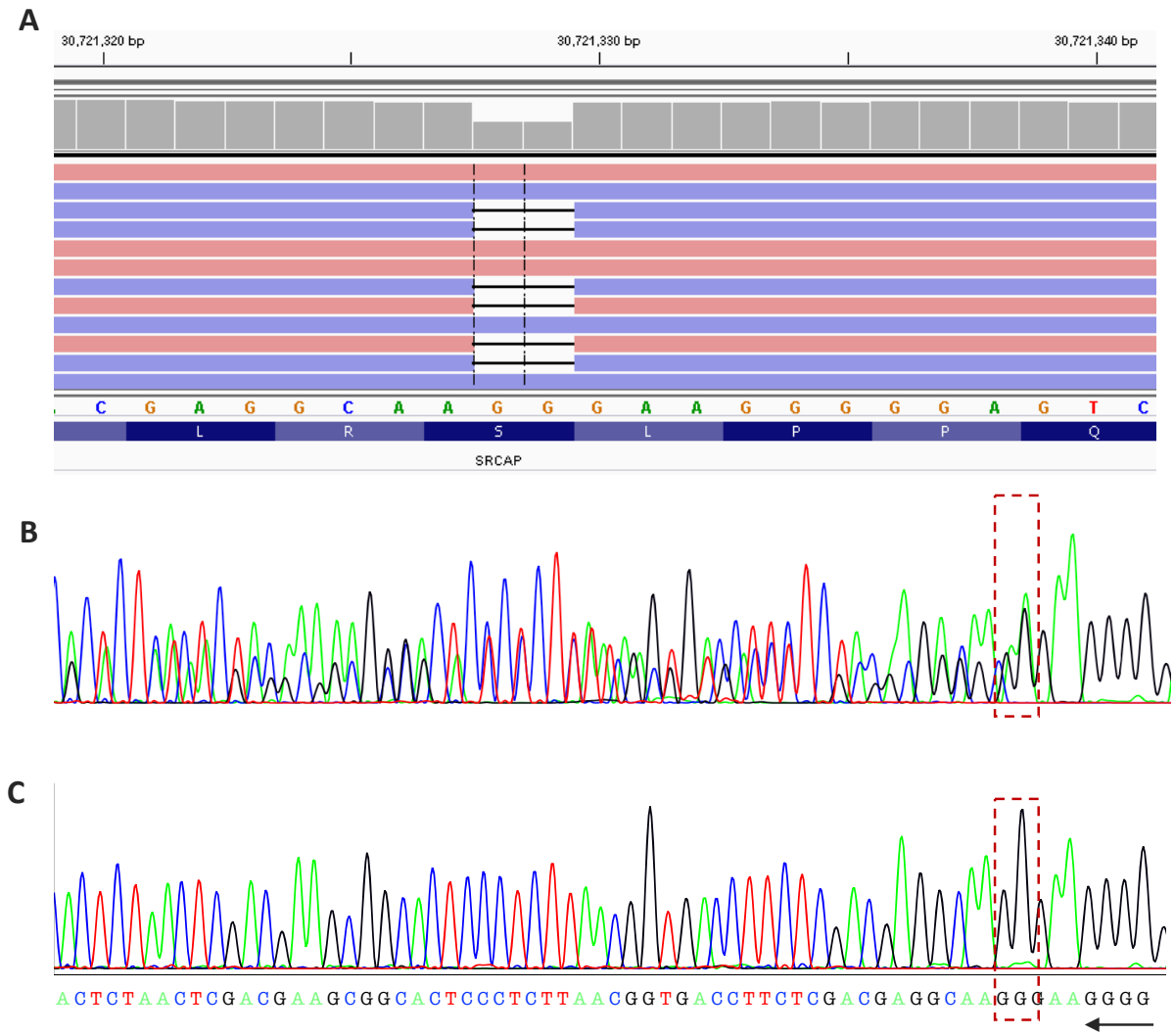
Supplemental Figure S29: Swimmer plots based on 5-fold cross validation of optimized lasso models of relapse-free (A) and overall survival (B) according to clinical features, functional groups and chromosomal aberrations in the entire CBF-AML cohort. A higher linear predictor is associated with a higher risk of death as indicated by clustering of events in the upper half of plots. Each bar represents one patient, the length of the bar the observation time since diagnosis, the color of the bar the corresponding linear predictor of the model and the light blue dots indicate death.

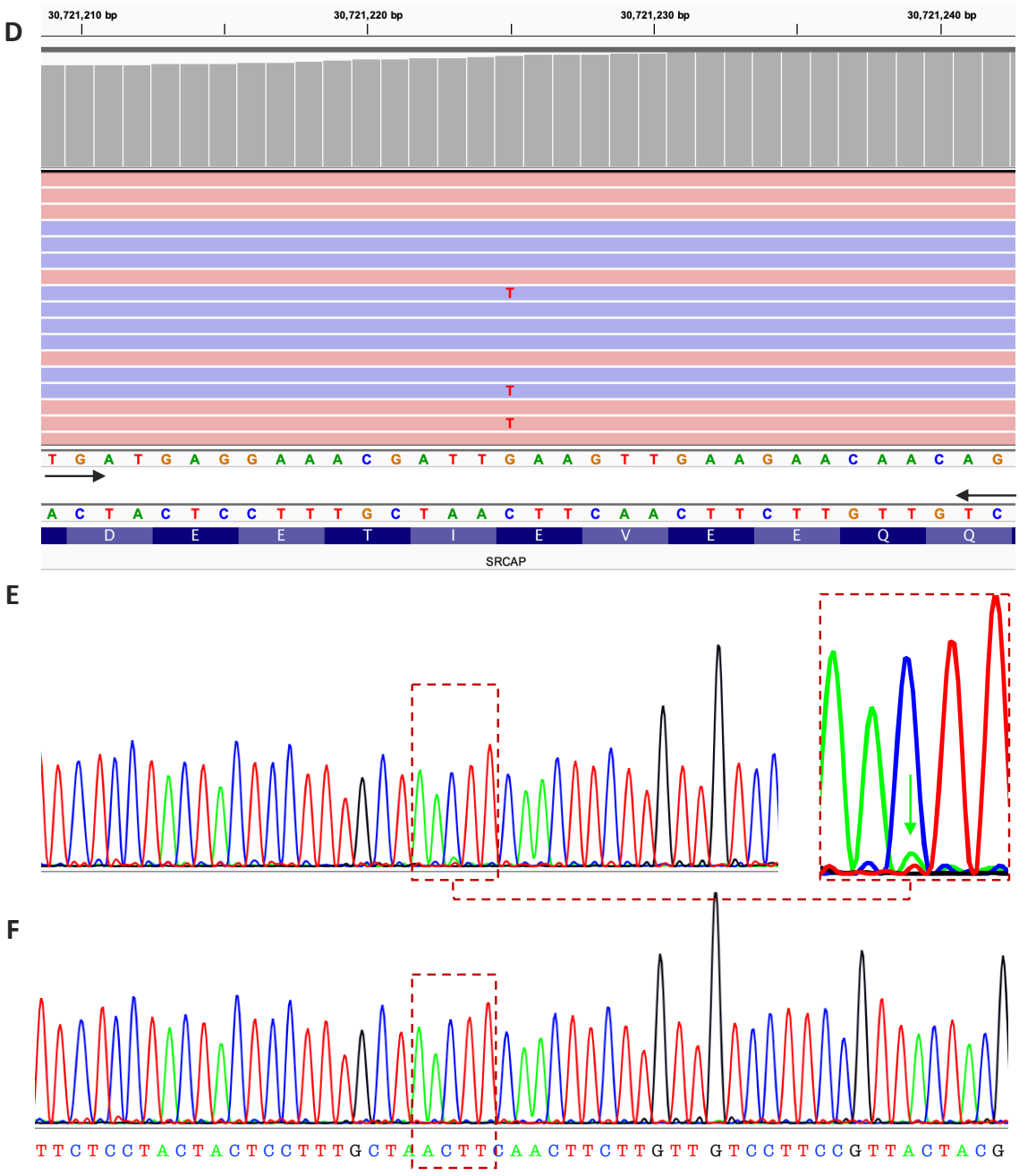


Supplemental Figure S30: Validation of somatic origin of *SRCAP* variants in two randomly selected patients.

A) *SRCAP* p.S338fs (VAF: 39%) mutation (reference sequence NM_006662) of patient CBF_270 (NGS sequence) and B) corresponding Sanger sequence (read direction flipped) at diagnosis. C) Corresponding wildtype sequence (read direction flipped) of the same patient at molecular remission as germline control (molecular remission determined by RQ-PCR analysis of CBF fusion transcripts).

D) *SRCAP* p.E304X (c.G910T; VAF 9%) of patient CBF_34 (NGS sequence) and E) corresponding reverse Sanger sequence at diagnosis. Right dashed box: magnification of altered region. Green arrow indicates substitution of C by A, corresponding to T on forward strand (smaller peaks background noise) F) Corresponding wildtype reverse Sanger sequence of the same patient at molecular remission as germline control (molecular remission determined by RQ-PCR analysis of CBF fusion transcripts).

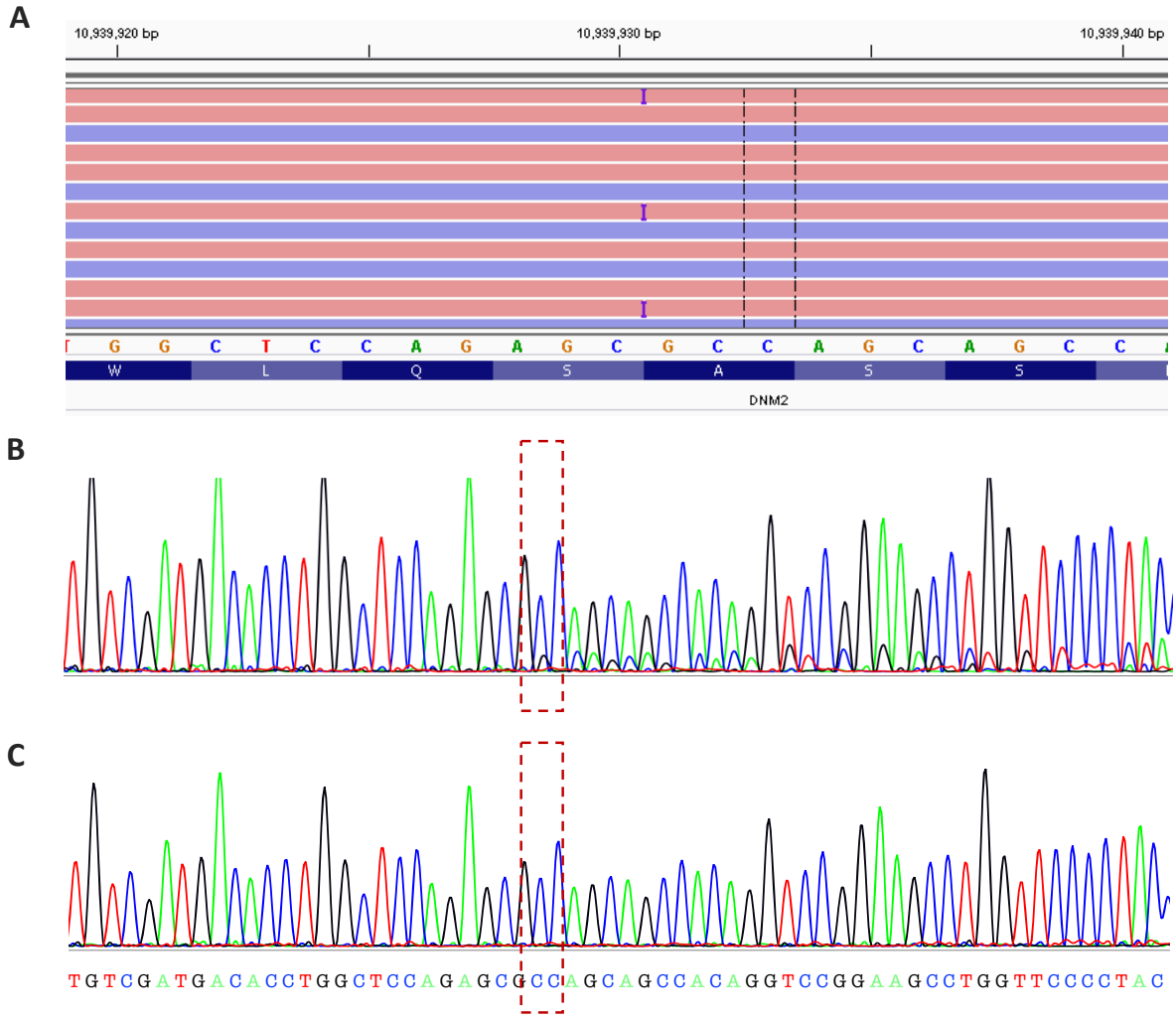



D
30,721,210 bp
30,721,220 bp
30,721,230 bp
30,721,240 bp
T
T
T
T G A T G A G G A A A C G A T T G A A G T T G A A G A A C A A C A G
A C T A C T C C T T T G C T A A C T T C A A C T T C T T G T T C C T T C C G T T A C T A C G
D E E T I E V E E Q Q
SRCAP
E
F
T T C T C C T A C T A C T C C T T T G C T A A C T T C A A C T T C T T G T T G T C C T T C C G T T A C T A C G

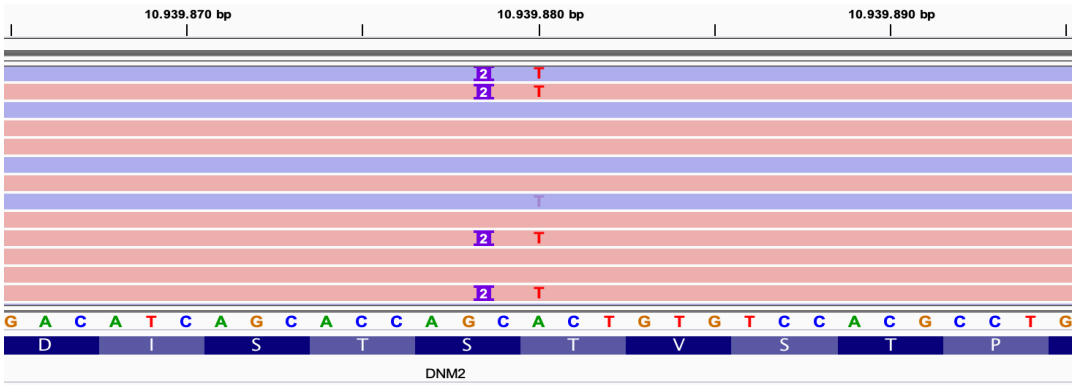
Supplemental Figure S31: Validation of somatic origin of *DNM2* variants in two randomly selected patients.

A) *DNM2* p.S755fs (VAF 12%) mutation (reference sequence NM_004945) of patient CBF_69 (NGS sequence) and B) corresponding Sanger sequence at diagnosis. C) Corresponding wildtype sequence of the same patient at molecular remission as germline control (molecular remission determined by RQ-PCR analysis of CBF fusion transcripts).

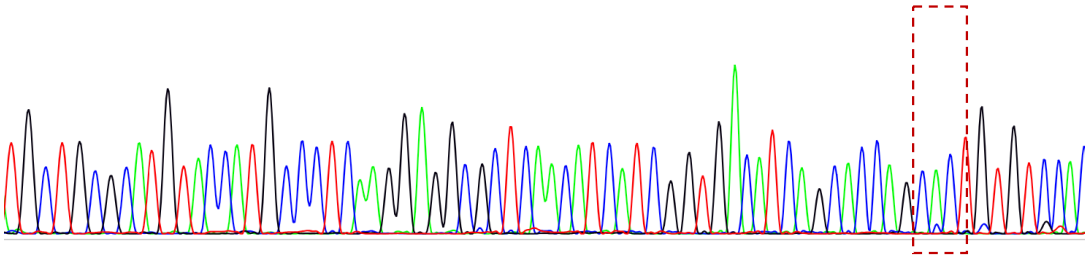
D) *DNM2* p.738fs (VAF 7%) mutation (reference sequence NM_004945) of patient CBF_61 (NGS sequence) and E) corresponding Sanger sequence at diagnosis. F) Corresponding wildtype sequence of the same patient at molecular remission as germline control (molecular remission determined by RQ-PCR analysis of CBF fusion transcripts).



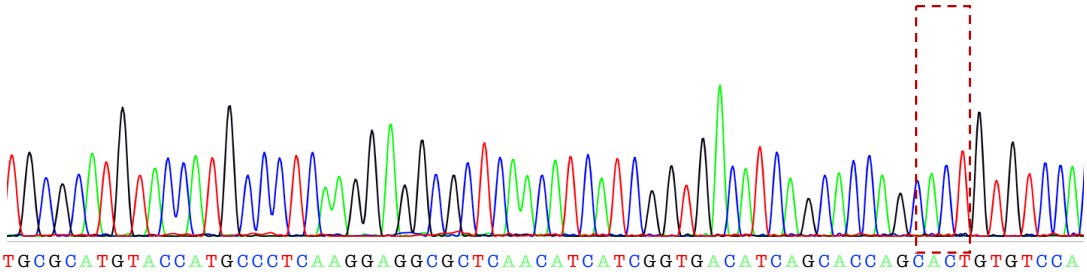
D


10.939.870 bp
10.939.880 bp
10.939.890 bp
G A C A T C A G C A C C A G C A C T G T G T C C A C G C C T G
D I S T S T V S T P
DNM2

E

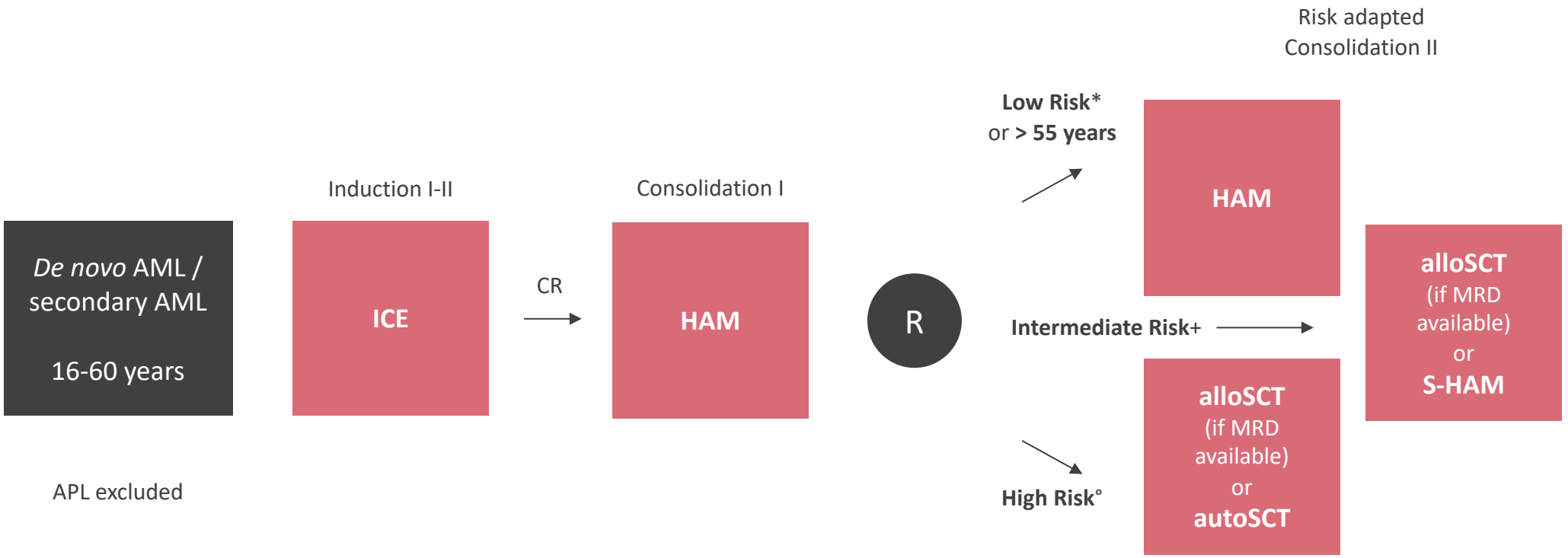


F


TGCGCATGTACCATGCCCTCAAGGAGGCGCTCAACATCATCGGTGACATCAGCACCAGCACTGTGTCCA

Supplemental Figure S32: AML HD93 trial

AML HD93



*low risk: t(8;21), inv(16); +intermediate risk: normal karyotype; °high risk: all other chromosomal abnormalities

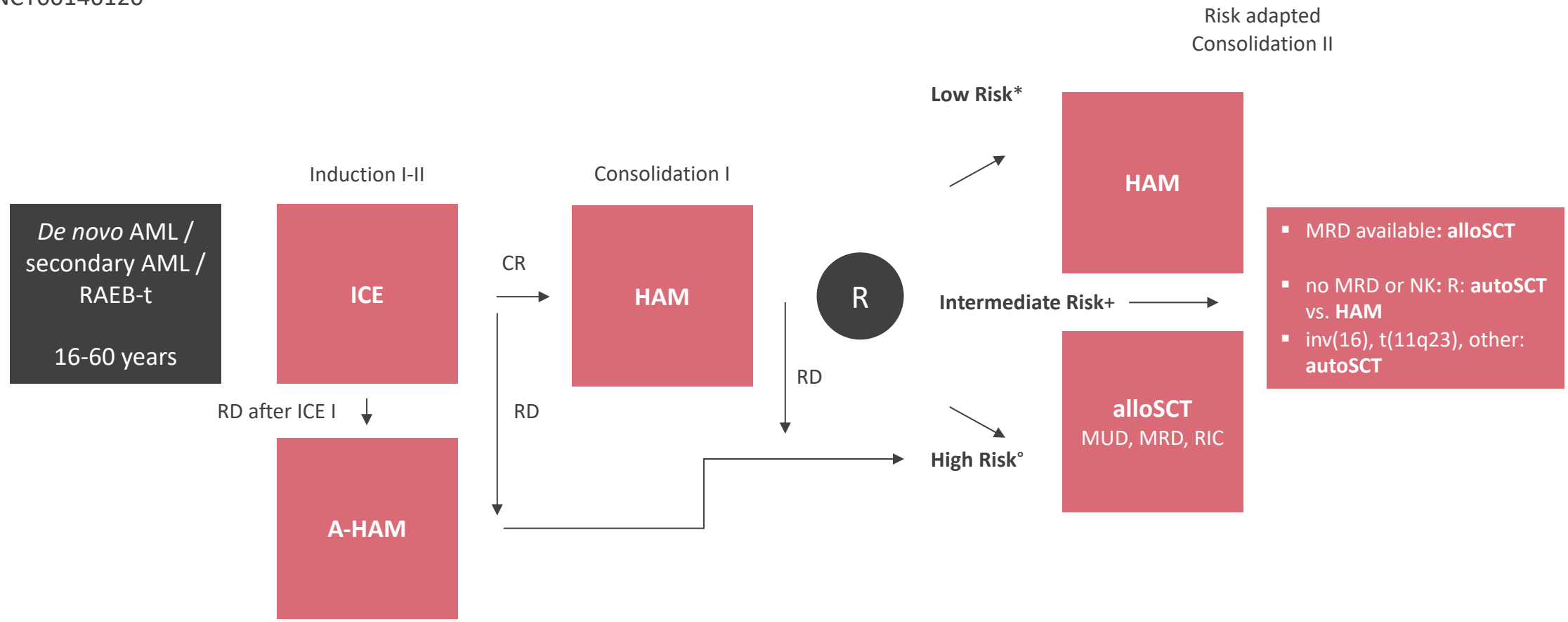
ICE: idarubicin 12 mg/m² d1,3,5; cytarabine 100 mg/m² d1-7; etoposide 100 mg/m² d1-3
HAM: cytarabine 3 g/m²/12h d1-3; mitoxantrone 12 mg/m² d2-3
S-HAM: cytarabine 3 g/m²/12 h, d1, 2, 8 ,9; mitoxantrone 10 mg/m² d 3,4,10,11

Abbreviations: alloSCT, allogeneic transplant; autoSCT, autologous transplant; CR, complete remission; MRD, matched related donor, R, randomization

Supplemental Figure S33: AML HD98A trial

AML HD98A

NCT00146120

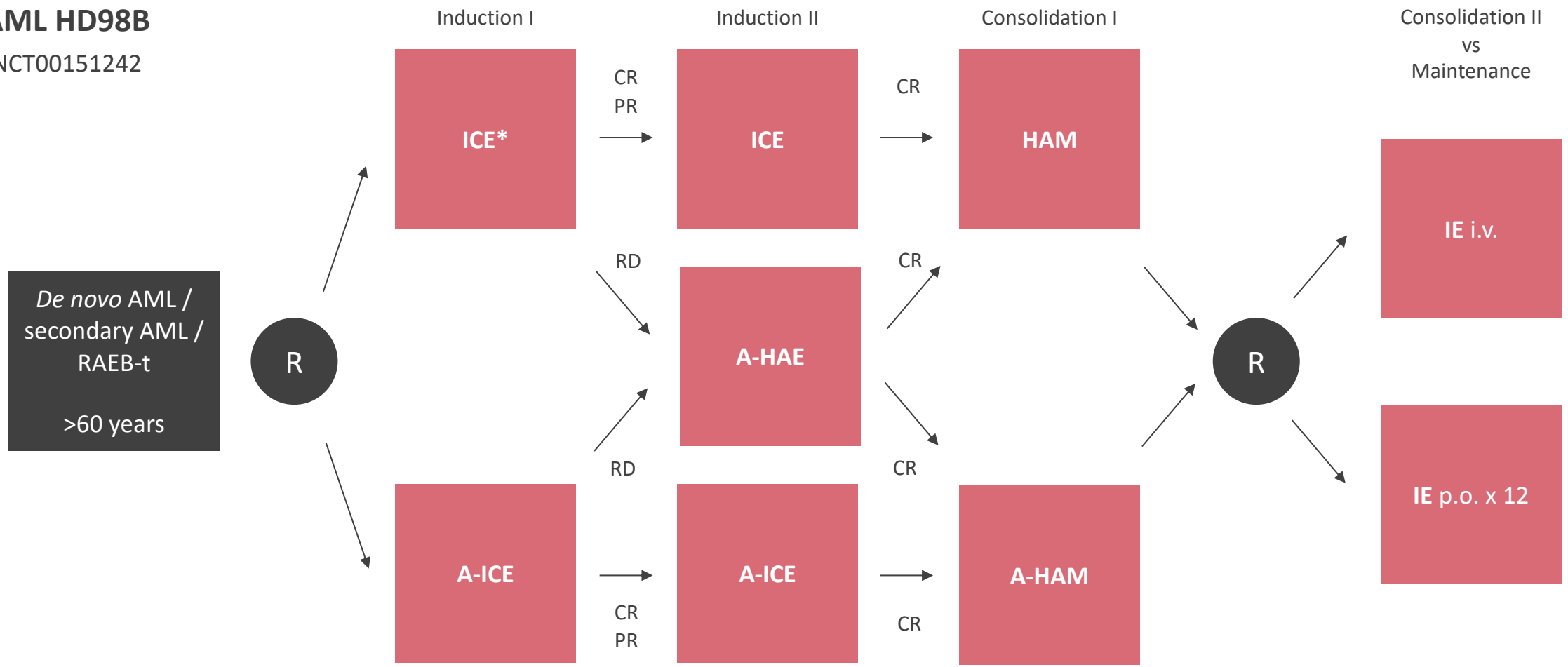


*low risk: t(8;21), t(15;17); +intermediate risk: normal karyotype, inv(16), t(11q23), other (< 3 aberrations); °high risk: abn(3q), -5/5q-, -7/7q-, abn(12p), ≥3 aberrations, RD

ICE: idarubicin 12 mg/m² d1,3,5; cytarabine 100 mg/m² d1-7; etoposide 100 mg/m² d1-3 / **HAM**: cytarabine 3 g/m²/12h d1-3; mitoxantrone 12 mg/m² d2-3 /
A-HAM: HAM plus all-trans retinoic acid 45 mg/m² d3-5, 15 mg/m² d6-28

Abbreviations: alloSCT, allogeneic transplant; autoSCT, autologous transplant; CR, complete remission; MRD, matched related donor; MUD, matched unrelated donor; NK, normal karyotype; RAEB-t, refractory anemia with excess of blasts in transformation; R, randomization; RD, refractory disease; RIC, reduced intensity conditioning

Supplemental Figure S34: AML HD98B trial



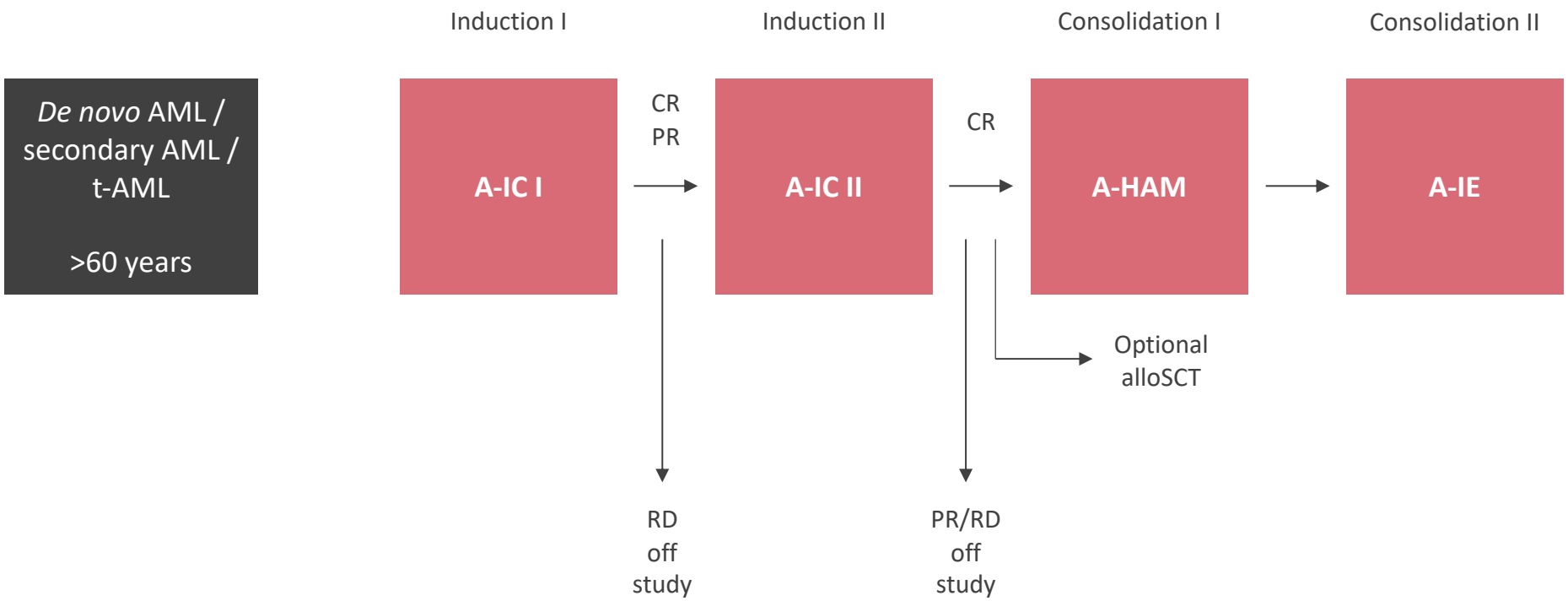
*Aida + HAM+ AID Apo in case of APL with t(15;17)
ICE: idarubicin 12 mg/m² d1,3; cytarabine 100 mg/m² d1-5; etoposide 100 mg/m² d1,3 / **A-ICE**: ICE plus plus all-trans retinoic acid 45 mg/m² d3-5, 15 mg/m² d6-28
HAM: cytarabine 0.5 g/m²/12h d1-3; mitoxantrone 10 mg/m² d2-3 / **A-HAM**: HAM plus all-trans retinoic acid 45 mg/m² d3-5, 15 mg/m² d6-28
IE i.v.: idarubicin 12mg/m² d1,2; etoposide 100 mg/m² d1-5 / **IE p.o.**: idarubicin 5 mg d1,4,7,10,13; etoposide 100mg d1,13
A-HAE: cytarabine 0.5 g/m²/12h d1-3; etoposide 250 mg/m² d4,5; all-trans retinoic acid 45 mg/m² d3-5, 15 mg/m² d6-28

Abbreviations: CR, complete remission; RAEB-t, refractory anemia with excess of blasts in transformation; PR, partial remission; R; randomization; RD, refractory disease

Supplemental Figure S35: AMLSG 06-04 trial

AMLSG 06-04

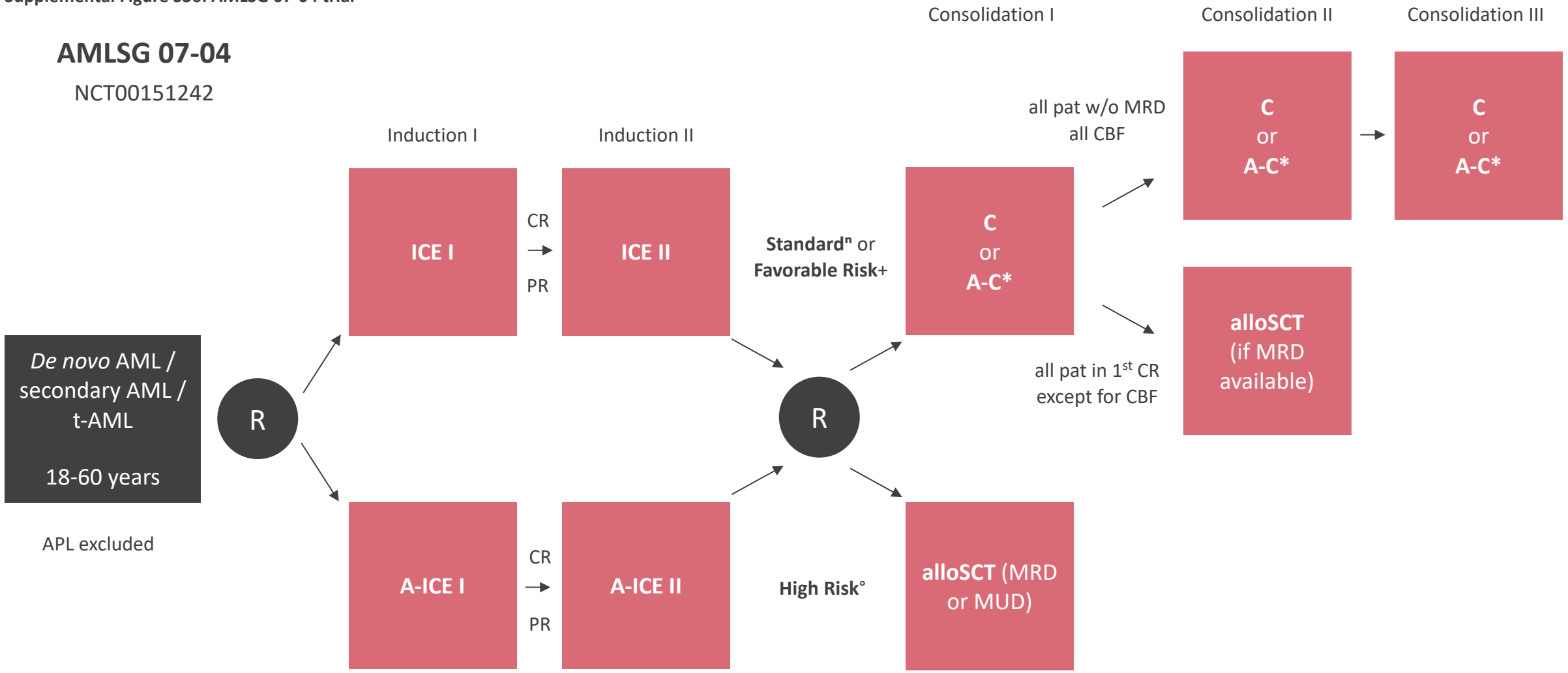
NCT00151255



A-IC: idarubicin 12 mg/m² d1,3; cytarabine 100 mg/m² d1-5; all-trans retinoic acid 45 mg/m² d4-6, 15 mg/m² d7-28
A-HAM: cytarabine 1 g/m²/12h d1-3; mitoxantrone 10 mg/m² d2-3; all-trans retinoic acid 15 mg/m² d4-28
A-IE: idarubicin 12 mg/m² d1,3; etoposide 100 mg/m² d1-5; all-trans retinoic acid 15 mg/m² d4-28

Abbreviations: alloSCT, allogeneic transplant; CR, complete remission; PR, partial remission; RD, refractory disease; t-AML, therapy-related AML

Supplemental Figure S36: AMLSG 07-04 trial



+favorable risk :CBF AML: inv(16), t(8;21); ⁿstandard risk: cytogenetic abnormalities not classified as favorable or adverse; °high risk: *FLT3*-ITD, inv(3) or t(3;3), t(6;9); t(v;11), -5/5q-, -7/7q-, abn(17p), ≥3 aberrations, RD

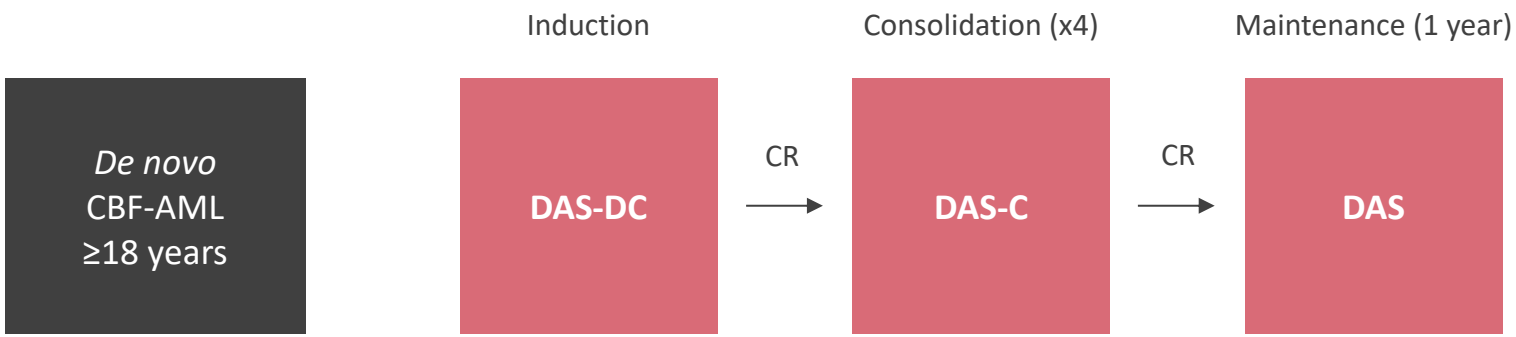
ICE: idarubicin 12 mg/m² d1,3; cytarabine 100 mg/m² d1-5; etoposide 100 mg/m² d1,3 / **A-ICE**: ICE plus all-trans retinoic acid 45mg/m² d6-8, 15 mg/m² d9-21
C: cytarabine 3 g/m²/12h, days 1-3 / **A-C** : C plus all-trans retinoic acid 15 mg/m² d4-21, *only pat. previously randomized to A-ICE arm

Abbreviations: CR, complete remission; MRD, matched related donor; MUD, matched unrelated donor; PR, partial remission; R; randomization; RD, refractory disease; t-AML, therapy-related AML

Supplemental Figure S37: AMLSG 11-08 trial

AMLSG 11-08

NCT00850382



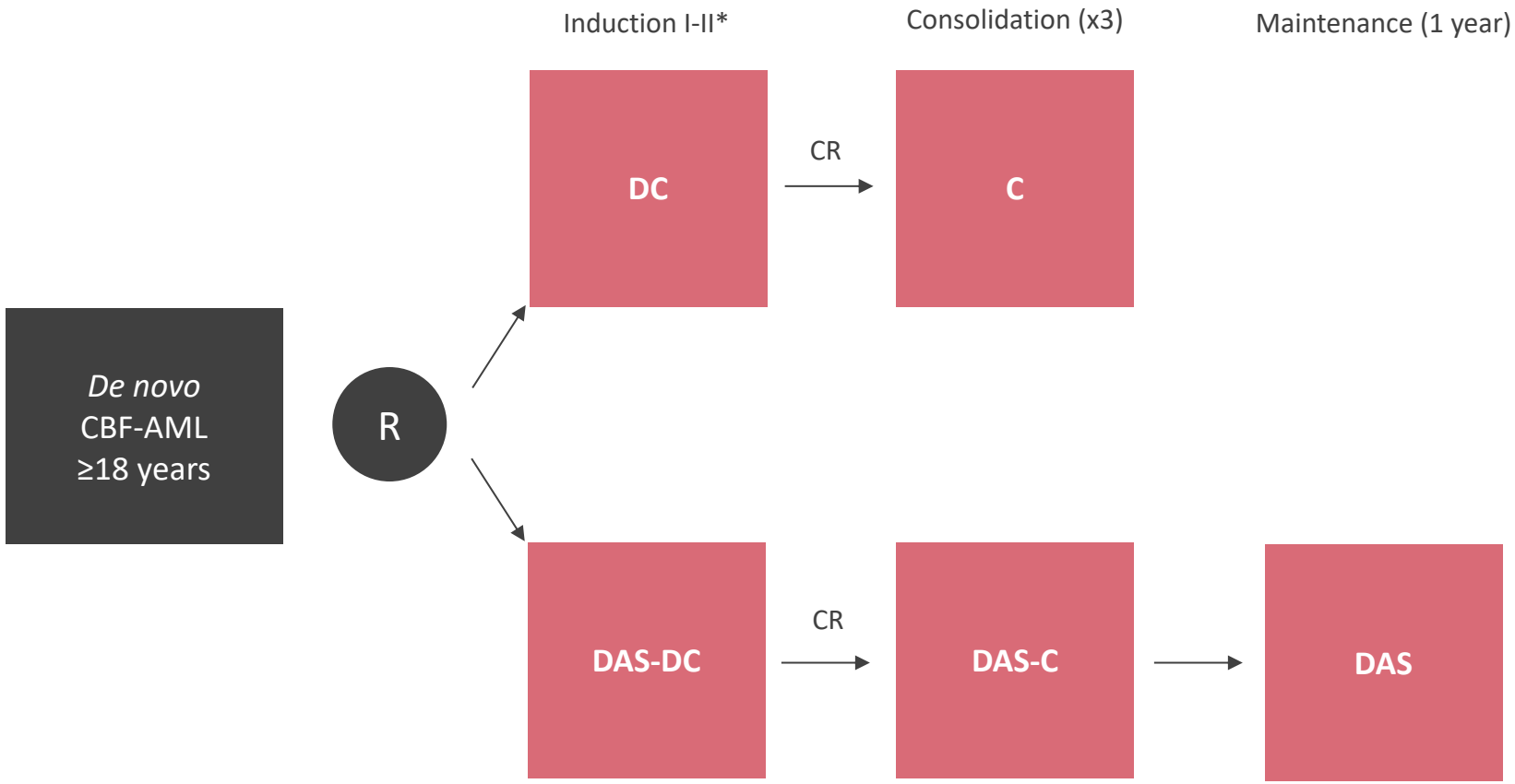
DAS-DC: daunorubicin 60 mg/m² d1–3; cytarabine 200 mg/m² d 1–7; dasatinib 100 mg p.o. d8-21
DAS-C: cytarabine 3000 mg/m²/12h d 1,3,5 (>60 years: 1000 mg/m²); dasatinib 100 mg p.o. d6-21
DAS: dasatinib 100 mg p.o. for one year

Abbreviations: CR, complete remission; CBF: corebinding-factor AML [inv(16), t(8;21)]

Supplemental Figure S38: AMLSG 21-13 trial

AMLSG 21-13

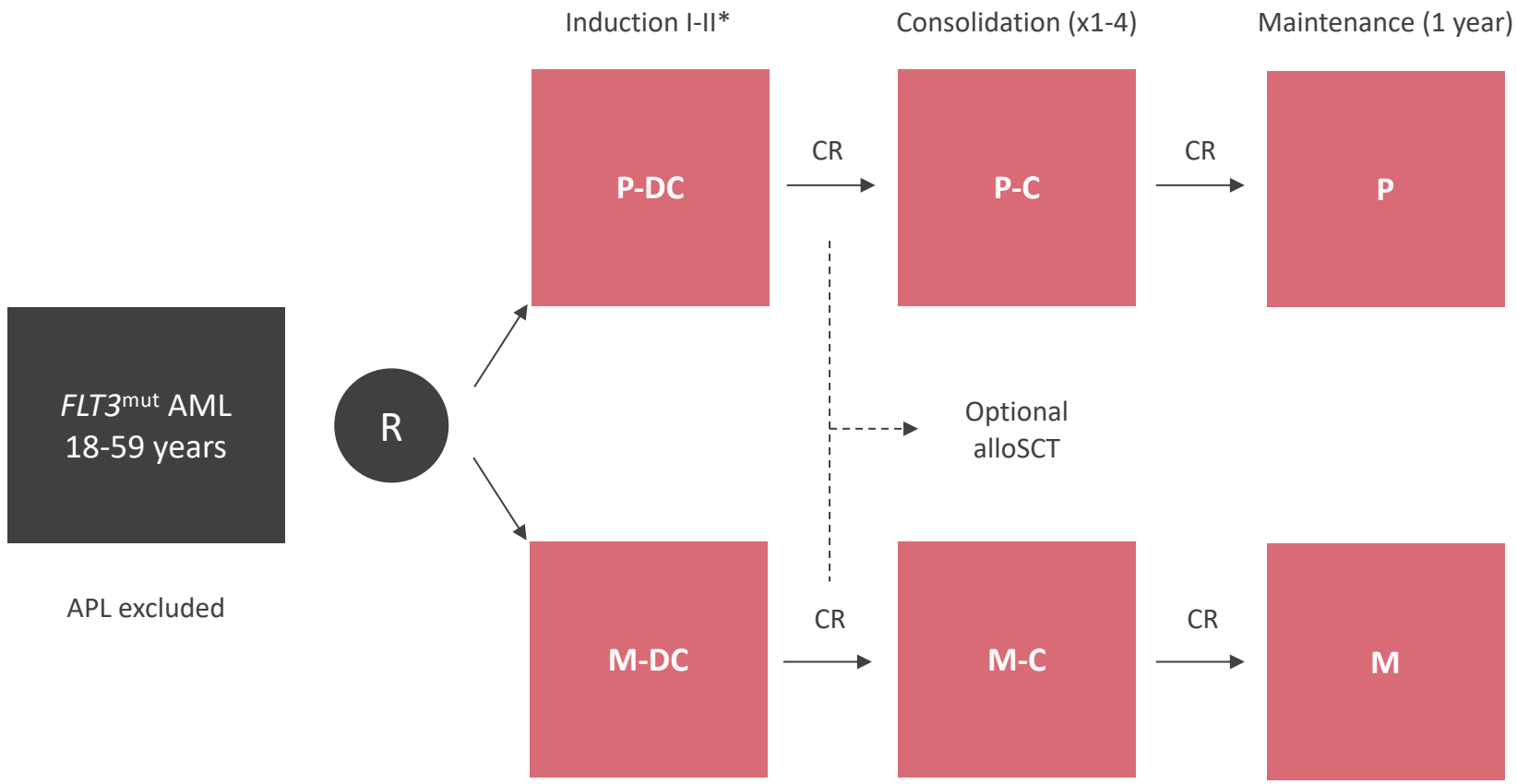
NCT02013648



DC: daunorubicin 60 mg/m² d1–3; cytarabine 200 mg/m² d 1–7 / **DAS-DC**: DC plus dasatinib 100 mg p.o. d8-21
*pat achieving PR after cycle 1: 2nd induction cycle with daunorubicin 50 mg/m² d1-3; cytarabine 200 mg/m²d1-5; dasatinib 100 mg d6-21
C: cytarabine 3000 mg/m²/12h d 1-3 (>60 years: 1000 mg/m²) / **DAS-C**: C plus dasatinib 100 mg p.o. d4-21
DAS: dasatinib 100 mg p.o. for one year

Abbreviations: CR, complete remission; CBF: corebinding-factor AML [inv(16), t(8;21)], PR, partial remission; R, randomization

Supplemental Figure S39: RATIFY trial



P-DC: daunorubicin 60 mg/m² d1–3; cytarabine 200 mg/m² d 1–7; placebo bid d8-21 / **M-DC**: daunorubicin 60 mg/m² d1–3; cytarabine 200 mg/m² d 1–7; midostaurin 50 mg bid d8-21
*pat achieving PR after cycle 1: 2nd induction cycle with daunorubicin 60 mg/m² d1-3; cytarabine 200 mg/m²d1-5; plus M or P
P-C: cytarabine 3000 mg/m²/12h d 1,3,5; placebo bid d8-21 / **M-C**: cytarabine 3000 mg/m²/12h d 1,3,5; midostaurin 50 mg bid d8-21
P: placebo bid for one year / **M**: midostaurin 50 mg bid for one year

Abbreviations: alloSCT, allogeneic transplant; bid, twice daily; CR, complete remission, PR, partial remission; R, randomization