SUPPLEMENTARY INFORMATION FOR:

The Genomic Landscape of Core-Binding Factor Acute Myeloid Leukemias

Zachary J. Faber[1,7], Xiang Chen[2,7], Amanda Larson Gedman[1,7], Kristy Boggs[2], Jinjun Cheng[1], Jing Ma[1], Ina Radtke[1], Jyh-Rong Chao[1], Michael P. Walsh[1], Guangchun Song[1], Anna K. Andersson[1‡], Jinjun Dang[1], Li Dong[1], Yu Liu[2], Robert Huether[2], Zhongling Cai[1], Heather Mulder[2], Gang Wu[2], Michael Edmonson[2], Michael Rusch[2], Chunxu Qu[2], Yongjin Li[2], Bhavin Vadodaria[2], Jianmin Wang[2], Erin Hedlund[2], Xueyuan Cao[3], Donald Yergeau[2], Joy Nakitandwe[1], Stanley B. Pounds[3], Sheila Shurtleff[1], Robert S. Fulton[5], Lucinda L. Fulton[5], John Easton[2], Evan Parganas[1], Ching-Hon Pui[6], Jeffrey E. Rubnitz[6], Li Ding[5], Elaine R. Mardis[5], Richard K. Wilson[5], Tanja A. Gruber[6], Charles G. Mullighan[1], Richard F. Schlenk[4], Peter Paschka[4], Konstanze Döhner[4], Hartmut Döhner[4], Lars Bullinger[4*], Jinghui Zhang[2*], Jeffery M. Klco[1*], James R. Downing[1*]

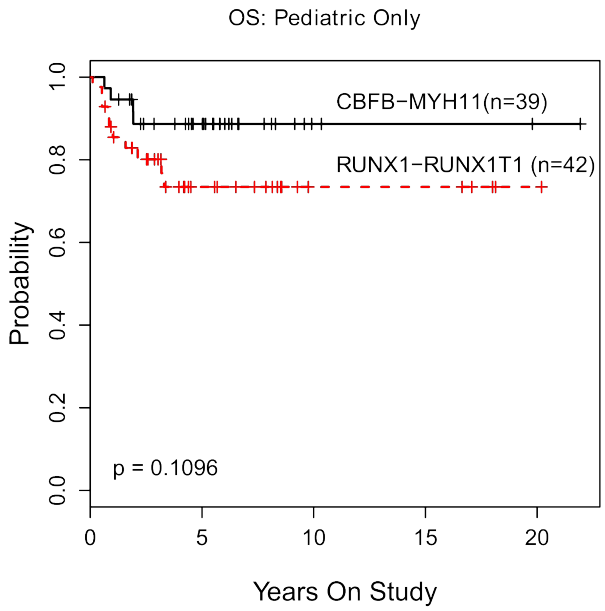
[1]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA
[2]Department of Computational Biology and Bioinformatics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA
[3]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA
[4]Department of Internal Medicine III, University of Ulm, Ulm, Germany
[5]McDonnell Genome Institute at Washington University, St. Louis, Missouri, USA
[6]Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA
[7]These authors contributed equally to this work

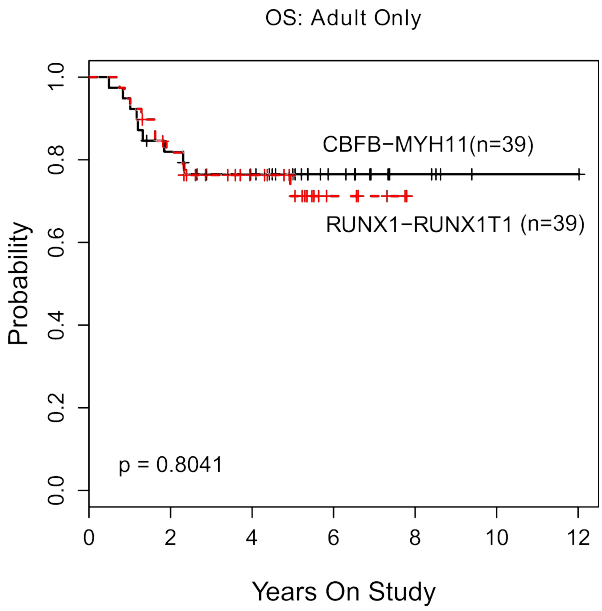
‡ Current affiliation: Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Lund, Sweden

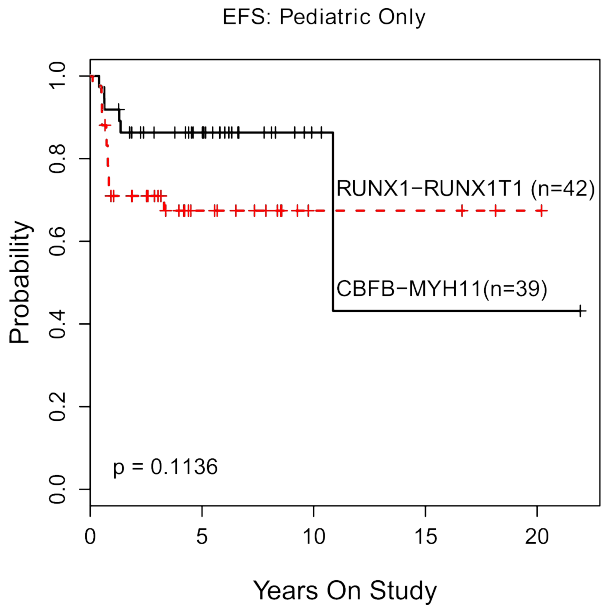
[*]Corresponding Authors: Lars Bullinger (lars.bullinger@uniklinik-ulm.de), Jinghui Zhang (Jinghui.Zhang@StJude.org), Jeffery M. Klco (Jeffery.Klco@StJude.org), and James R. Downing (James.Downing@StJude.org)

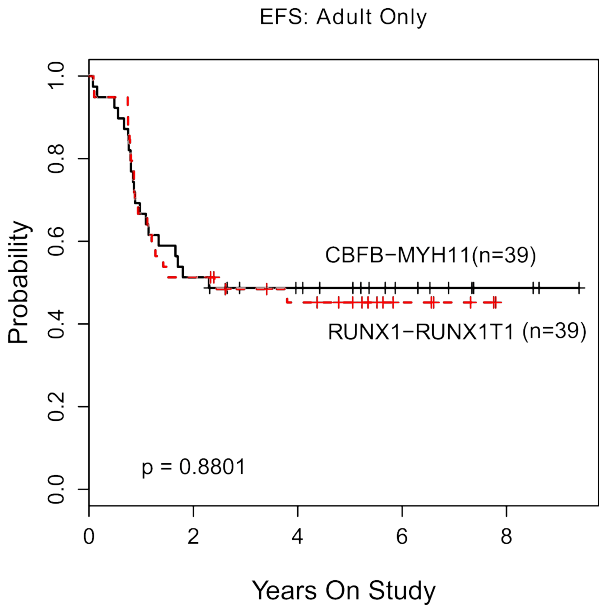
TABLE OF CONTENTS

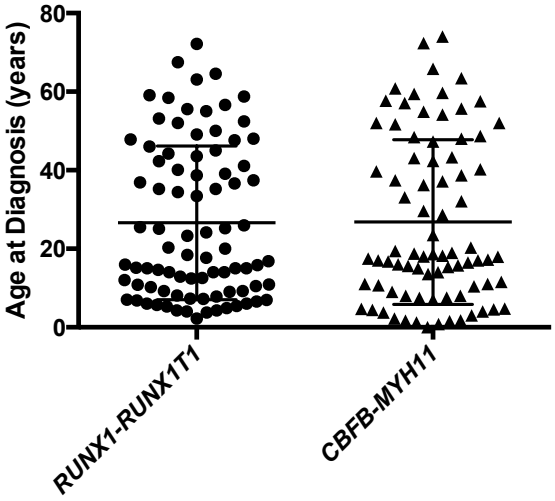
Supplementary Figure 1. Comparison of outcomes and ages in CBF-AML. Kaplan-Meier curves depicting overall survival (OS) and event free survival (EFS) are shown for RUNX1-RUNX1T1 and CBFB-MYH11 cases in adult and pediatric cases. Number of cases in each group is shown in parentheses. P values were determined by log rank test.


OS: Pediatric Only
CBFB−MYH11(n=39)
RUNX1−RUNX1T1 (n=42)
Probability
p = 0.1096
Years On Study

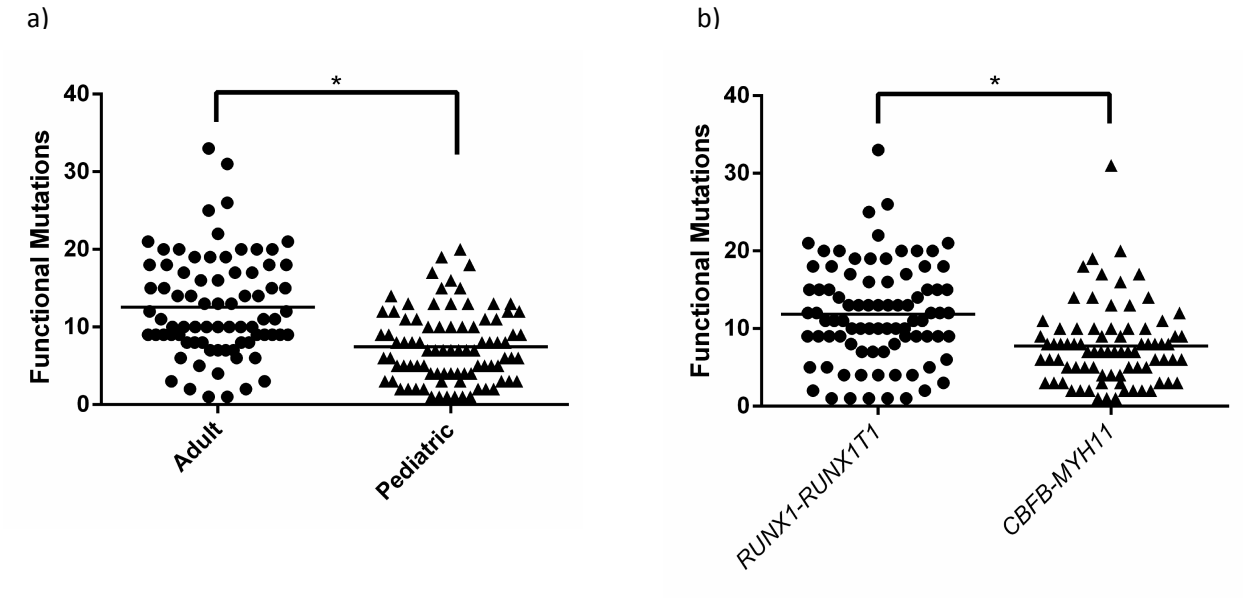

OS: Adult Only
CBFB−MYH11(n=39)
RUNX1−RUNX1T1 (n=39)
Probability
p = 0.8041
Years On Study


EFS: Pediatric Only
RUNX1−RUNX1T1 (n=42)
CBFB−MYH11(n=39)
Probability
p = 0.1136
Years On Study


EFS: Adult Only
CBFB−MYH11(n=39)
RUNX1−RUNX1T1 (n=39)
Probability
p = 0.8801
Years On Study

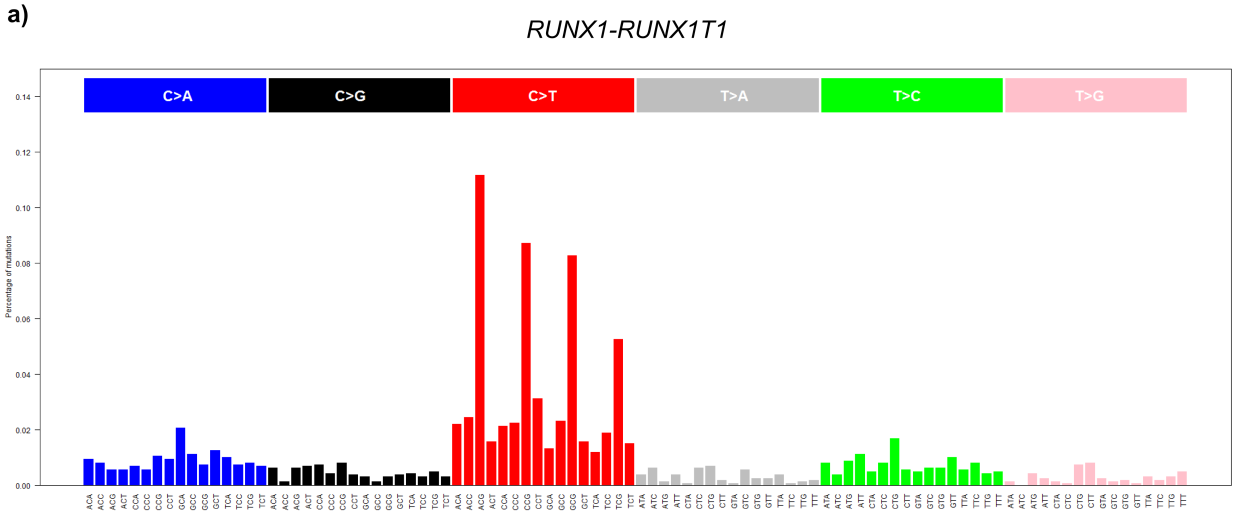

Age at Diagnosis (years)
RUNX1-RUNX1T1
CBFB-MYH11

Supplementary Figure 2. Mutation number in CBF-AML cases. The number of functional somatic mutations is shown for each case broken down by both a) age group and b) fusion status. Functional mutations are defined as coding missense, nonsense, and indel mutations as well as mutations at splice sites. The mean of the data is shown. * p <0.0001, exact Wilcoxon rank-sum test based on 10,000 permutations.

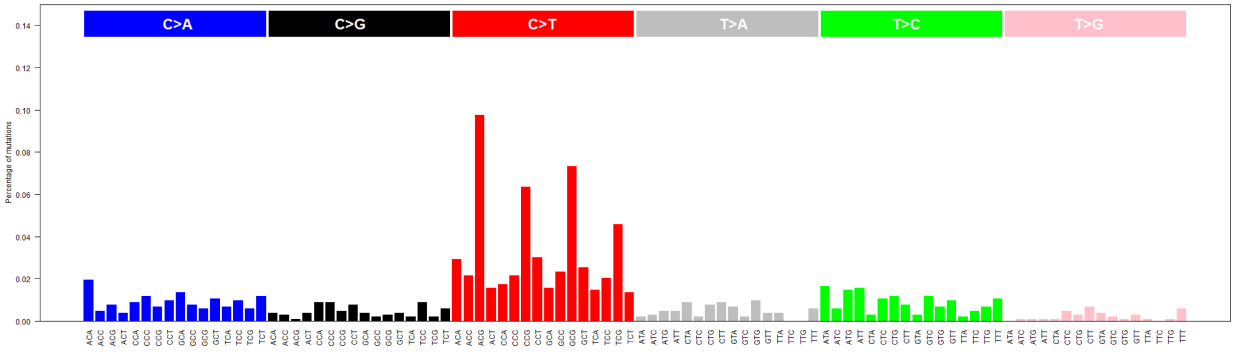


Supplementary Figure 3: Trinucleotide mutational spectrum of CBF AML. Patterns of substitutions for the identified mutation signature in a) *RUNX1-RUNX1T1* AML and b) *CBFB-MYH11* AML, which is displayed as a percentage of mutations in the context of each of the 96 possible combinations of three nucleotides in the human genome where the center base is the location of the mutation. The mutation types are displayed on the horizontal axes, while vertical axes depict the percentage of mutations attributed to a specific mutation type. The depicted profiles are similar between *RUNX1-RUNX1T1* and *CBFB-MYH11*. Using the published computational framework for mutation signature identification [http://www.mathworks.com/matlabcentral/fileexchange/38724]. Only one mutational signature was identified, which is consistent with Signature 1B (Alexandrov et al, Nature, 2013) and was observed in other AMLs in the Alexandrov study.

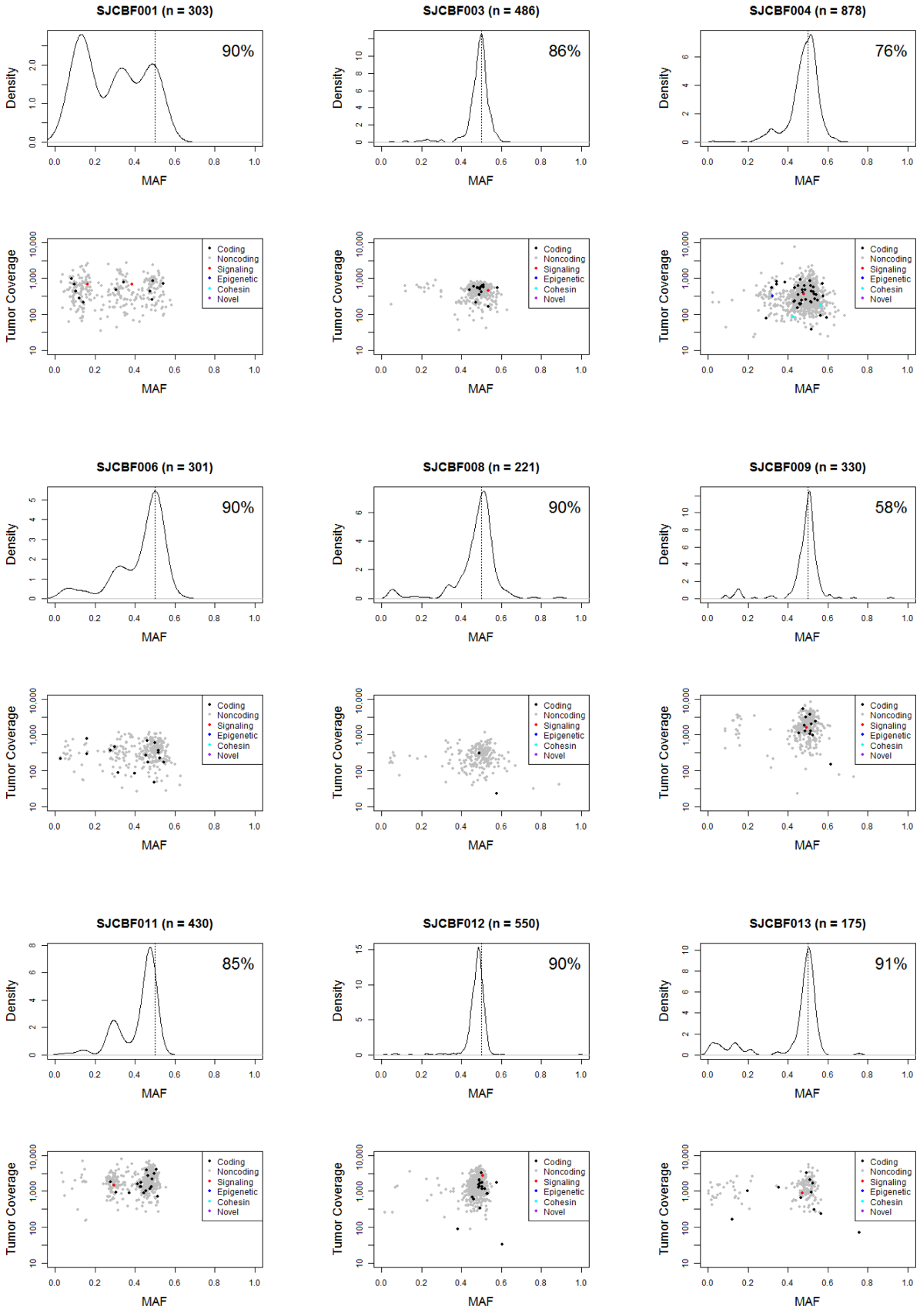
a)

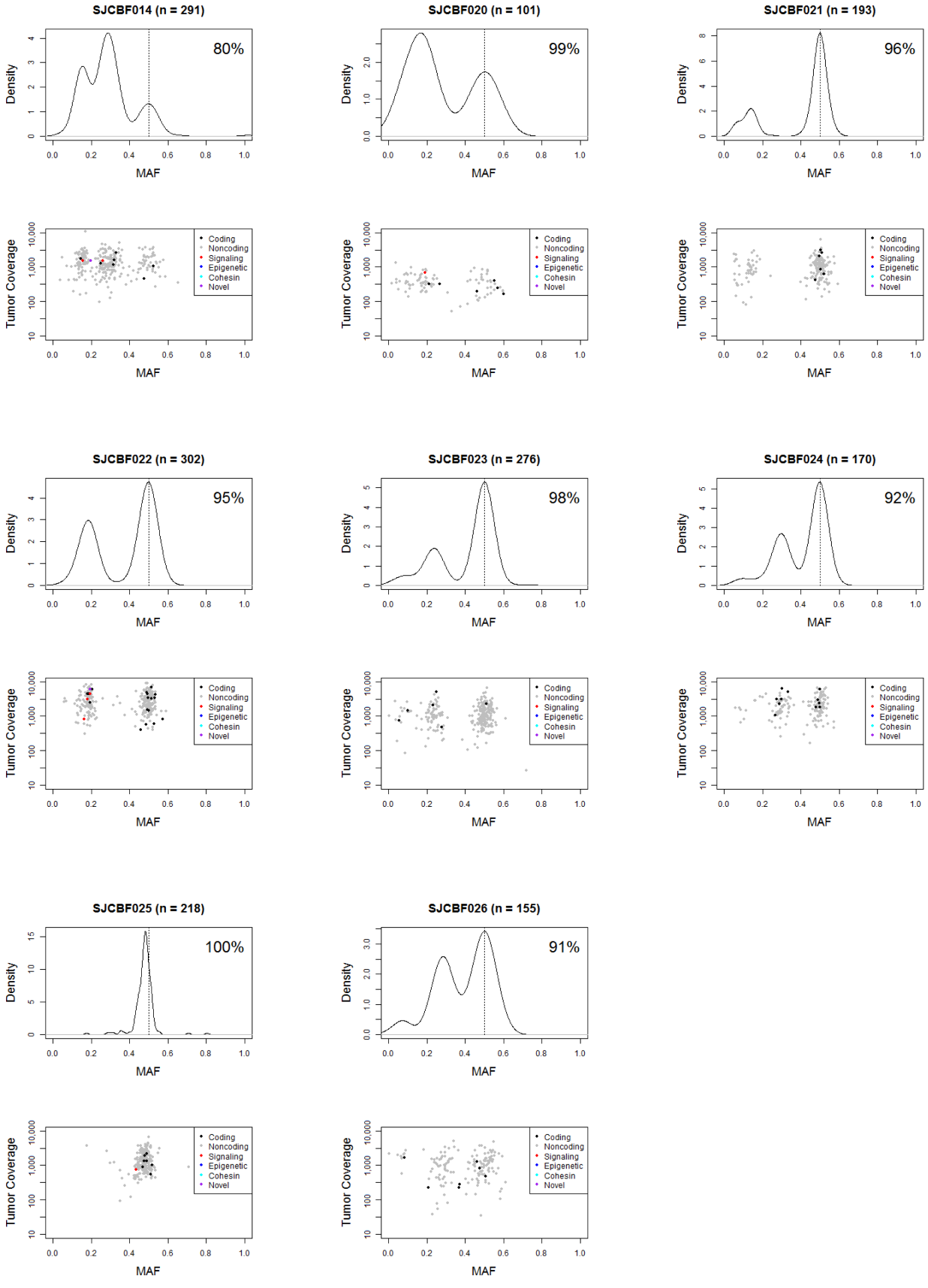


b)

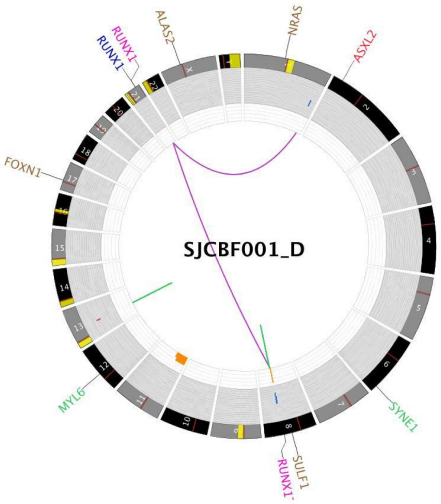


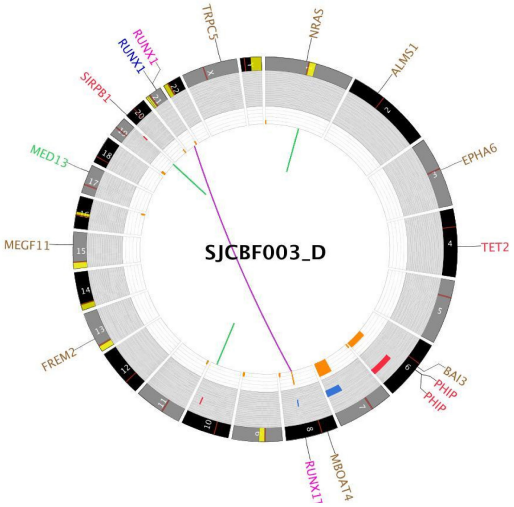
Supplementary Figure 4. Clonal architecture in CBF-AML. Mutant allele frequency (MAF) of each of the 17 cases that underwent whole genome sequencing is shown. Case I.D. and number of somatic variants identified by WGS (n=#) are shown above each histogram. Estimated tumor purity is also shown for each case (top right). Coverage at each variant is shown in the bottom plot. The class of each variant is indicated by the color (black – coding, grey – noncoding, red – signaling, blue – epigenetic, cyan – cohesin, and purple – novel). The signaling, epigenetic, cohesin, and novel gene groups as the same as those shown in Figure 1.

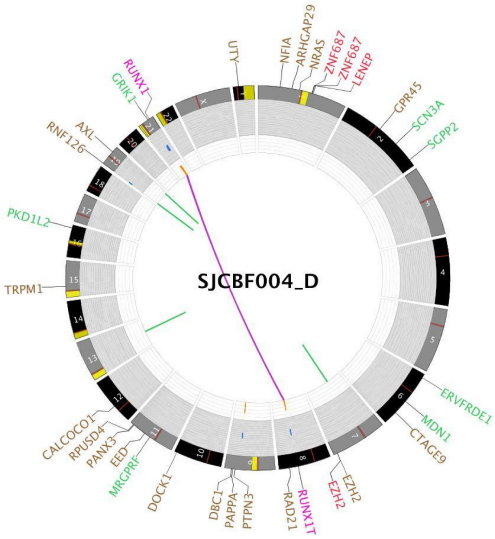

SJCBF001 (n = 303)
90%
SJCBF003 (n = 486)
86%
SJCBF004 (n = 878)
76%
SJCBF006 (n = 301)
90%
SJCBF008 (n = 221)
90%
SJCBF009 (n = 330)
58%
SJCBF011 (n = 430)
85%
SJCBF012 (n = 550)
90%
SJCBF013 (n = 175)
91%
Density
MAF
Tumor Coverage
Coding
Noncoding
Signaling
Epigenetic
Cohesin
Novel

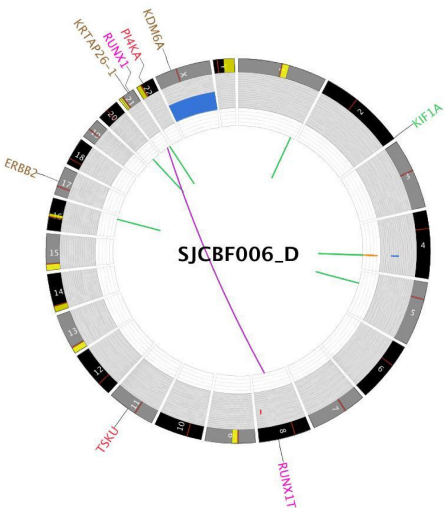

SJCBF014 (n = 291)
80%
SJCBF020 (n = 101)
99%
SJCBF021 (n = 193)
96%
SJCBF022 (n = 302)
95%
SJCBF023 (n = 276)
98%
SJCBF024 (n = 170)
92%
SJCBF025 (n = 218)
100%
SJCBF026 (n = 155)
91%
Density
MAF
Tumor Coverage
Coding
Noncoding
Signaling
Epigenetic
Cohesin
Novel

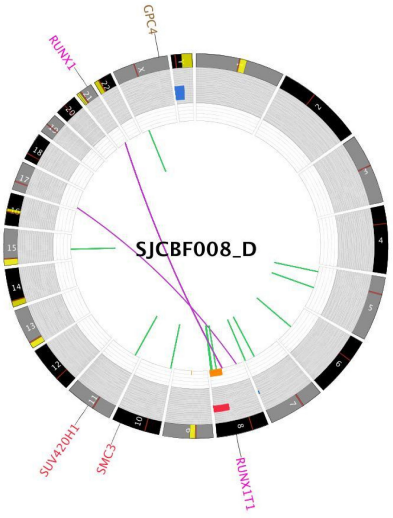
Supplementary Figure 5. Circos plots of genetic alterations in CBF-AML. Loss-of-heterozygosity structural variations, copy number variations, and mutation data from the 17 WGS cases.

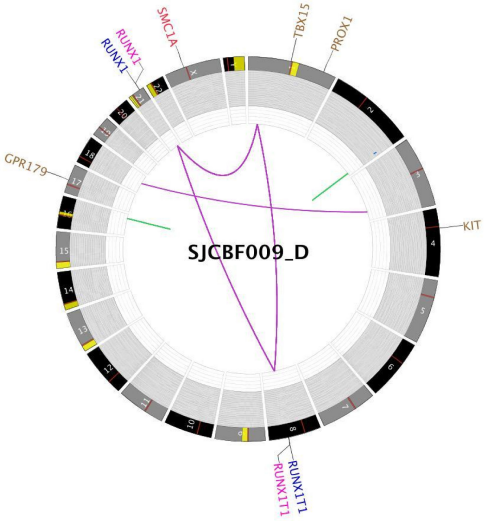

SJCBF001_D
NRAS
ASXL2
SYNE1
RUNX1
MYL6
FOXN1
RUNX1
RUNX1
ALAS2

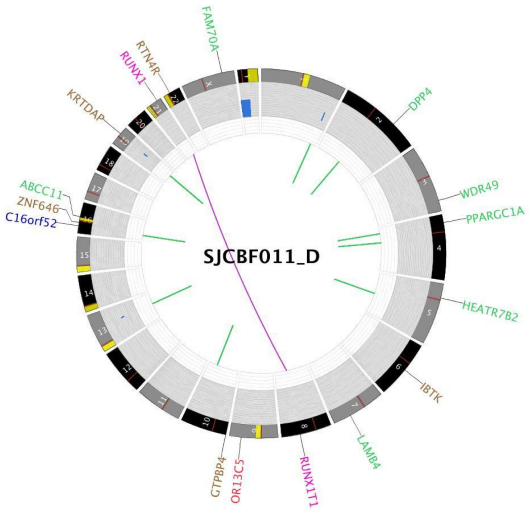

SJCBF003_D
TRPC5
NRAS
ALMS1
EPHA6
TET2
BAI3
PHIP
PHIP
MBOAT4
RUNX1T1
FREM2
MEGF11
MED13
SIRPB1
RUNX1
RUNX1

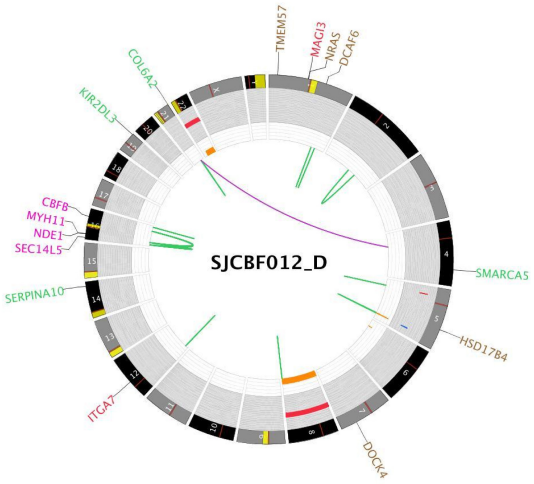

SJCBF004_D
UTY
NFIA
ARHGAP29
NRAS
ZNF687
ZNF687
LENEP
GPR45
SCN3A
SGPP2
ERVFRDE1
MDN1
CTAGE9
EZH2
RUNX1T1
RAD21
PTPN3
PAPPA
DBC1
DOCK1
MRGPRF
EED
PANX3
RPUSD4
CALCOCO1
PKD1L2
TRPM1
RNF126
AXL
GRIK1
RUNX1


SJCBF006_D
KDM6A
PI4KA
RUNX1
KRTAP26-1
KIF1A
ERBB2
RUNX1T1
TSKU

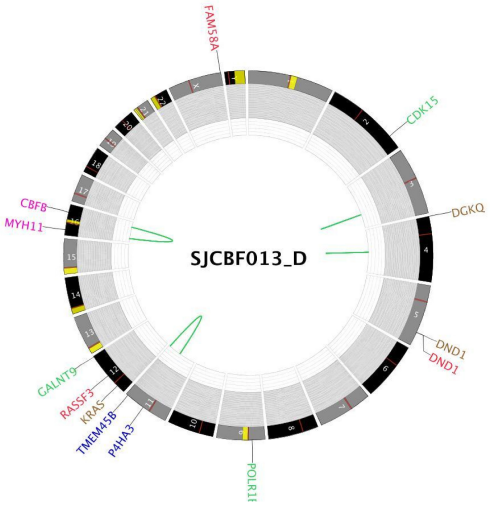

SJCBF008_D
GPC4
RUNX1
RUNX1T1
SMC3
SUV420H1

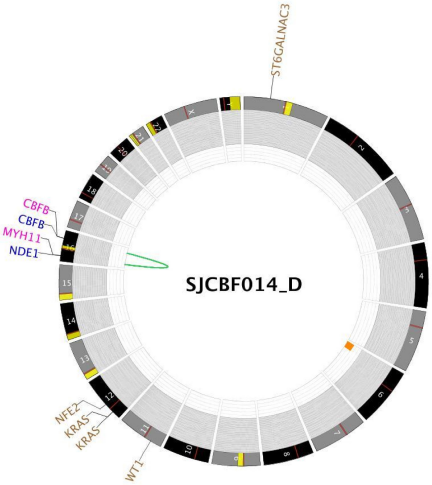

SJCBF009_D
SMC1A
TBX15
PROX1
RUNX1
RUNX1
KIT
GPR179
RUNX1T1
RUNX1T1

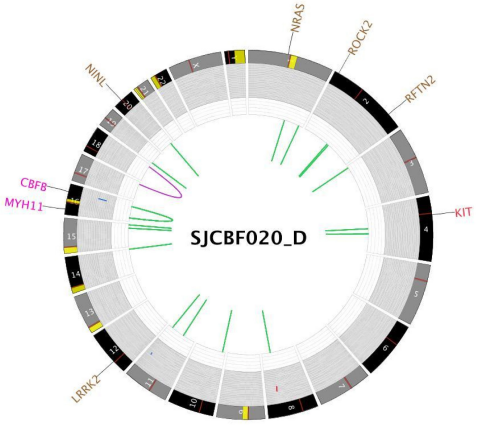


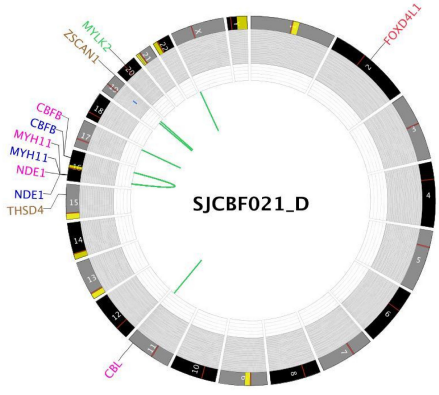


*labels for gene disrupting SVs removed

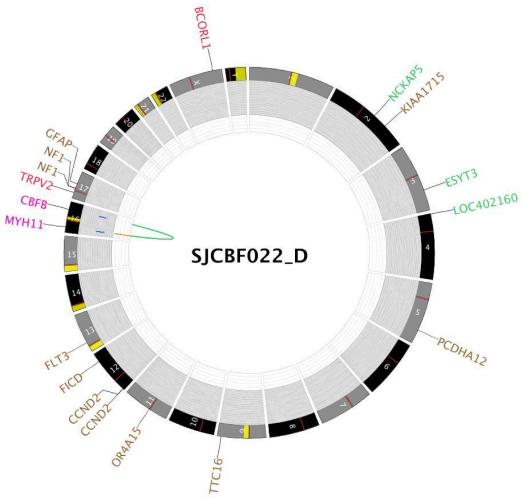


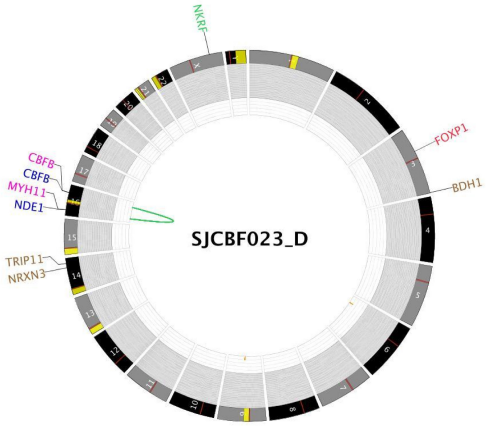


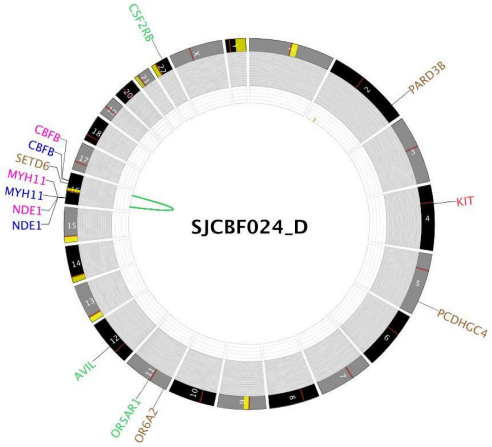


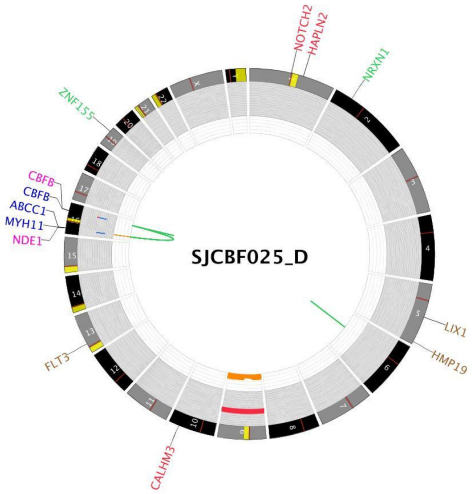


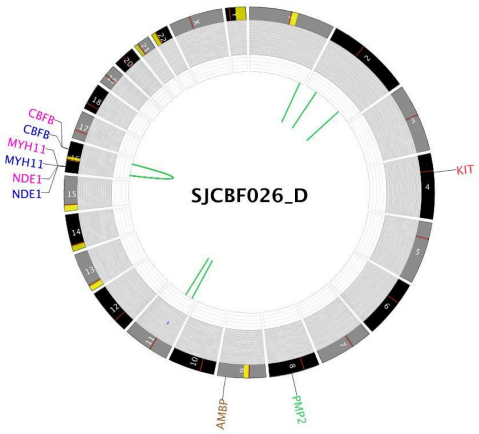
*labels for gene disrupting SVs removed

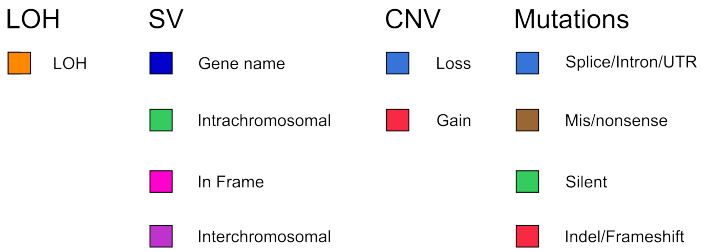

SJCBF022_D
BCORL1
NCKAP5
KIAA1715
ESYT3
LOC402160
PCDHA12
TTC16
OR4A15
CCND2
CCND2
FICD
FLT3
CBFB
MYH11
TRPV2
NF1
NF1
GFAP


SJCBF023_D
NKRF
FOXP1
BDH1
NDE1
MYH11
CBFB
CBFB
TRIP11
NRXN3


SJCBF024_D
CSF2RB
PARD3B
KIT
PCDHGC4
OR6A2
OR5AR1
AVIL
NDE1
NDE1
MYH11
MYH11
SETD6
CBFB
CBFB

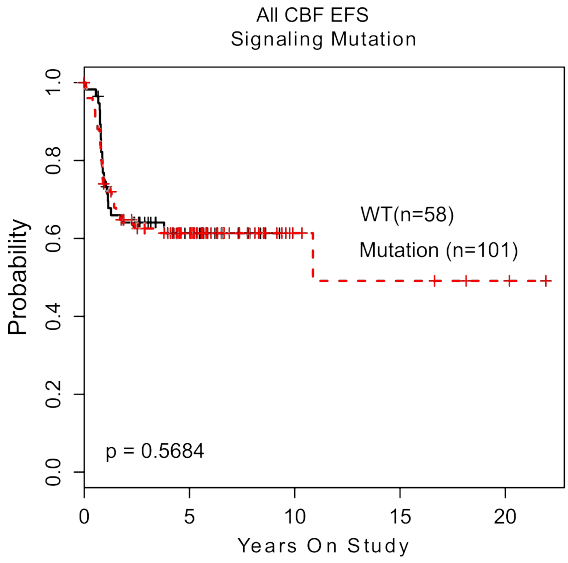

SJCBF025_D
NOTCH2
HAPLN2
NRXN1
LIX1
HMP19
CALHM3
FLT3
NDE1
MYH11
ABCC1
CBFB
CBFB
ZNF155

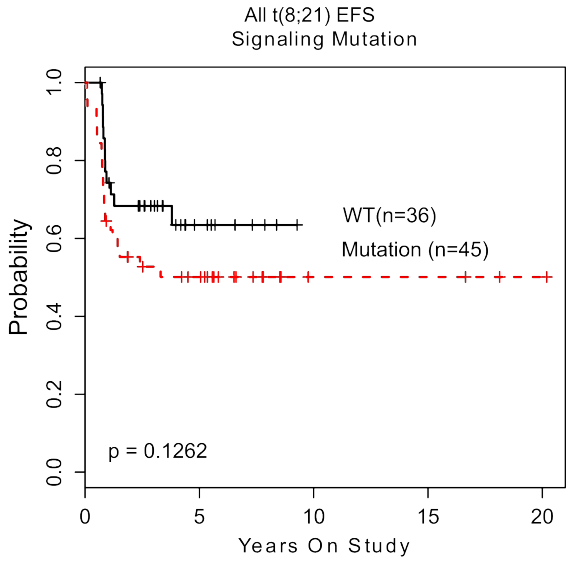

SJCBF026_D
KIT
PMP2
AMBP
NDE1
NDE1
MYH11
MYH11
CBFB
CBFB

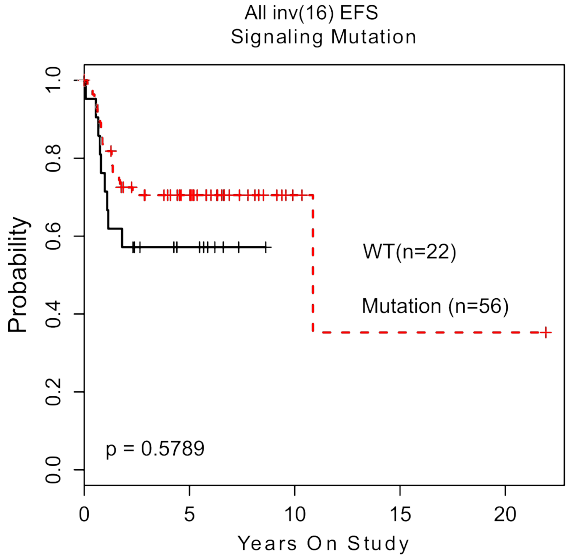

LOH
LOH
SV
Gene name
Intrachromosomal
In Frame
Interchromosomal
CNV
Loss
Gain
Mutations
Splice/Intron/UTR
Mis/nonsense
Silent
Indel/Frameshift

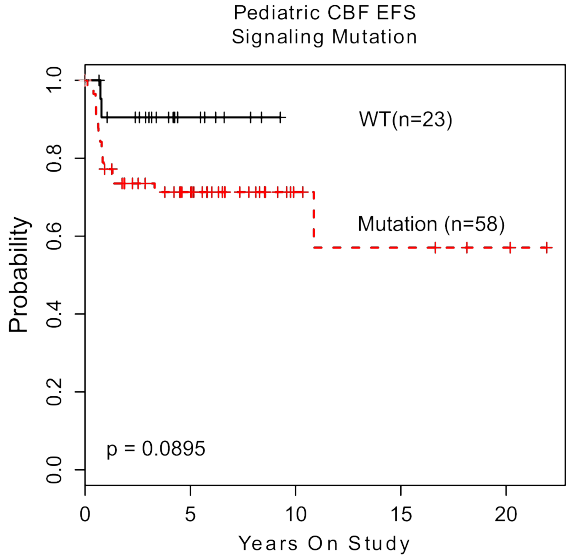
Supplementary Figure 6. Impact of signaling mutations on outcome in CBF-AML. Outcome analysis for CBF-AML based on signaling mutations. The genes in the signaling group are: *KRAS, NRAS, FLT3, KIT, NF1,* and *PTPN11.* a) Event free survival. b) Overall survival. P values correspond to log rank tests. Association of indicated mutations with c) event free survival (EFS) or d) overall survival (OS). e) Multivariable Cox Model. HR – hazard ratio, HR95CI – 95% confidence interval for the hazard ratio.

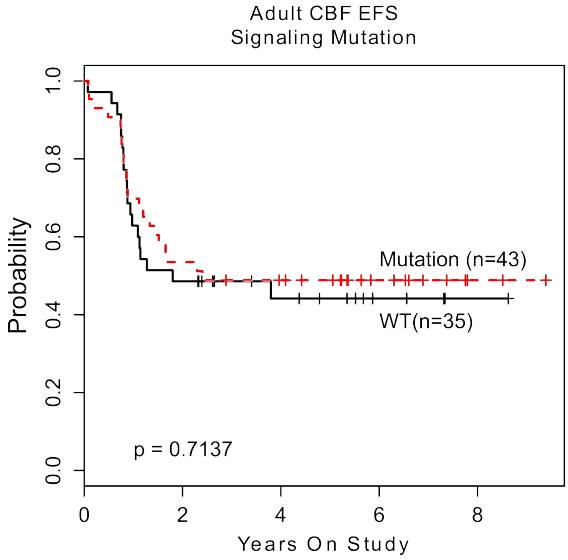
a)


All CBF EFS
Signaling Mutation
Probability
Years On Study
WT(n=58)
Mutation (n=101)
p = 0.5684

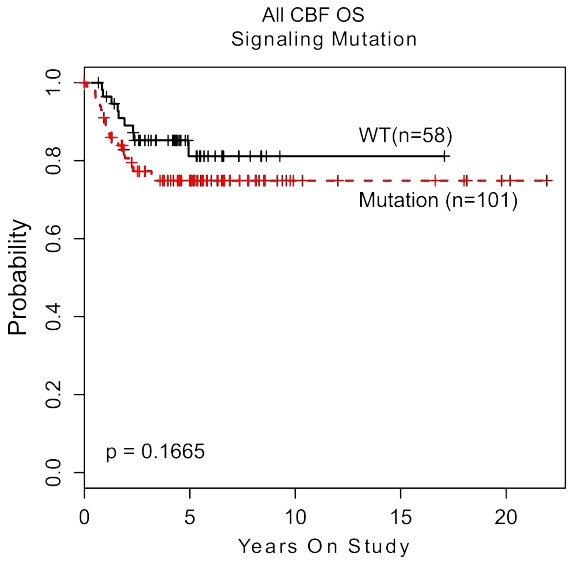

All t(8;21) EFS
Signaling Mutation
Probability
Years On Study
WT(n=36)
Mutation (n=45)
p = 0.1262

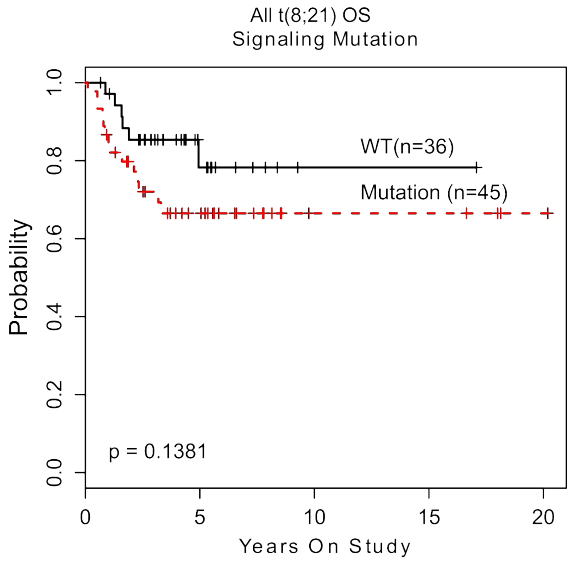

All inv(16) EFS
Signaling Mutation
Probability
Years On Study
WT(n=22)
Mutation (n=56)
p = 0.5789

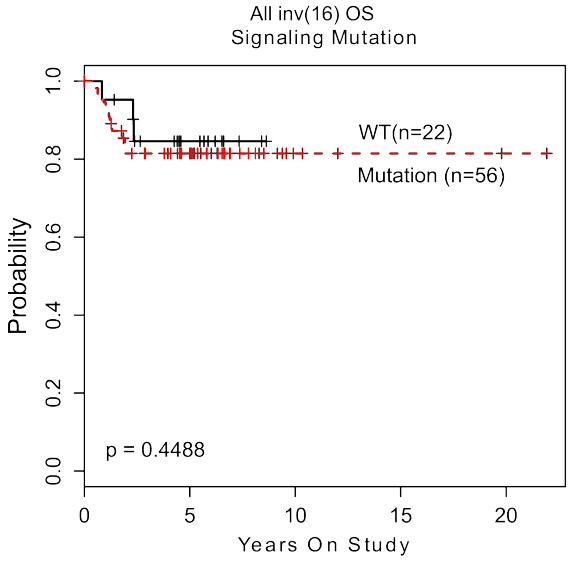

Pediatric CBF EFS
Signaling Mutation
Probability
Years On Study
WT(n=23)
Mutation (n=58)
p = 0.0895

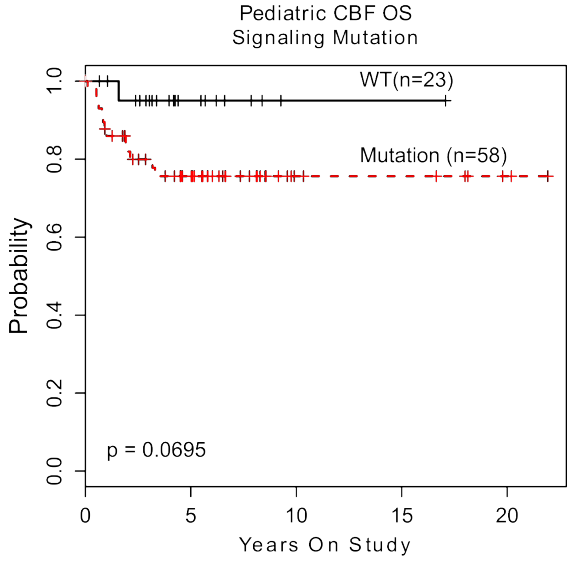

Adult CBF EFS
Signaling Mutation
Probability
Years On Study
Mutation (n=43)
WT(n=35)
p = 0.7137

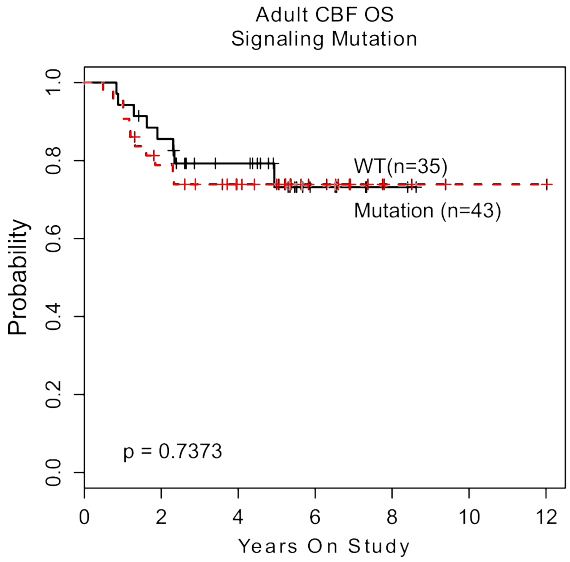
b)


All CBF OS
Signaling Mutation
WT(n=58)
Mutation (n=101)
p = 0.1665
Probability
Years On Study


All t(8;21) OS
Signaling Mutation
WT(n=36)
Mutation (n=45)
p = 0.1381
Probability
Years On Study


All inv(16) OS
Signaling Mutation
WT(n=22)
Mutation (n=56)
p = 0.4488
Probability
Years On Study


Pediatric CBF OS
Signaling Mutation
WT(n=23)
Mutation (n=58)
p = 0.0695
Probability
Years On Study


Adult CBF OS
Signaling Mutation
WT(n=35)
Mutation (n=43)
p = 0.7373
Probability
Years On Study

c)

Association of mutation with EFS in CBF-AML

Mutation	HR	HR95CI	P value	Log Rank P value
NRAS	0.8731	0.4954-1.5387	0.6388	0.6380
KIT	1.5541	0.8608-2.8058	0.1436	0.1388
KIT (Exon 17)	2.3150	1.2001-4.4657	0.0123	0.0097
Any Signaling	1.1685	0.6823-2.0012	0.5704	0.5684

d)

Association of mutation with OS in CBF-AML

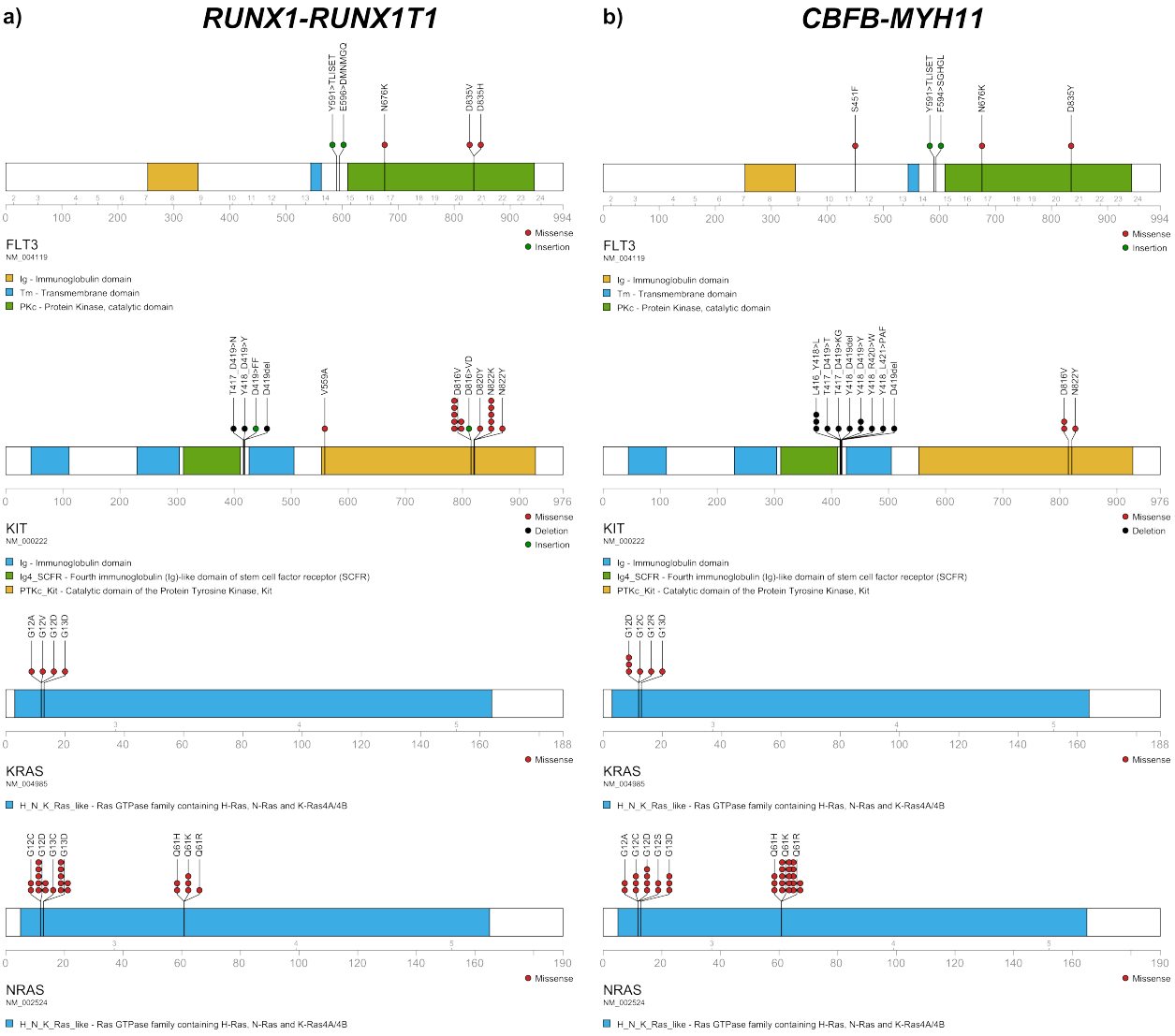
Mutation	HR	HR95CI	P value	Log Rank P value
NRAS	0.9953	0.4685-2.1146	0.9903	0.9903
KIT	1.6143	0.7482-3.4831	0.2223	0.2179
KIT (Exon 17)	2.5533	1.1001-5.9259	0.0291	0.0237
Any Signaling	1.7185	0.7911-3.7329	0.1713	0.1665

e)

Multivariable Cox Model

	Predictor	HR	HR95CI	P value
EFS	*KIT* (Exon 17)	2.271	1.1514-4.4793	0.0179
	RUNX1-RUNX1T1	1.1693	0.6893-1.9835	0.562
	Pediatric	0.4016	0.2304-0.6998	0.0013
OS	*KIT* (Exon 17)	2.4096	1.0123-5.7358	0.0469
	RUNX1-RUNX1T1	1.3684	0.6666-2.8093	0.3927
	Pediatric	0.7912	0.3946-1.5864	0.5094

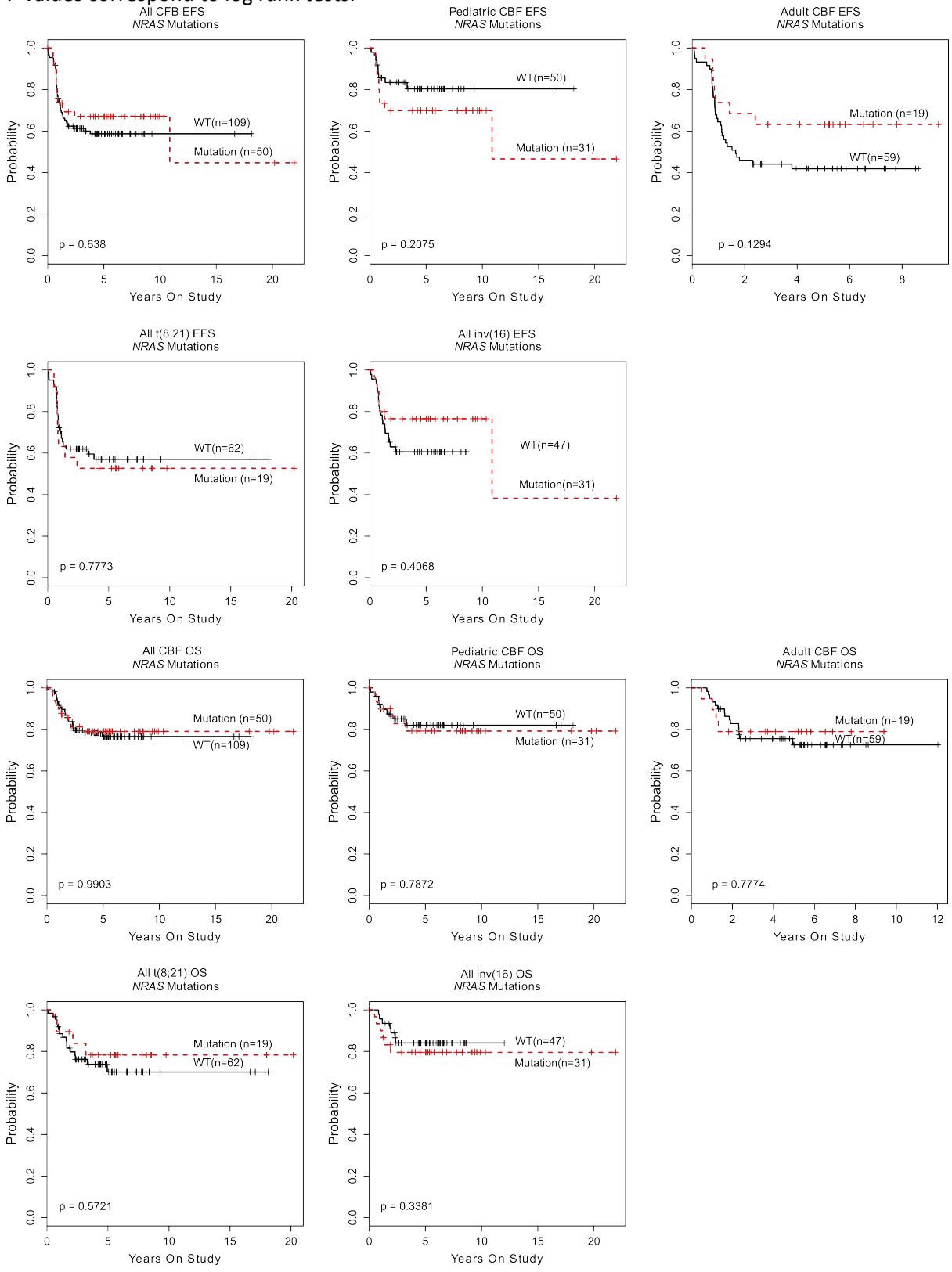
Supplementary Figure 7. Distribution of signaling mutations in CBF-AML subgroups. Graphic representation of mutations in signaling genes in a) *RUNX1-RUNX1T1* and b) *CBFB-MYH11* CBF AML. c) Enrichment of mutations in each fusion cohort. P values calculated using Fisher's exact test.



c)

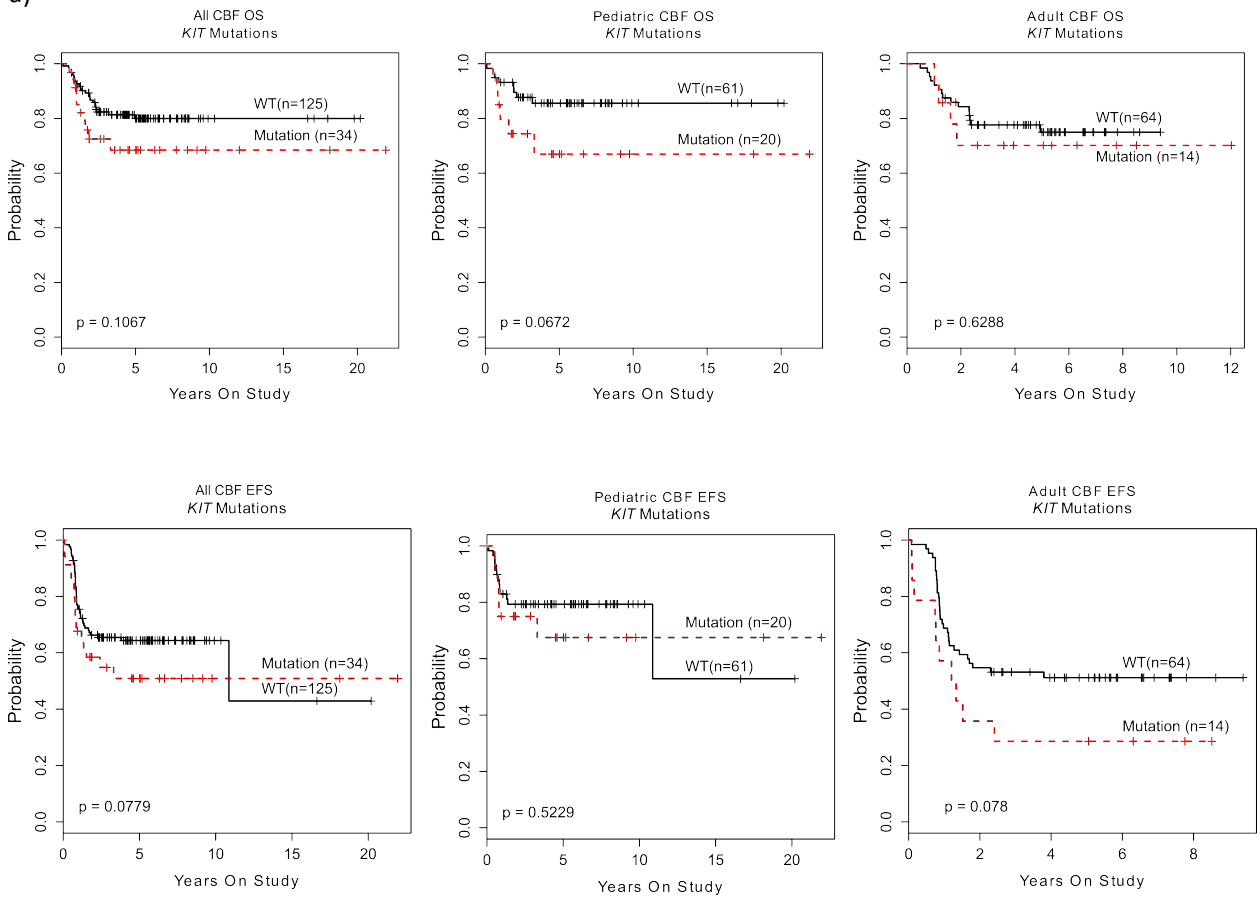
Gene/Mutation	*RUNX1-RUNX1T1*	*CBFB-MYH11*	p
FLT3	5	5	1.00
FLT3-ITD	2	2	0.99
FLT3-TKD	3	2	0.9871
FLT3 other	0	1	0.7869
KRAS	5	5	1.00
NRAS (total)	22	34	0.032
NRAS (G12/G13)	16	14	0.843
NRAS (Q61)	6	20	0.0023
KIT (total)	23	15	0.267
KIT (exon 8)	4	11	0.057
KIT (exon 17)	15	3	0.0051

Supplementary Figure 8. Impact of *NRAS* mutations on outcome in CBF-AML. Mutations in *NRAS* had no effect on outcome in pediatric or adult patients in either *RUNX1-RUNX1T1* or *CBFB-MYH11* leukemia. P values correspond to log rank tests.

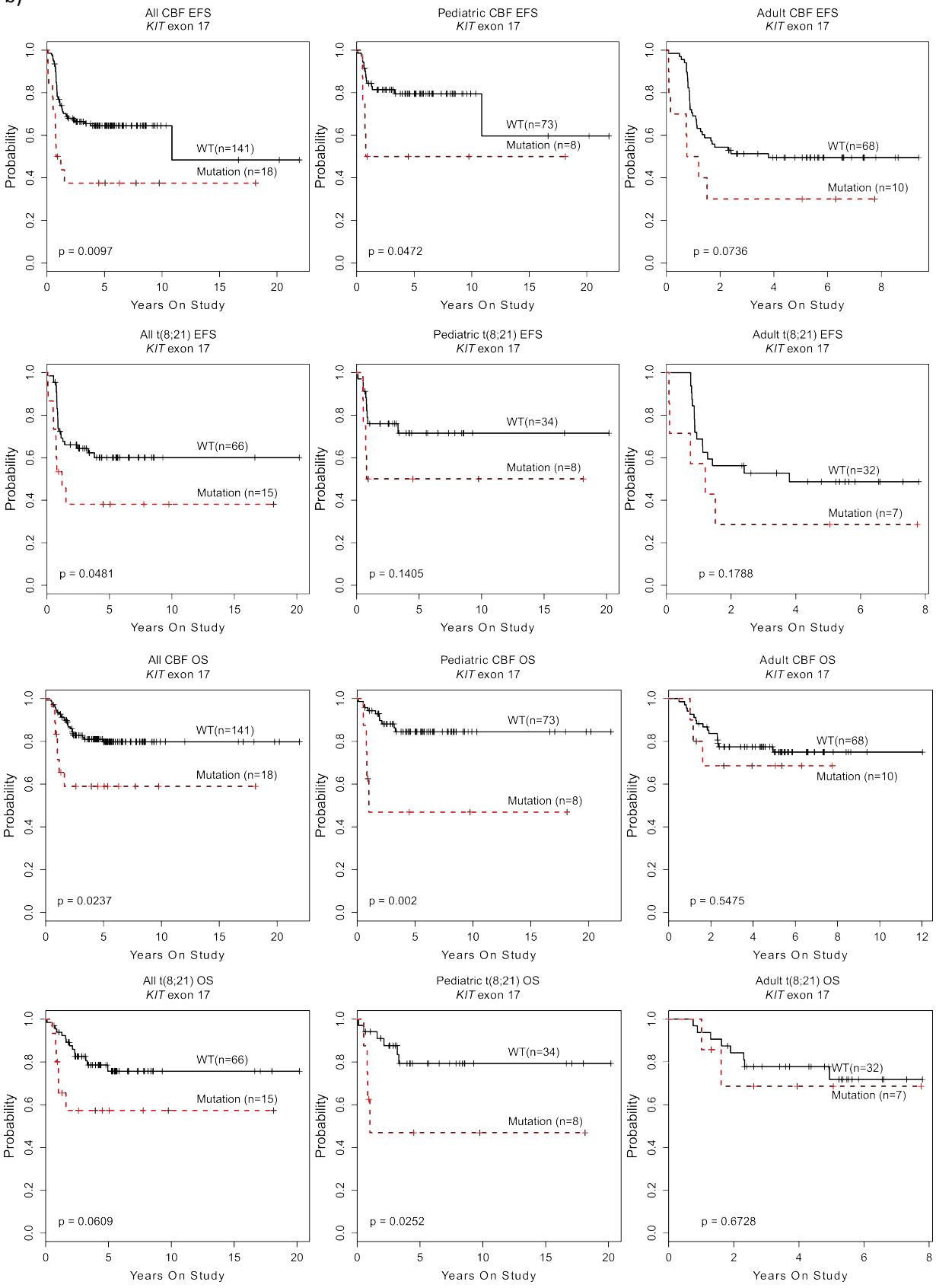


Supplementary Figure 9. Impact of KIT mutations on outcome in CBF-AML.
a) Outcome analysis of CBF patients in our cohort based on the presence of any *KIT* mutation. b) Patients with an exon 17 *KIT* mutation have a worse outcome. P values correspond to log rank tests.

a)

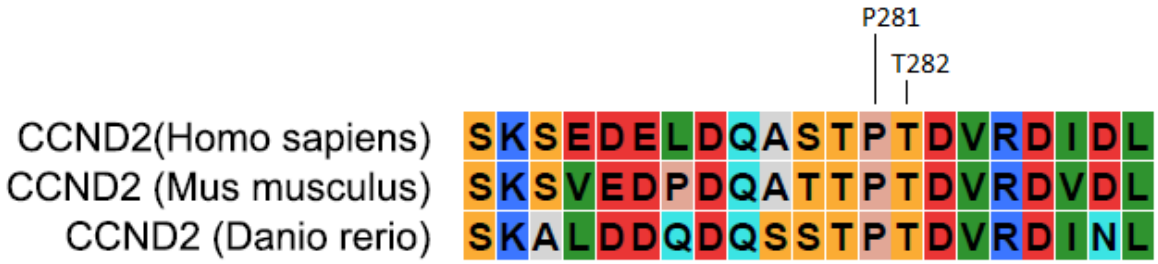


b)

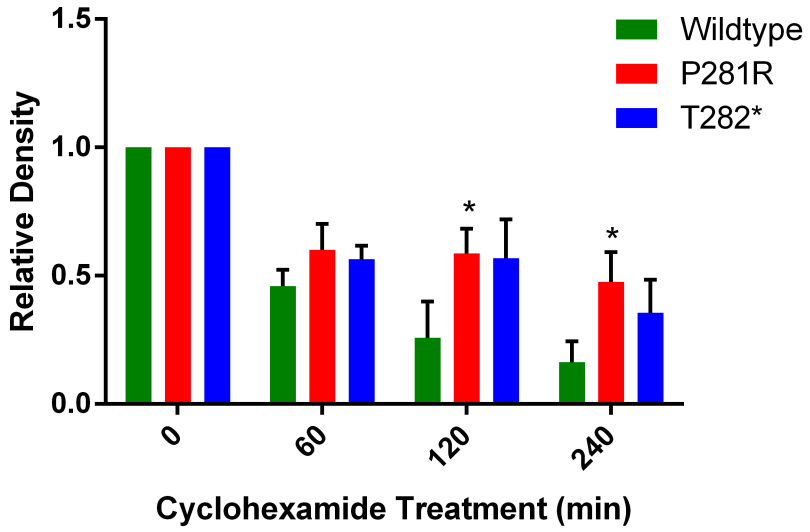

All CBF EFS
KIT exon 17
WT(n=141)
Mutation (n=18)
p = 0.0097
Pediatric CBF EFS
KIT exon 17
WT(n=73)
Mutation (n=8)
p = 0.0472
Adult CBF EFS
KIT exon 17
WT(n=68)
Mutation (n=10)
p = 0.0736
All t(8;21) EFS
KIT exon 17
WT(n=66)
Mutation (n=15)
p = 0.0481
Pediatric t(8;21) EFS
KIT exon 17
WT(n=34)
Mutation (n=8)
p = 0.1405
Adult t(8;21) EFS
KIT exon 17
WT(n=32)
Mutation (n=7)
p = 0.1788
All CBF OS
KIT exon 17
WT(n=141)
Mutation (n=18)
p = 0.0237
Pediatric CBF OS
KIT exon 17
WT(n=73)
Mutation (n=8)
p = 0.002
Adult CBF OS
KIT exon 17
WT(n=68)
Mutation (n=10)
p = 0.5475
All t(8;21) OS
KIT exon 17
WT(n=66)
Mutation (n=15)
p = 0.0609
Pediatric t(8;21) OS
KIT exon 17
WT(n=34)
Mutation (n=8)
p = 0.0252
Adult t(8;21) OS
KIT exon 17
WT(n=32)
Mutation (n=7)
p = 0.6728
Probability
Years On Study

Supplementary Figure 10. CCND2 stabilization. (a) Alignment of CCND2 amino acid sequence across human, mouse, and zebrafish highlights the sequence conservation near residue T280. (b) Quantification of CCND2 levels. Mean + s.d. shown (n=3). * P< 0.05, two-tailed unpaired Student's t-test (P281R vs WT 120 minutes – p=0.029, t=3.334, df=4, P281R vs WT 240 minutes – p=0.0185, t=3.8362, df=4). (c) Full size western blots for CCND2 stabilization.

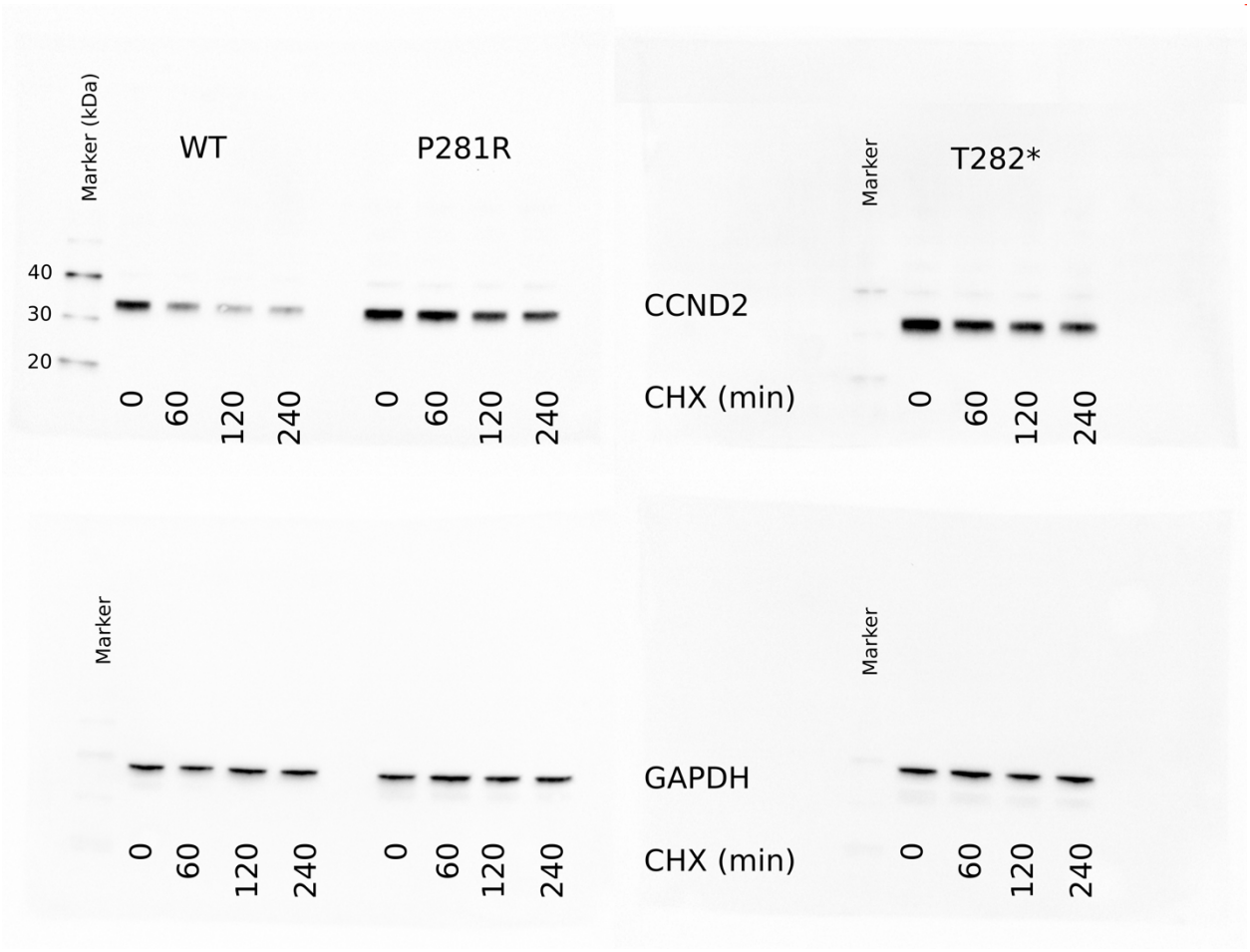
a)



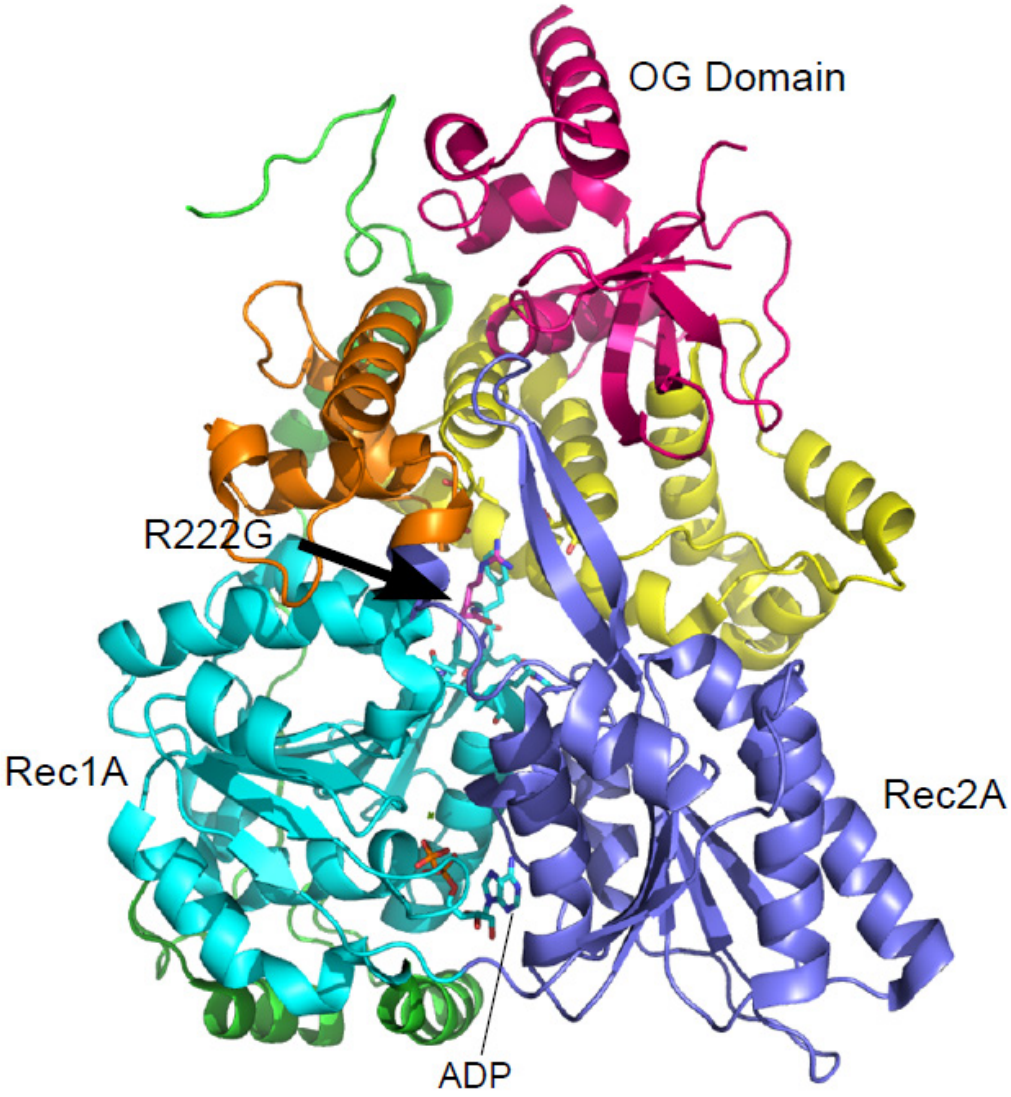
b)



c)

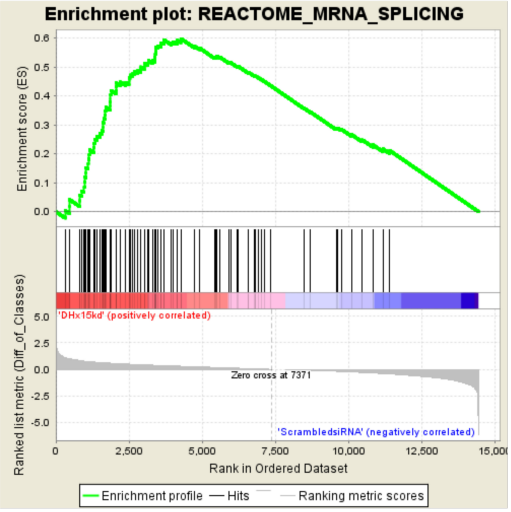

Marker (kDa)
WT
P281R
Marker
T282*
40
30
20
CCND2
CHX (min)
0
60
120
240
Marker
Marker
GAPDH
CHX (min)

Supplementary Figure 11. DHX15 p. R222G mutation in CBF-AML. A structural model of yeast homologue of DHX15 (Prp43:PDB 2XAU) with the location of the R222G mutation highlighted by a black arrow. Other domains are colored as follows: N-terminal, green; Rec1A, teal; Rec2A, blue; WH, orange; ratchet, yellow; c-terminal/OG, red. The protein is shown with bound ADP.



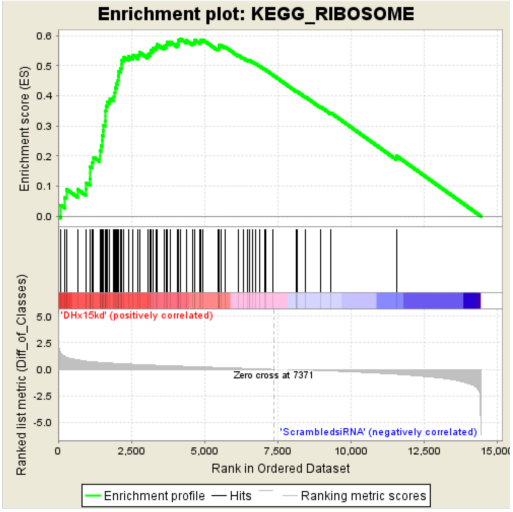
Supplementary Figure 12. Gene Set Enrichment Analysis of DHX15 depleted cells. siRNA-mediated depletion of DHX15 leads to differential expression of genes involved in a) mRNA splicing and b) ribosome biogenesis.

a)



NES: 2.609
FDR q-value: 0.0

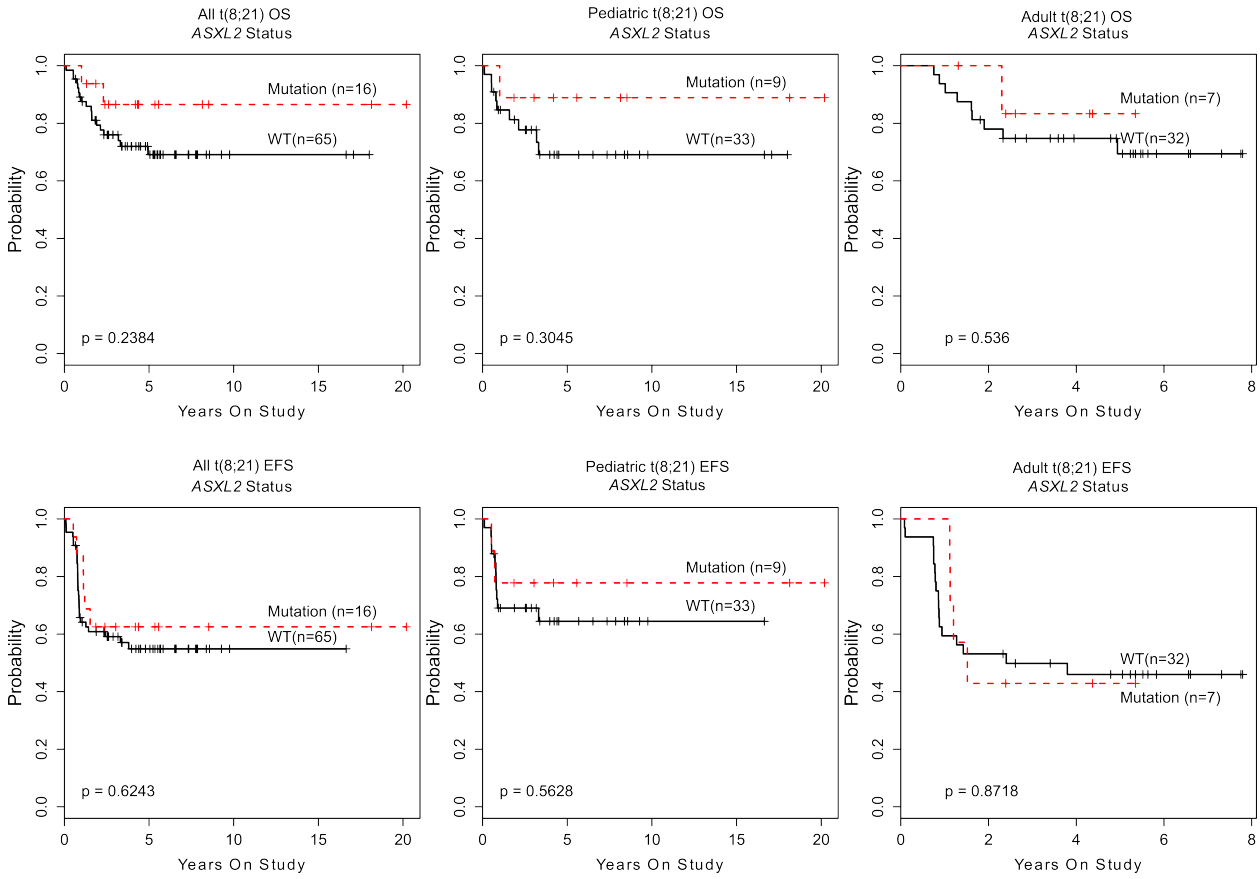
b)



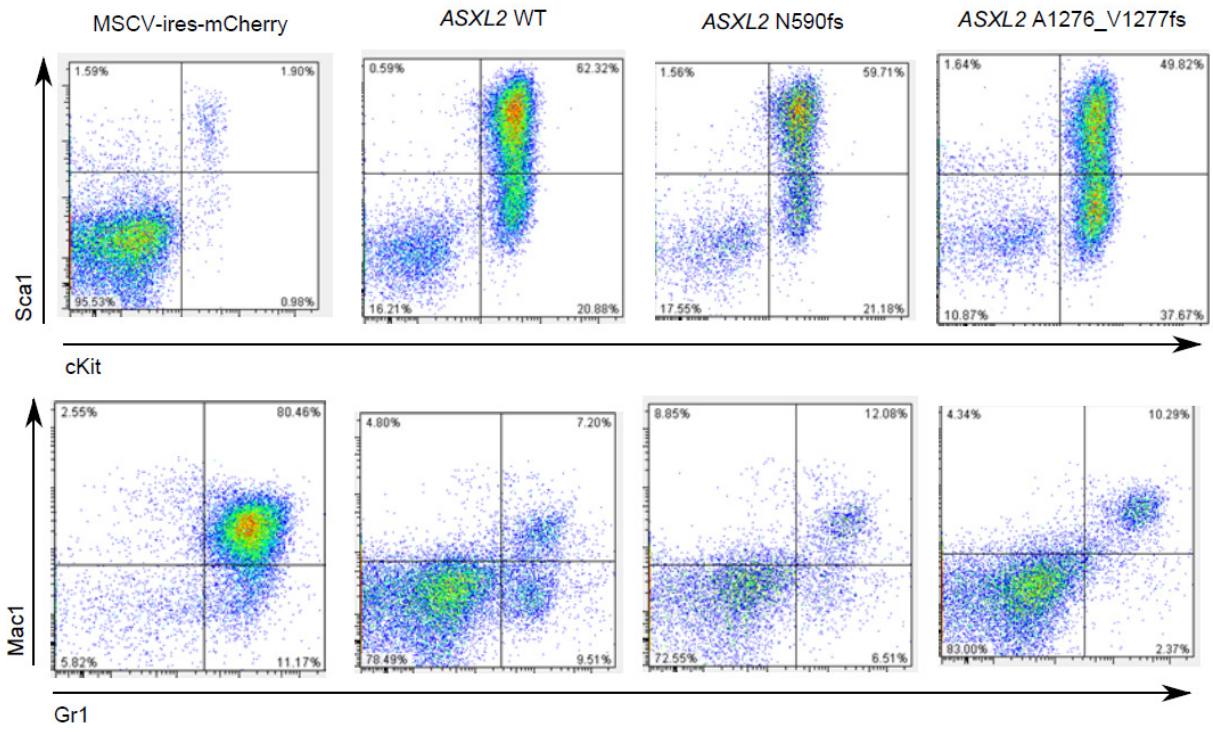
NES: 2.498
FDR q-value: 0.0

Supplementary Figure 13. Mutations in epigenetic regulators found in CBF-AML. Graphical representation of the mutations found in a) *CREBBP*, b) *KDM6A*, c) *EED*, d) *SETD2*, e) *KMT2D*, f) *KMT2C* g) *EZH2*, and h) *ASXL2*.

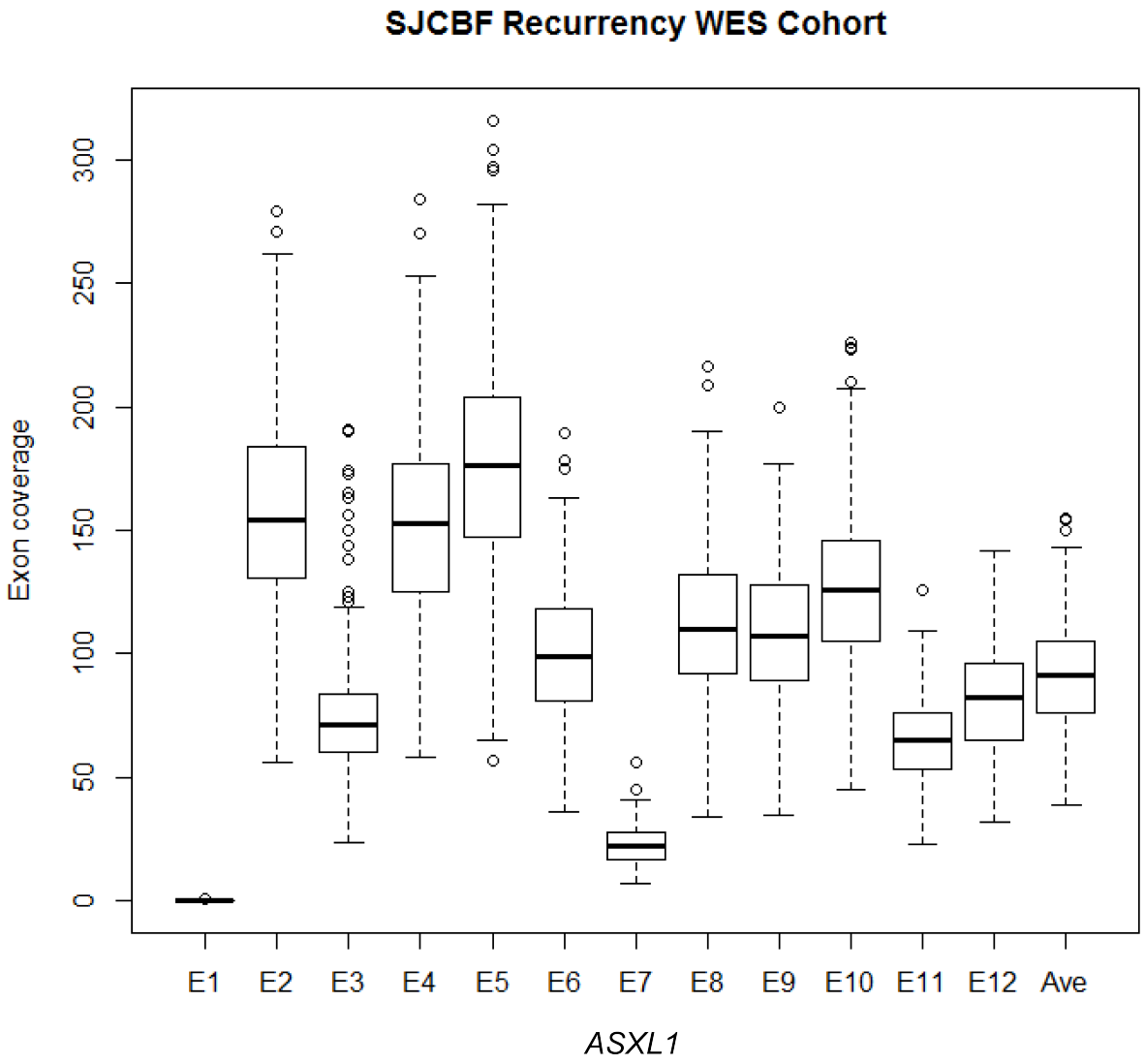
Supplementary Figure 14. Impact of *ASXL2* mutations on outcome in *RUNX1-RUNX1T1* rearranged CBF-AML. Mutations in *ASXL2* are not significantly correlated with outcome in CBF-AML. P values correspond to log rank tests.



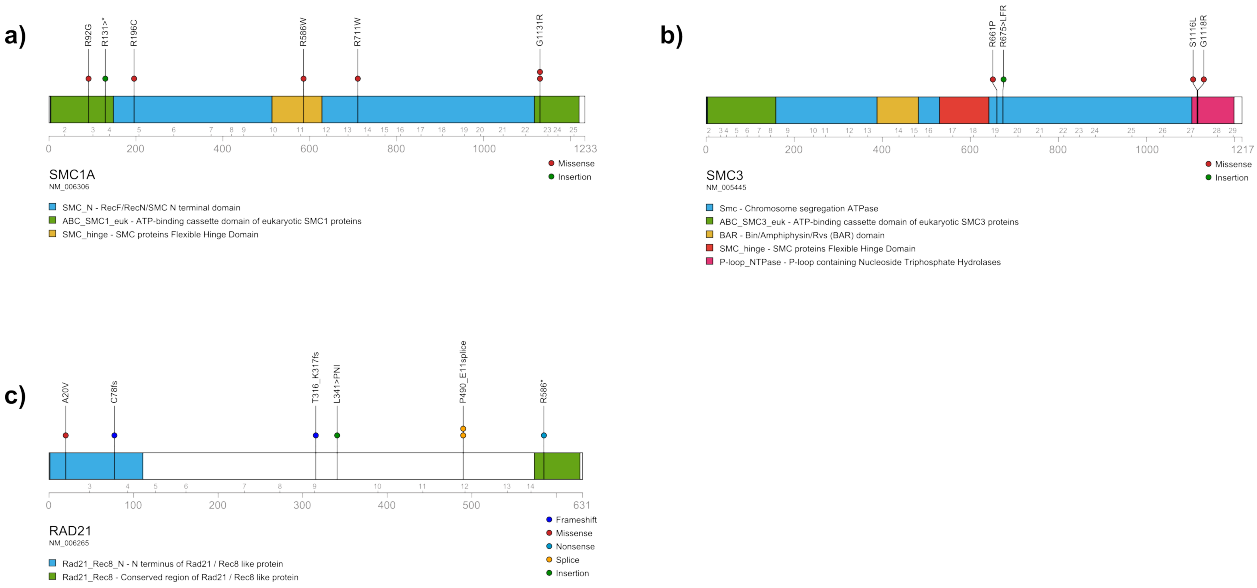
Supplementary Figure 15. *ASXL2* mutations alter myeloid differentiation. Lineage depleted mouse bone marrow cells were transduced with the indicated murine stem cell virus (MSCV) ires mCherry based vectors. After sorting for mCherry positive cells, cultures were grown for 14 days in cytokine-enriched media. Representative (of three independent experiments) day 14 FACS plots are shown indicating an increase in cKit/Sca1 double positive cells as well as an increase in Mac1/Gr1 double negative cells.



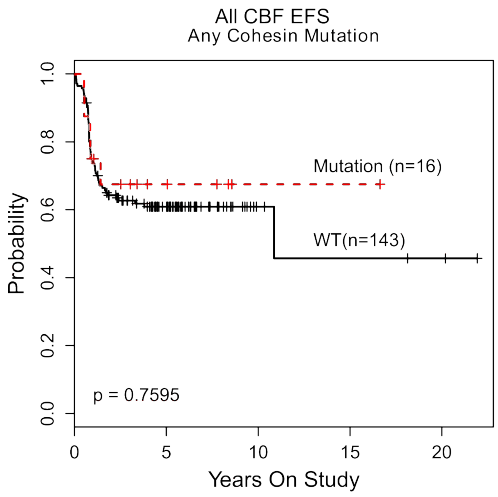
Supplementary Figure 16. Exon coverage of *ASXL1*. The coverage level of each exon of *ASXL1* is shown as well as average coverage over all exons for WES cases. The coverage in exon 12, which frequently harbors somatic mutations, is 80.7x. Exons 1 and 7 of *ASXL1* have high GC content (0.77 and 0.71, respectively), which explains their low coverage.

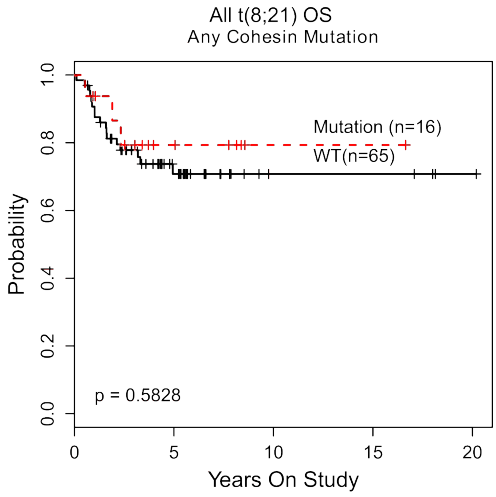


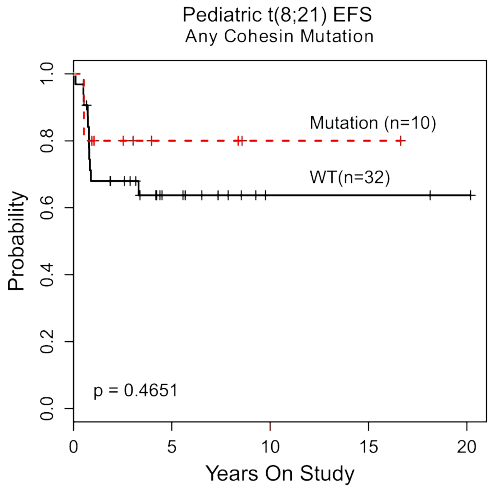
Supplementary Figure 17. Cohesin mutations in CBF-AML. Representative diagrams of the mutations found in a) *SMC1A*, b) *SMC3*, and c) *RAD21*.

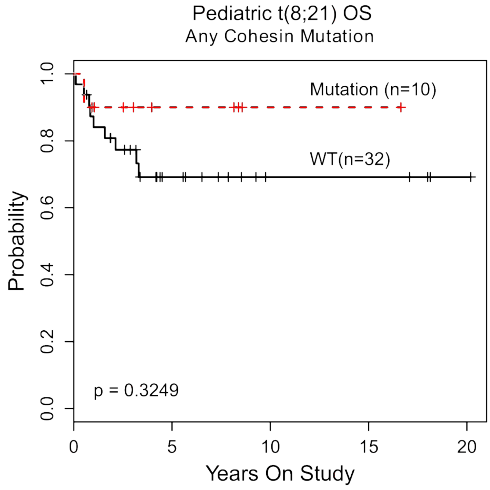


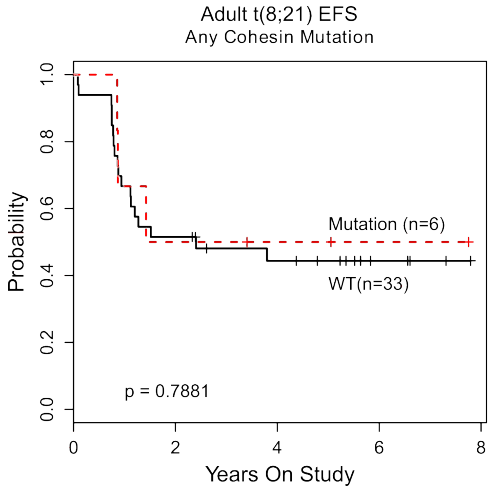
Supplementary Figure 18. Impact of cohesin mutations on outcome in CBF-AML. Mutations in *SMC1A, SMC3, or RAD21* had no effect on outcome in pediatric or adult patients in either *RUNX1-RUNX1T1* or *CBFB-MYH11* leukemia. P values correspond to log rank tests.

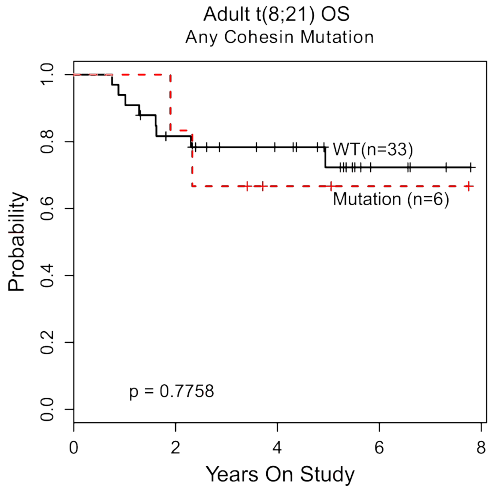






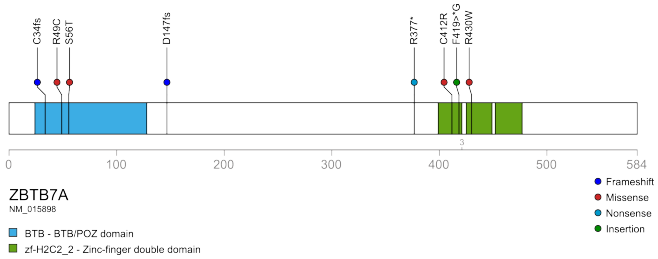






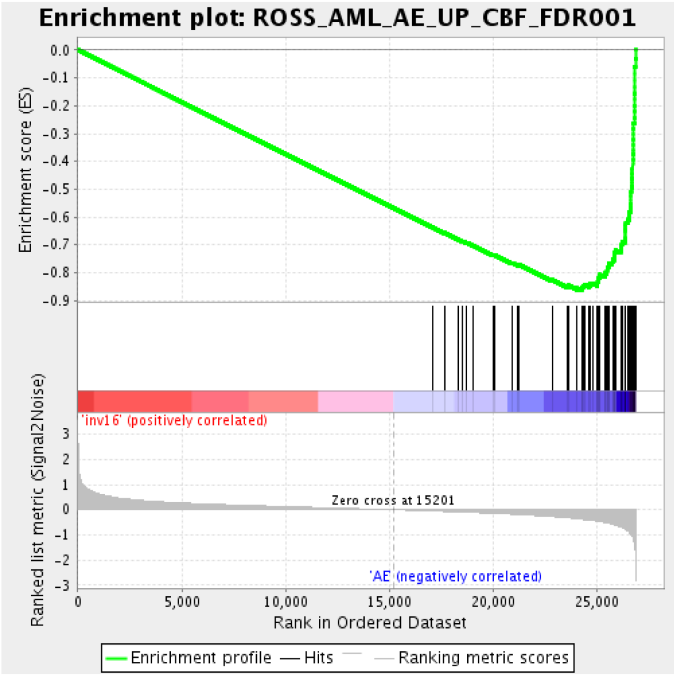
Supplementary Figure 19. *ZBTB7A* mutations in CBF AML. Graphical representation of the mutations found in *ZBTB7A*.

a)

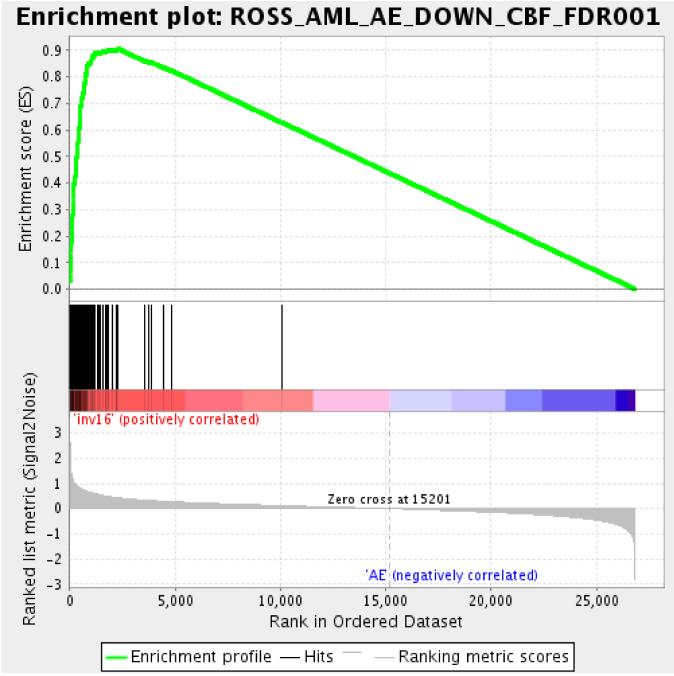


Supplementary Figure 20: Gene set enrichment analysis. Comparison of transcriptome analysis with previously published array data of CBF leukemia showed similar a) *RUNX1-RUNX1T1* and b) *CBFB-MYH11* gene signatures.

a)

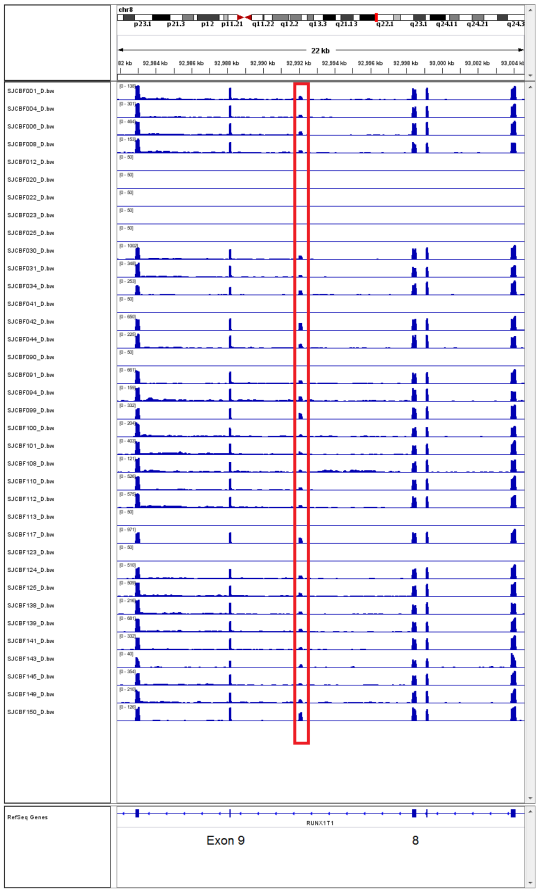


b)

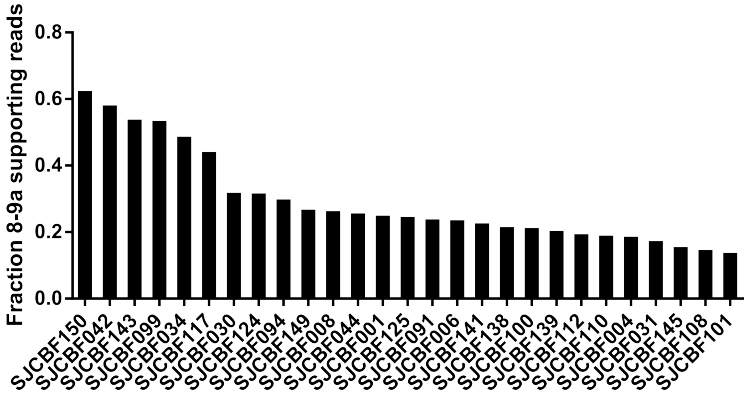


Supplementary Figure 21: The 9a isoform of *RUNX1T1* was detected in all cases of *RUNX1-RUNX1T1* CBF leukemia. The alternatively spliced 9a isoform could be detected in our transcriptome analysis. a) The location of the 9a exon is indicated by the red box. b) The fraction of reads supporting the 8-9a junction is shown for each case.

a)



b)



Supplementary Figure 22. Clonal evolution between diagnosis and relapse in CBF-AML. a) *De novo* tumor (red) and relapse (blue) MAFs are shown for each individual case. All validated variants are shown and were sequenced after PCR amplification to a median coverage of >25000x. b) Copy number analysis of the paired samples shows differences between the diagnosis and relapsed sample.

a)

SJCBF028

Mutant	MAF_D	MAF_R
TGM2_E13_UTR_3	0.45	0.32
UBE2W_E7_UTR_3	0.45	0.26
KRAS_G13D	0.42	0.28
KIT_N822K	0.16	0.00
EPS8L1_G537R	0.00	0.28
CDHR1_V713M	0.00	0.25
PTPRE_D488E	0.01	0.22
COL7A1_Q246*	0.00	0.21

SJCBF030

Mutant	MAF_D	MAF_R
ZBTB47_D277E	0.52	0.30
FAM81B_F23S	0.50	0.28
ASXL2_N590fs	0.49	0.27
CCND2_T282*	0.38	0.33
KIT_N822K	0.18	0.29
SPOCK1_E11_UTR_3	0.17	0.02
DDRGK1_E255D	0.15	0.00
KIT_Y418_D419>Y	0.11	0.01
RASSF2_E12_UTR_3	0.00	0.21
RUNX1_C81R	0.00	0.26

SJCBF043

Mutant	MAF_D	MAF_R
CHFR_G531S	0.45	0.40
EPHB4_Y467Y	0.52	0.41
DSG1_R26Q	0.44	0.39
CCND2_P281A	0.43	0.00
RUVBL1_E389D	0.38	0.00
NRAS_G12C	0.23	0.00
COL3A1_R1434C	0.00	0.39
PIP5K1A_G504V	0.00	0.36
EPHA6_A380S	0.00	0.35
WDR46_Y28*	0.00	0.31

SJCBF044

Mutant	MAF_D	MAF_R
CUL9_R1211C	0.52	0.42
DHX15_R222G	0.51	0.44
KMT2C_E33_splice	0.34	0.38
ATRX_E1_UTR_5	0.35	0.21
GRIK2_A657T	0.18	0.00
MGA_F907fs	0.11	0.00
SMC5_E25_UTR_3	0.00	0.50
KIT_D816Y	0.01	0.47
ASXL1_G710fs	0.00	0.22
DAG1_R371W	0.00	0.13

SJCBF078

Mutant	MAF_D	MAF_R
NXT2_E5_UTR_3	0.95	0.00
DHX15_R222G	0.43	0.00
NUDT16L1_E3_UTR_3	0.43	0.00
HTR1A_C109C	0.39	0.00

SJCBF155

Mutant	MAF_D	MAF_R
BMP7_R157L	0.53	0.58
ADAM18_N711K	0.46	0.51
ELOVL3_S217C	0.42	0.49
RPS6KA4_V295V	0.11	0.45
KIT_D812V	0.07	0.49
ABCB1_I1158F	0.22	0.00
SRD5A2_R168C	0.19	0.00
PDK4_E11_UTR_3	0.17	0.00
SH2B3_E8_UTR_3	0.00	0.48
TET2_F1104fs	0.00	0.36

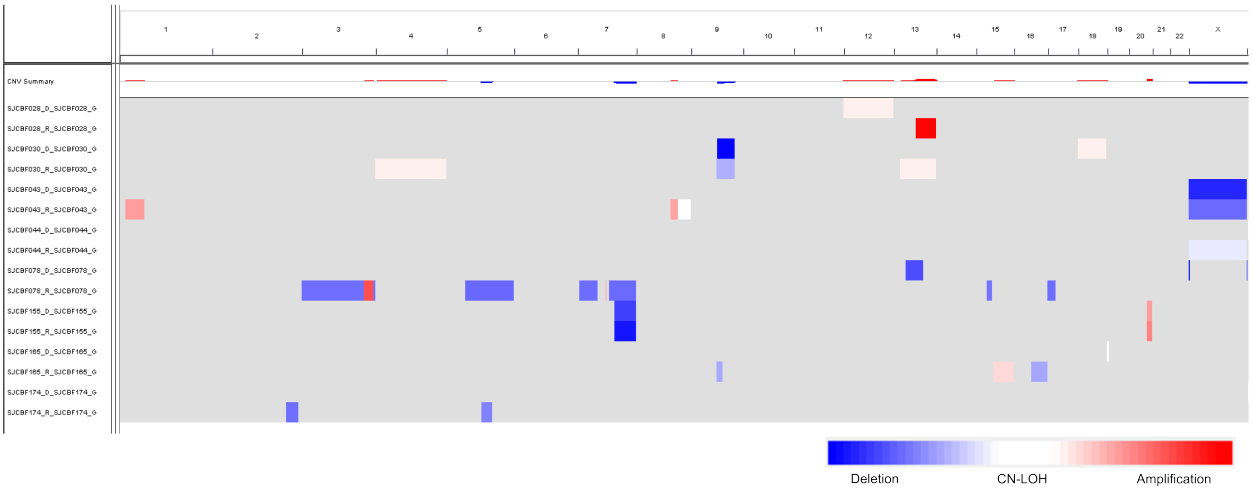
SJCBF165

Mutant	MAF_D	MAF_R
CCDC88A_D399V	0.45	0.33
PCDHB14_E137*	0.44	0.31
PLCE1_F839F	0.43	0.01
GLUL_E7_UTR_3	0.24	0.00
EXD3_E2_UTR_5	0.20	0.00
NEBL_G465V	0.19	0.00
KDM6A_T1323fs	0.02	0.00
RABEP1_E18_UTR_3	0.09	0.27
NMS_W130*	0.06	0.29
ZBTB7A_A205fs	0.07	0.00

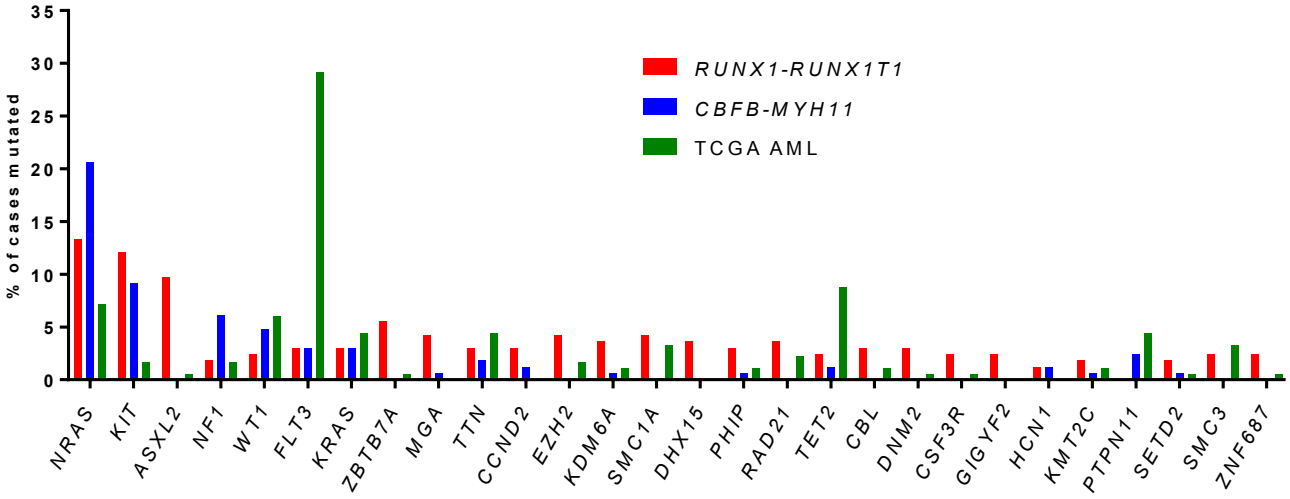
SJCBF174

Mutant	MAF_D	MAF_R
COL11A1_T905I	0.47	0.38
GUCY1A2_K516N	0.45	0.39
ATG2B_D1013N	0.44	0.38
PHIP_E40_UTR_3	0.43	0.20
KIT_D812>VD	0.23	0.20
MEIS2_W326*	0.00	0.44
LRRC3_R38P	0.00	0.41
SIN3A_Y325C	0.00	0.39
APTX_E9_UTR_3	0.00	0.35
ARHGAP30_E12_UTR_3	0.00	0.16

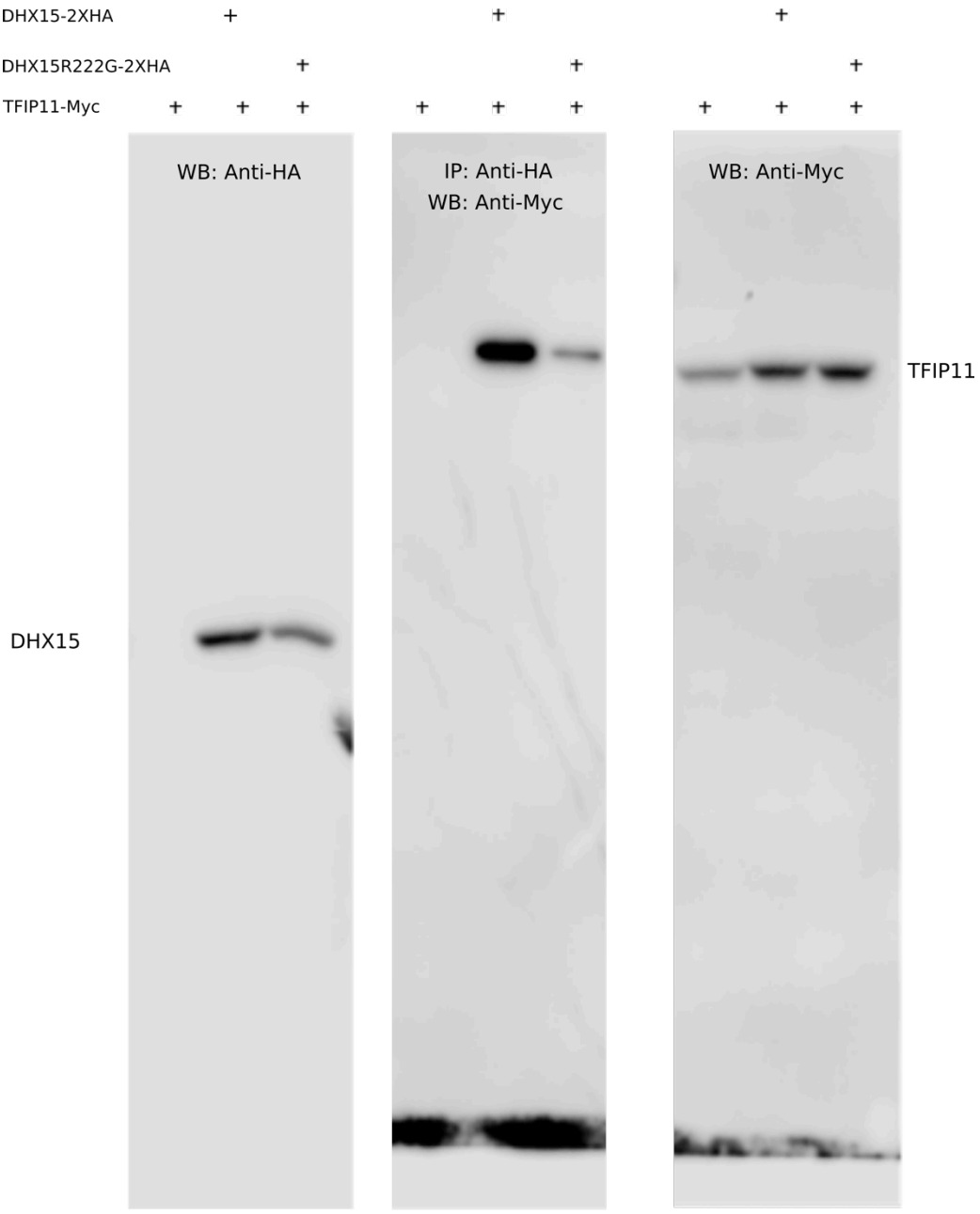
b)


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
CNV Summary
SJCBF028_D_SJCBF028_G
SJCBF028_R_SJCBF028_G
SJCBF030_D_SJCBF030_G
SJCBF030_R_SJCBF030_G
SJCBF043_D_SJCBF043_G
SJCBF043_R_SJCBF043_G
SJCBF044_D_SJCBF044_G
SJCBF044_R_SJCBF044_G
SJCBF078_D_SJCBF078_G
SJCBF078_R_SJCBF078_G
SJCBF155_D_SJCBF155_G
SJCBF155_R_SJCBF155_G
SJCBF165_D_SJCBF165_G
SJCBF165_R_SJCBF165_G
SJCBF174_D_SJCBF174_G
SJCBF174_R_SJCBF174_G
Deletion
CN-LOH
Amplification

Supplementary Figure 23. CBF AML mutation spectrum vs TCGA AMLs. The mutation status of the recurrently mutated genes reported in this study was compared to the 182 cases of non-CBF AML in the TCGA data set. The percentage of cases within the entire cohort with a mutation in the given gene is displayed.



Supplementary Figure 24. CoIP full length blots.



Supplementary Tables

Supplementary Table 1: Clinical information for all 165 sequenced tumor samples. D, G, and R indicate diagnosis, germline, and relapse samples respectively.

Supplementary Table 2: Coding region SNVs per case of CBF-AML

Supplementary Table 3: Coverage data for WGS cohort

Supplementary Table 4: Coverage data for WES cohort

Supplementary Table 5: Validated mutations found in the 17 WGS cases

GeneName: HUGO gene name
VarType: SUB – substitution, INDEL – insertion or deletion
SJQuality: mutation quality score
Sample: sample identifier
Fusion: AE – *RUNX1-RUNX1T1*, inv(16) – *CBFB-MYH11*
Chr: Chromosome
WU_HG19_Pos: nucleotide position in hg19
Class: mutation class
AAChange: amino acid change resulting from the mutation
ProteinGI: NCBI protein GI number
mRNA_acc: RefSeq mRNA accession number
#Mutant_in_Tumor: number of sequencing reads containing the mutation in the tumor
#Total_in_Tumor: number of total sequencing reads covering that nucleotide position
#Mutant_in_Normal: number of sequencing reads containing the mutation in the matched germline sample
#Total_in_Normal: total number of reads at the site in the matched germline sample
MAF: mutant allele frequency (#Mutant_in_Tumor/#Total_in_Tumor)
Reference Allele: allele in the reference genome, insertions are denoted with "-"
Mutant Allele: non-reference allele
Flanking: nucleotide sequence surrounding the mutation site [reference/mutant]
Validation Status: status of the mutation (somatic, uncovered, wild type, or germline)
SIFTResult: prediction of the mutations effect on the protein
SIFTScore: numerical score from the SIFT algorithm
pph2result: PolyPhen-2 prediction of the mutations effect
pph2score: score from the PolyPhen-2 algorithm

Supplementary Table 6: Validated WES mutations

Columns in the same order as Supplementary Table 5.

Supplementary Table 7: Recurrent somatic mutations in discovery and recurrency cohorts

Columns in the same order as Supplementary Table 5.

Supplementary Table 8: Copy number alterations identified by WGS

Sample: Sample identifier
HG_Version: reference genome build
Chromosome Location: Chromosome number and cytoband

Start: nucleotide position of the 5' end of the CNA
End: nucleotide position of the 3' end of the CNA
CNAs:
Size (kb): size in kilobases of the alteration
Number of Genes: number of genes contained within the CNA region
Genes: genes contained within the region (only shown if the number <10)
# of miRNAs: number of micro RNA genes found within the region
# of CGC Genes: number of genes within the region that are on the cancer gene consensus list
CGC Genes: cancer gene consensus list genes

Supplementary Table 9: Validated structural variants for the 17 pediatric CBF-AML discovery cases
Sample: Sample identifiers
ChrA: chromosome for breakpoint A
PosA: position of breakpoint A
OrientationA: "+" region left of the breakpoint in included in the mutation or "–" region right of the breakpoint is included in the mutation
SclipA:
ChrB, PosB, OrientationB, SclipB: same as with A, but the other end of the breakpoint
Type: INS – insertion, CTX – interchromosomal translocation, DEL – deletion, INV – inversion, ITX – intrachromosomal translocation
Usage: Intronic/intergenic – neither endpoint was in a gene, genic – both endpoints were in genes, truncating – one endpoint led to a truncation of a gene, half_intergenic – one endpoint was in a gene, but no truncation results

Supplementary Table 10: Rank ordered list of genes differentially expressed in DHX15 knockdown cells and their enrichment in the Reactome mRNA splicing gene set

Supplementary Table 11: Rank ordered list of genes differentially expressed in DHX15 knockdown cells and their enrichment in the KEGG ribosome gene set

Supplementary Table 12: GSEA enriched gene sets after *DHX15* knockdown

Supplementary Table 13: IP mass spectrometry spectral counts
SC: Spectral counts
TP: Total peptides

Supplementary Table 14: RNAseq *RUNX1-RUNX1T1* upregulated genes
Gene: HUGO gene name
FoldChange: Expression change between RUNX1-RUNX1T1 and CBFB-MYH11
Mean_inv16: Mean expression in CBFB-MYH11 cases
Mean_AE: Mean expression in RUNX1-RUNX1T1 cases
FDR: False discovery rate

Supplementary Table 15: RNAseq *RUNX1-RUNX1T1* downregulated genes
Columns in the same order as Supplementary Table 11

Supplementary Table 16: GSEA gene list

Supplementary Table 17: Copy Number Analysis of de novo-relapse pairs

Chrom: Chromosome
loc.start: Start of copy number alteration
loc.end: End of copy number alteration
num.mark: number of variable-length windows contained in the segment called
seg.mean: log2[Tumor coverage/germline coverage]
observed.CN: Copy number converted from log2 ratio values listed in "seg.mean"
comments: Comments regarding type of alteration
Cytoband: Chromosomal location based on cytogenetic banding
num_gene: number of genes involved
genes: specific genes involved

Supplementary Table 18: Deep sequencing read counts in diagnosis, germline, and relapse trios

GeneName: HUGO gene name
Sample: Patient sample identifier
Chr: Chromosome
WU_HG19: nucleotide position in hg19
Class: mutation class
AAChange: amino acid change resulting from the mutation
#Mutant_in_D: number of sequencing reads containing the mutation in the diagnosis sample
#Total_in_D: number of total sequencing reads covering that nucleotide position in the diagnosis sample
#Mutant_in_normal: number of sequencing reads containing the mutation in the matched germline sample
#Total_in_normal: total number of reads at the site in the matched germline sample
#Mut_in_R: number of sequencing reads containing the mutation in the relapse sample
#Total_in_R: number of total sequencing reads covering that nucleotide position in the relapse sample
MAF_D: Mutant allele frequency in the diagnosis sample
MAF_G: Mutant allele frequency in the germline sample
MAF_R: Mutant allele frequency in the relapse sample
Forward Primer: Sequence of forward primer
Reverse Primer: Sequence of reverse primer
Forward Primer Tm: Melting temperature of forward primer
Reverse Primer Tm: Melting temperature of reverse primer
Flanking Sequence (SNV pos = []): Sequence flanking the SNV, which is marked by brackets
Percent GC target: Percentage of GC base pairs in the target sequence
Predicted Amplicon Size: Size of predicted amplicon using indicated primers

Supplementary Table 19: Oligonucleotides used in this study

Identifier: Catalog number for Stealth or Silencer siRNAs, internal identifier for site-directed mutagenesis primers
Sequence: Sequence of indicated oligonucleotide

Supplementary Note

Signaling mutations in CBF AML

In our analysis, 109 out of 165 cases harbored mutations in signaling genes *NF1*, *PTPN11*, *KRAS*, *FLT3*, *KIT*, and *NRAS*. Mutations in these genes were enriched in *CBFB-MYH11* AMLs, as 56.4% of *RUNX1-RUNX1T1* cases had mutations in these genes, compared to 76.2% of *CBFB-MYH11* cases (Figure 1; p=0.0086). Alterations in these 6 genes were not cumulatively associated with outcome (see Supplementary Tables). Although mutations in these 6 genes were less common in *RUNX1-RUNX1T1,* cases of that subtype did had mutations in *CCND2, MGA* and *MYC* more frequently than *CBFB-MYH11* cases (16.4% vs 3.75%), which also were not associated with outcome (data not shown).

We detected *NRAS* and *KIT* mutations in 56 and 33 *de novo* cases, respectively. There was a significant difference in the distribution of *NRAS* Q61 and G12/G13 mutations between the two CBF-AML subgroups, with more codon 61 alterations in *CBFB-MYH11* cases (p=0.0023). Mutation status of *NRAS* did not correlate with outcomes (Supplementary Figure 8). All these mutations lead to enhanced Ras pathway signaling, ultimately resulting in increased MAP kinase activation[80] and thus cells harboring these mutations are predicted to be sensitive to MEK inhibitors[81].

KIT mutations are relatively common in CBF-AML with reported frequencies between 17% and 48%[82,83]. Similar to *NRAS*, the spectrum of KIT mutations differed between the two subgroups, with exon 8 mutations being more common in *CBFB-MYH11* patients and exon 17 mutations having an increased prevalence in *RUNX1-RUNX1T1* patients (p=0.0051). Mutations in exon 17, but not exon 8, were associated with inferior outcome. One case (SBCBF030) lost a subclone with an exon 8 mutation and retained an exon 17 mutation at relapse. These data are relevant as the kinase inhibitor imatinib has shown efficacy as a KIT antagonist, although it is ineffective against the constitutively active D816 mutation[83]. However, the N822K mutation is not resistant to imatinib and is as sensitive as wild type KIT to inhibition by imatinib[83]. Both mutations show varying sensitivity to dasatinib[84].

References

80. Dhillon, A.S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279-90 (2007).
81. Burgess, M.R. *et al.* Preclinical efficacy of MEK inhibition in Nras-mutant AML. *Blood* **124**, 3947-55 (2014).
82. Boissel, N. *et al.* Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia* **20**, 965-70 (2006).
83. Wang, Y.Y. *et al.* AML1-ETO and C-KIT mutation/overexpression in t(8;21) leukemia: implication in stepwise leukemogenesis and response to Gleevec. *Proc Natl Acad Sci U S A* **102**, 1104-9 (2005).
84. Wang, Y.Y. *et al.* C-KIT mutation cooperates with full-length AML1-ETO to induce acute myeloid leukemia in mice. *Proc Natl Acad Sci U S A* **108**, 2450-5 (2011).