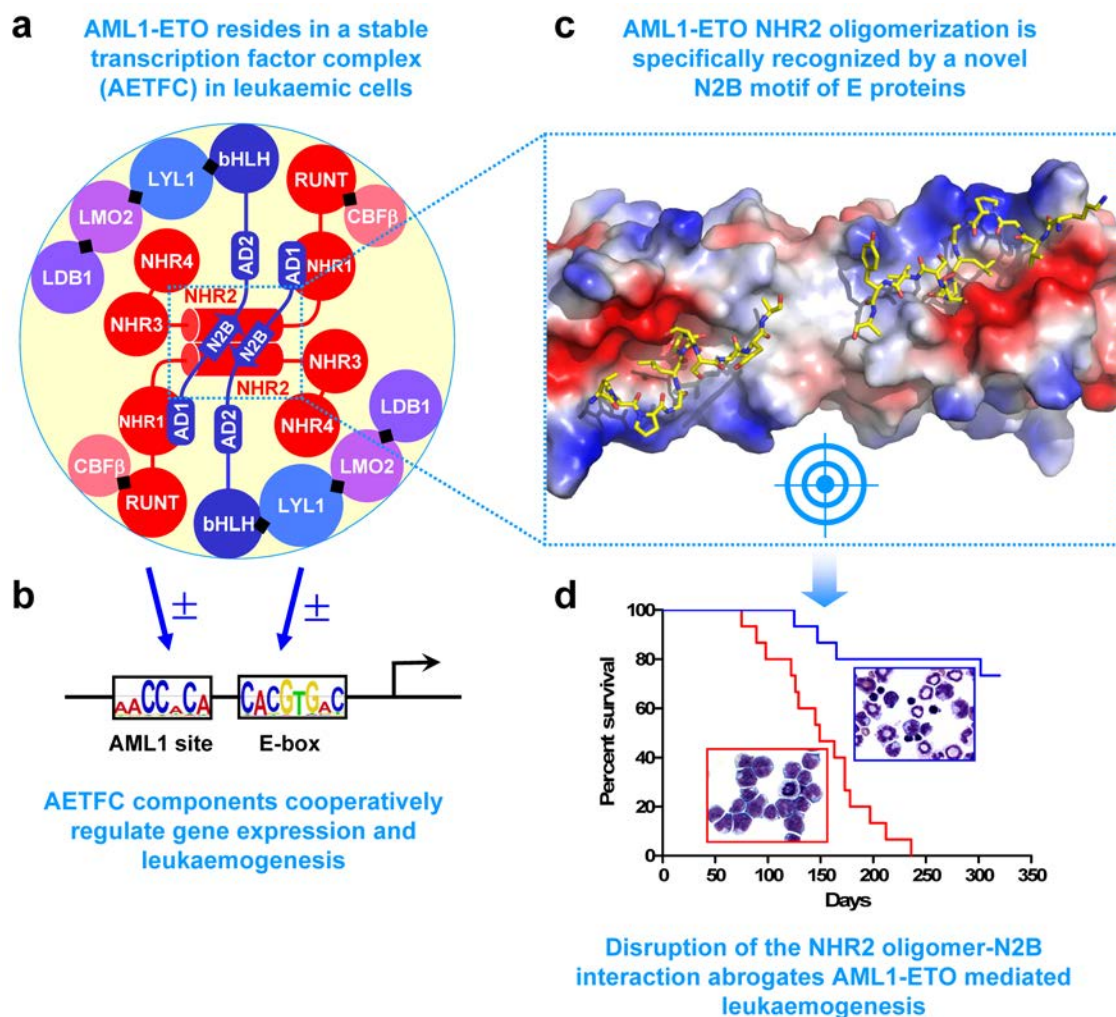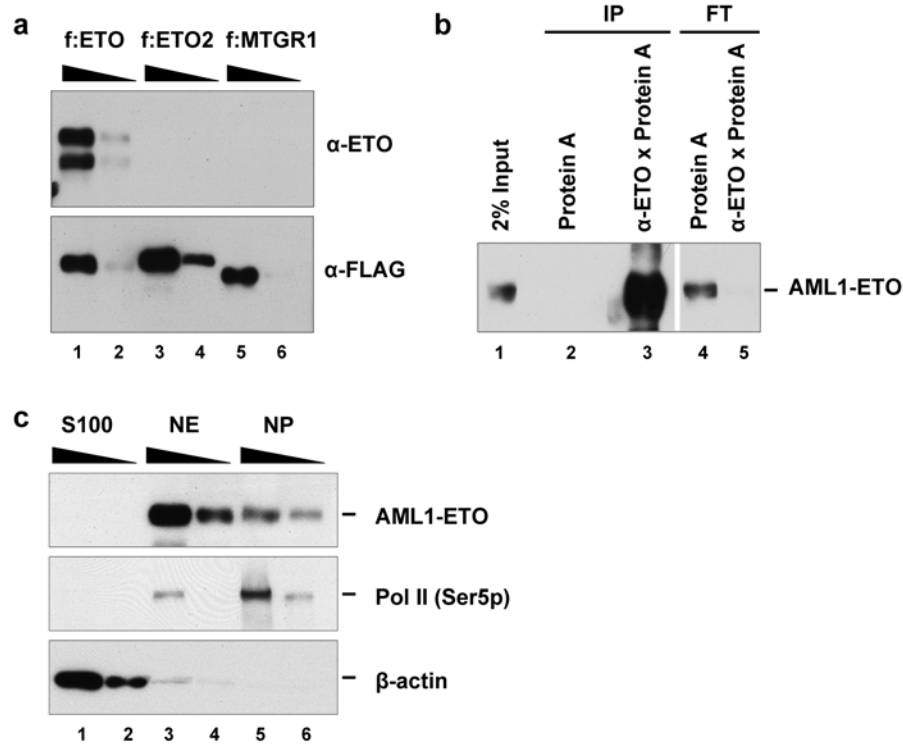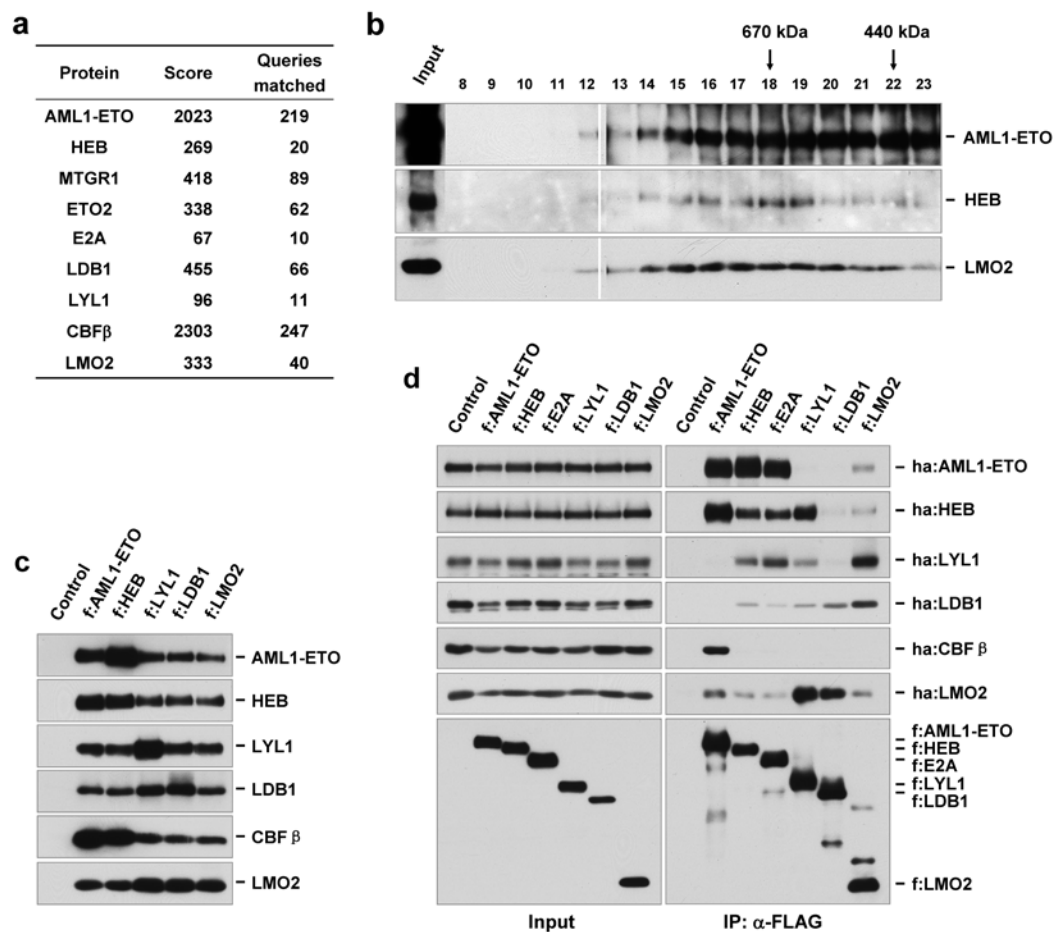**Supplementary Figures**

Supplementary Figure 1. Summary of the main findings. **a**, An unbiased biochemical analysis of the natural state of AML1-ETO in leukemic cells led to the identification and characterization of a novel endogenous complex (AETFC) containing AML1-ETO in association with several haematopoietic transcription factors and cofactors. Domain architectures of AML1-ETO (red) and E proteins (HEB or E2A; blue) are presented in detail, showing the direct NHR1-AD1 and NHR2-N2B interactions. Black diamonds denote other strong interactions among the essential AETFC components (some weak interactions are not shown). **b**, ChIP-seq and RNA-seq analyses, in conjunction with knockdowns of individual components and *in vivo* leukemogenesis assays, revealed that the AETFC components co-bind target genes (potentially

through different DNA-binding domains), cooperatively regulate gene expression and contribute to leukemogenesis. **c**, In relation to the critical requirement of the AML1-ETO NHR2 oligomerization domain for leukemogenesis, NHR2 oligomerization was shown to be required for its interaction with a newly defined motif (N2B) of E proteins. X-ray structural analyses demonstrated that a single N2B peptide makes direct contacts with two NHR2 α-helices as a dimer. This unique interaction pattern provides a novel model for dimeric/oligomeric transcription factors to create a new protein-binding interface through the dimerization/oligomerization. **d**, Targeted disruption of the NHR2 oligomer-N2B interaction abrogates AML1-ETO–mediated haematopoietic stem cell self-renewal and leukemogenesis. Survival curves and bone marrow morphological analyses of wild-type (red) and NHR2-mutated (blue) mice are presented. In contrast to the highly stable NHR2 oligomerization, which makes its therapeutic targeting very challenging, the NHR2 oligomer-N2B interaction potentially offers a new target.

Supplementary Figure 2. Characterization of the antigen-purified anti-ETO antibody and the Kasumi-1 cell derived nuclear extract. **a**, High specificity of the anti-ETO antibody. FLAG-tagged ETO, ETO2 and MTGR1 were over-expressed in 293T cells and probed by immunoblot with the anti-ETO antibody (upper) and the anti-FLAG antibody (bottom). Note that the anti-ETO antibody specifically recognizes ETO but not ETO2 or MTGR1. **b**, High affinity of the anti-ETO antibody. Immunoprecipitation (IP) was performed with Kasumi-1 nuclear extract and the anti-ETO antibody conjugated with protein A Sepharose (unconjugated protein A Sepharose as a control), and the precipitated complex and flow-through (FT) were subjected to immunoblot analysis. The anti-ETO antibody almost completely depletes the AML1-ETO protein from the nuclear extract. **c**, The majority of AML1-ETO protein resides in the Kasumi-1 nuclear extract. Immunoblot was performed with equal cell equivalents of cytoplasmic (S100), nuclear extract (NE) and nuclear pellet (NP) fractions. Serine 5-phosphorylated RNA polymerase II (Pol II) and β-actin were used as markers for NP and S100, respectively.

a

Protein	Score	Queries matched
AML1-ETO	2023	219
HEB	269	20
MTGR1	418	89
ETO2	338	62
E2A	67	10
LDB1	455	66
LYL1	96	11
CBFβ	2303	247
LMO2	333	40

Supplementary Figure 3. Characterization of the AETFC complex. **a**, Components identified by mass spectrometry analysis with high scores and numbers of queries matched. Among the components, the presence of the ETO homologues ETO2 and MTGR1 in the complex may reflect their heterodimerization with AML1-ETO. **b**, Superose 6 gel filtration analysis of the antigen-eluted AETFC complex. Each fraction (numbered on the top) was subjected to immunoblot with indicated antibodies. The molecular weight values were derived from a calibration curve. Since the apparent size (>670 kDa) of AETFC is much greater than the calculated size (~300 kDa) of the minimum complex formed by these components, AML1-ETO oligomerization likely plays a role in assembling AETFC. **c**, Intracellular reconstitution of AETFC. Sf9 cells were co-infected with baculoviruses expressing a single FLAG-tagged AETFC component (indicated on the top) and all of the components indicated on the right. Complexes in cell lysates (six, including the control with no FLAG-tagged protein vector) were

immunoprecipitated with anti-FLAG antibody and subjected to immunoblot with antibodies to the proteins indicated on the right. Note that, even under high stringency conditions (300 mM NaCl and 1% Triton X-100), the complex can be co-purified by tagging any single component. **d**, Determination of pairwise interactions among the AETFC components. Sf9 cells were co-infected with pairs of baculoviruses expressing a single FLAG-tagged component (top) and a single HA-tagged AETFC component (right). Anti-FLAG immunoprecipitates (42 including the controls) were analyzed by immunoblot with anti-HA antibody (upper panels) to score the pairwise interactions and with anti-FLAG antibody (lower panel) to score expression of the FLAG-tagged components.
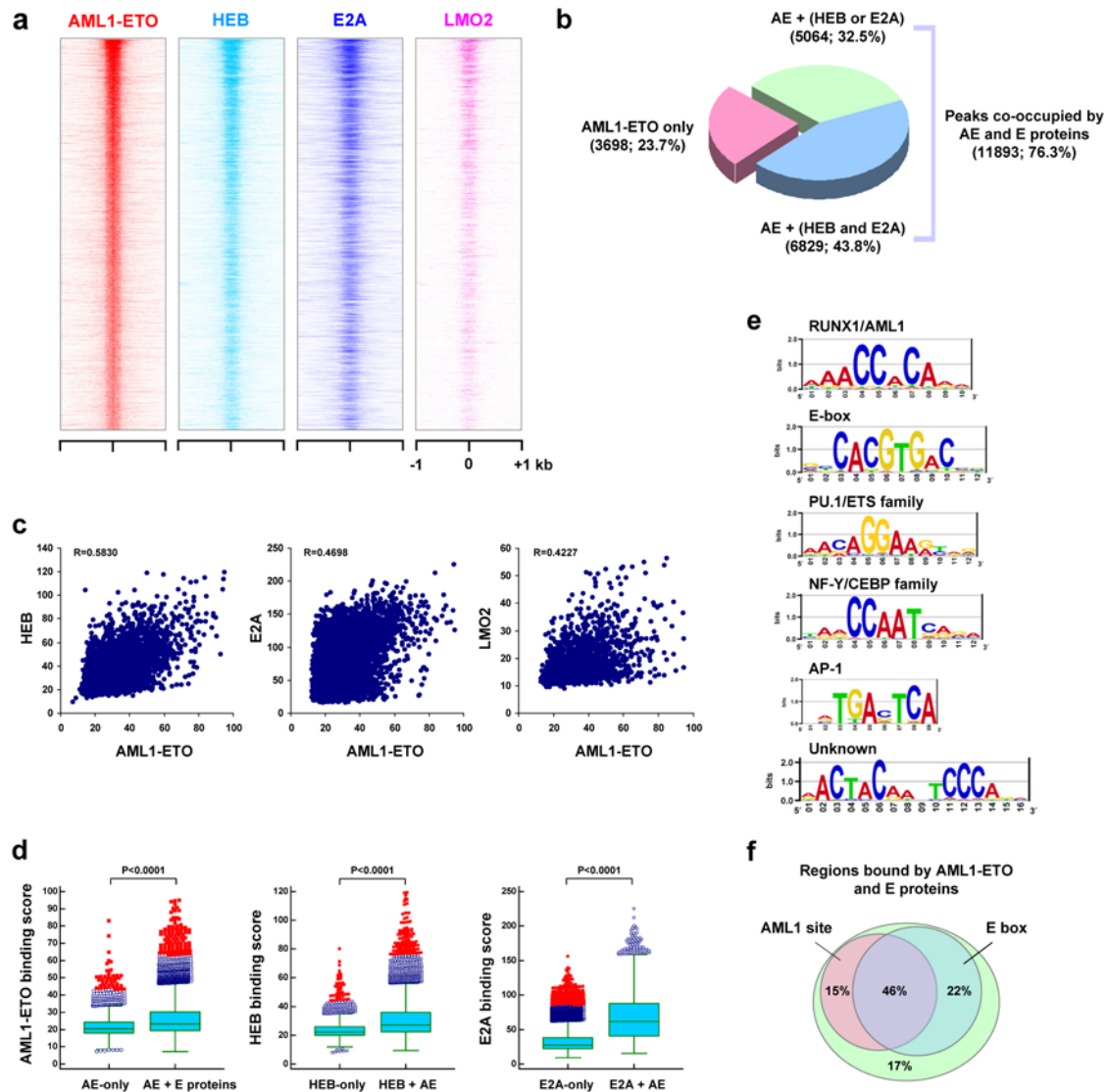
Supplementary Figure 4. AETFC components bind cooperatively to target genes.   **a**, Heat map of binding signals for AML1-ETO, HEB, E2A and LMO2 on regions from -1 kb to +1 kb surrounding the AML1-ETO binding sites, ordered by AML1-ETO binding strength. The top 3000 AML1-ETO binding sites are shown. **b**, AML1-ETO binding sites largely overlap the binding sites of HEB, E2A or both. **c**, Correlation analysis of the binding scores of AML1-ETO with HEB, E2A and LMO2 to target genes. **d**, Binding scores of AML1-ETO, HEB and E2A were compared between the sites co-bound by AML1-ETO and E proteins and the sites bound only by AML1-ETO or E proteins. Note that the binding of each protein is significantly higher on the co-bound sites, suggesting that the interactions among the components facilitate their binding to

the genome. **e**, Consensus sequences over-represented in the AML1-ETO binding regions (FDR<0.001) and their corresponding transcription factors. Besides the AML1 sites and E-boxes, motifs that correspond to other haematopoietic transcription factors may reflect previously reported, likely dynamic, interactions of these transcription factors with the AETFC components. This analysis also identified an unknown motif (bottom) whose corresponding transcription factor(s) and functional relevance remain to be studied. **f**, Motif mapping analysis reveals that the majority of genomic regions co-bound by AML1-ETO and E proteins contain AML1 sites and/or E-boxes: 46% of the regions contain both AML1 sites and E boxes, 15% contain an AML1 site alone, 22% contain E box alone, and 17% contain neither an AML1 site nor an E-box. Multiple DNA motif matrices corresponding to AML1 sites and E-boxes were obtained from the TRANSFAC database and used in the motif mapping. These results suggest that both AML1 sites and E-boxes contribute to the recruitment of AETFC to the genome, and that the coexistence of these two motifs may serve as a preferred recognition site for AETFC.
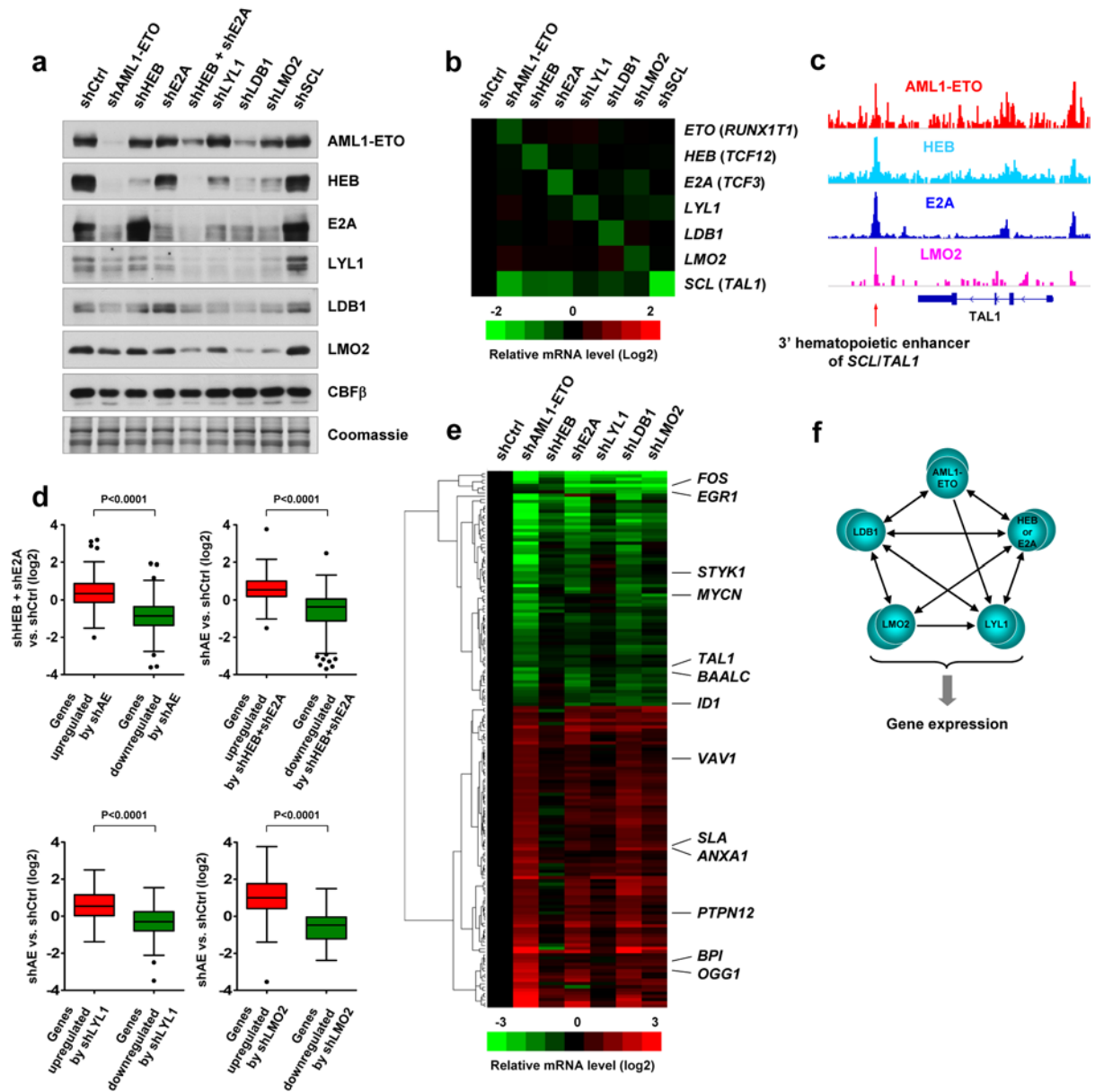
Supplementary Figure 5. AETFC components stabilize the complex and cooperatively regulate gene expression. **a**, AETFC protein levels following knockdowns of individual AETFC components in Kasumi-1 cells. Nuclear extracts were analyzed by immunoblot with antibodies against AETFC components indicated on the right. Note that the knockdowns of individual components decrease the protein levels of some of the other components, and that the double-knockdown of HEB and E2A shows the most dramatic effect. In contrast, none of the components are affected by the knockdown of SCL, a homologue of LYL1 not present in AETFC.

b, AETFC mRNA levels following knockdowns of AETFC components. Note that the individual knockdowns do not significantly affect the mRNA levels of the other components of the complex. However, they do decrease the *SCL* mRNA level, implicating *SCL* as a target gene cooperatively regulated by AETFC components. The expression data were extracted from RNA-seq results. **c**, The AETFC components AML1-ETO, HEB, E2A and LMO2 colocalize on the well-established 3' haematopoietic enhancer of the *SCL*/*TAL1* locus. **d**, Correlation between AML1-ETO knockdown and HEB/E2A double-knockdown in up- and down-regulation of gene expression (upper panels). Note that the genes changed by AML1-ETO knockdown are likewise regulated by HEB/E2A knockdown, and *vice versa*. Genes with greater than 2-fold changes relative to a negative control (shCtrl) were analyzed. Similar correlations were observed in comparison of AML1-ETO knockdown with LDB1, LYL1 and LMO2 knockdowns (bottom panels; and see Figure 1e). **e**, Hierarchical cluster and heat map of the genes co-bound and cooperatively regulated by AETFC components. Indicated are several key genes implicated in cancer-related functions. **f**, Model showing the stabilization network among the AETFC components and their cooperativity in regulating gene expression. Arrows indicate the stabilization relationships between the AETFC components according to the knockdown and immunoblot assays shown in (**a**).
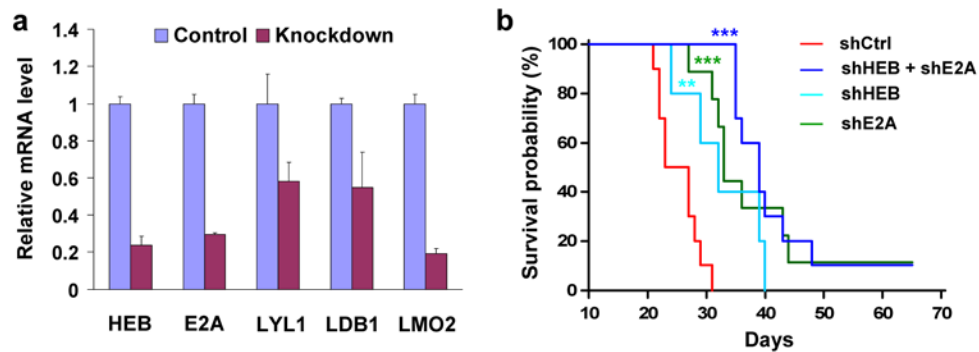
**Supplementary Figure 6. Knockdown of AETFC components delays leukemogenesis in mice.** **a**, RT-qPCR analysis of the knockdown efficiencies of AETFC components relative to a negative control in AML1-ETO9a–driven mouse leukemic cells. The transplantable mouse leukemic cells, which were originally derived from primary spleen cells of mice bearing AML1-ETO9a–induced leukemia, were infected with shRNA-expressing lentiviruses to knock down the AETFC components and then transplanted into recipient mice. Averages and standard deviations were calculated from two separate assays. **b**, Kaplan-Meier survival curves of mice bearing the leukemic cells with knockdowns of HEB, E2A and both. See Figure 1f for the effects of knockdown of other AETFC components. Note that double knockdown of HEB and E2A shows a more dramatic effect in the delay of leukemogenesis compared with the single knockdown of either HEB or E2A. ***P<0.001; **P<0.01.
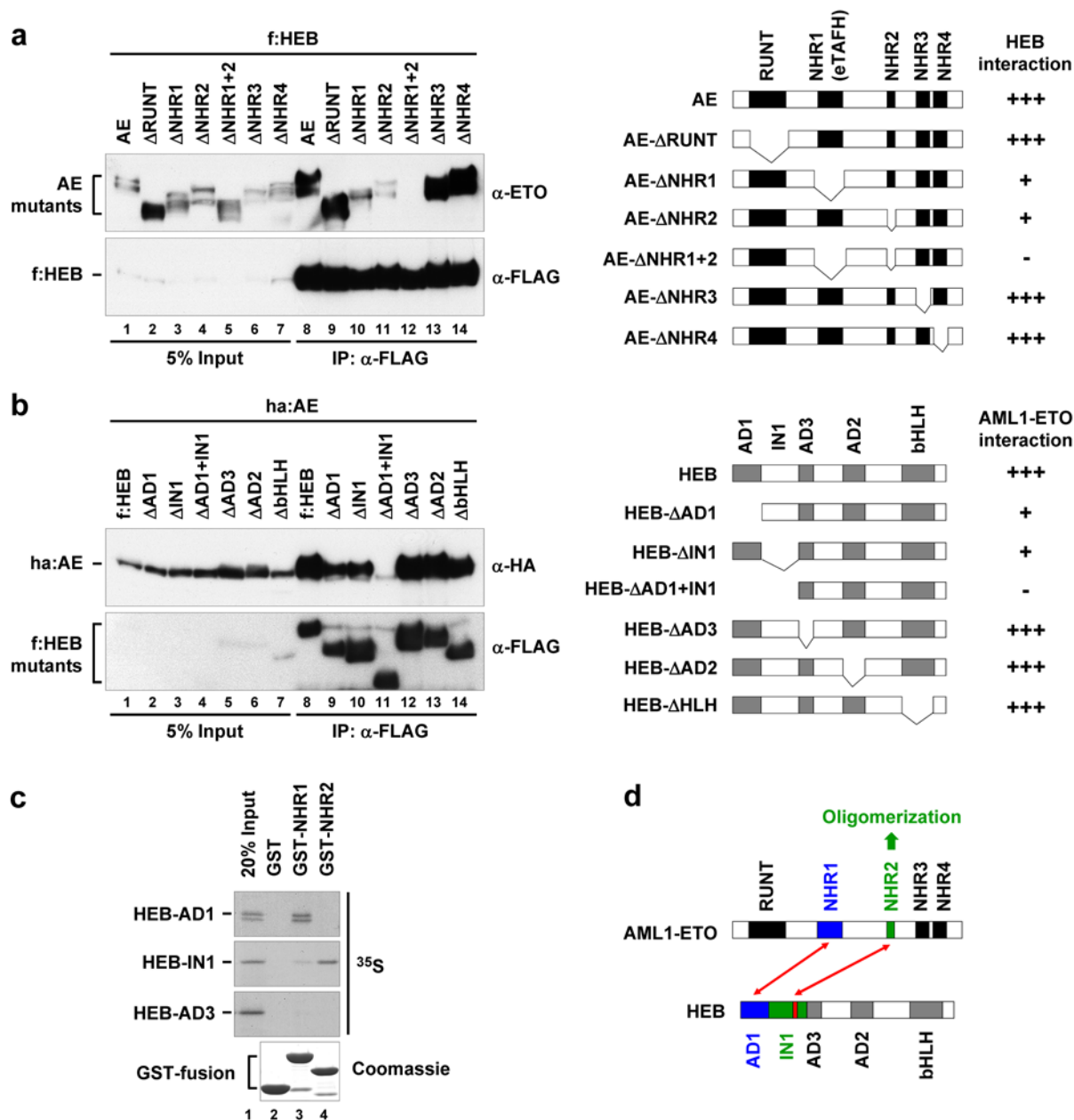
Supplementary Figure 7. Characterization of the direct interactions between AML1-ETO and E proteins. **a**, **b**, Coimmunoprecipitation assays mapping the interacting domains between AML1-ETO (AE) and HEB (left panels) and schematic presentations of the intact AML1-ETO and HEB proteins (with specific domains), the deletion mutants and the experimental results (right panels). The indicated wild-type (WT) and mutant proteins, with FLAG (f) or HA (ha) tags as indicated, were co-expressed in 293T cells. Anti-FLAG immunoprecipitates of cell lysates were subjected to immunoblot analyses with the indicated antibodies. **c**, Independent

NHR1-AD1 and NHR2-IN1 interactions. GST-pull down assays were conducted with GST-tagged AML1-ETO NHR1 and NHR2 domains and ^{35}S-labelled HEB AD1, IN1 and AD3 domains, with binding scored by autoradiography. **d**, Summary of the NHR1-AD1 and NHR2-IN1 interactions between AML1-ETO and HEB. The red block within IN1 represents the N2B domain described in Figure 3.
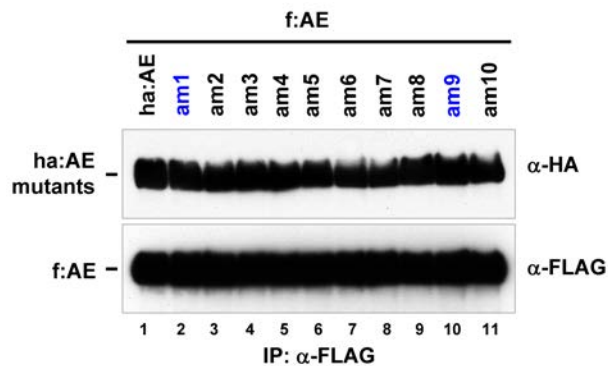
Supplementary Figure 8. AML1-ETO oligomerization is not affected by the herein screened NHR2 mutations. Indicated wild-type and mutant AML1-ETO (AE), with FLAG (f:) or HA (ha:) tags, were co-expressed in 293T cells. Anti-FLAG immunoprecipitates of cell lysates were subjected to immunoblot with indicated antibodies. Whereas am1 and am 9 (blue) dramatically disrupt the interaction between AE and HEB-ΔAD1 (see Fig. 2b), they do not disrupt AE oligomerization.
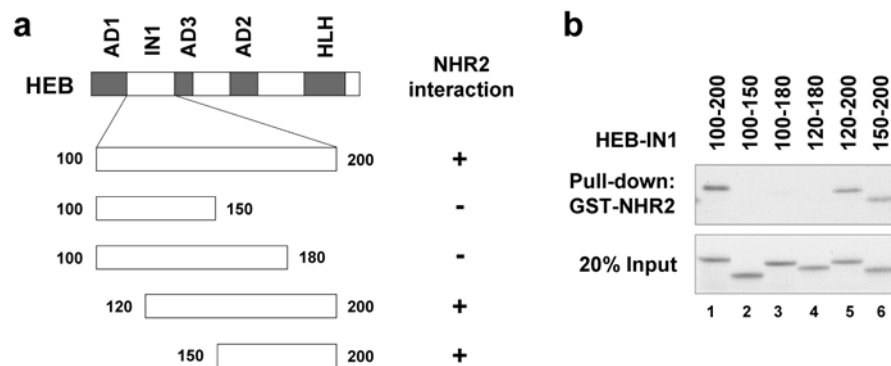
Supplemental Figure 9. Mapping the NHR2-binding region of HEB. **a**, Schematic presentation of HEB-IN1 deletion mutants and a summary of their interactions with NHR2. **b**, The 150-amino acid C-terminal region of HEB-IN1 is sufficient to bind NHR2. The binding of ^{35}S-labelled HEB-IN1 fragments (top) to GST-tagged NHR2 was scored by autoradiography.
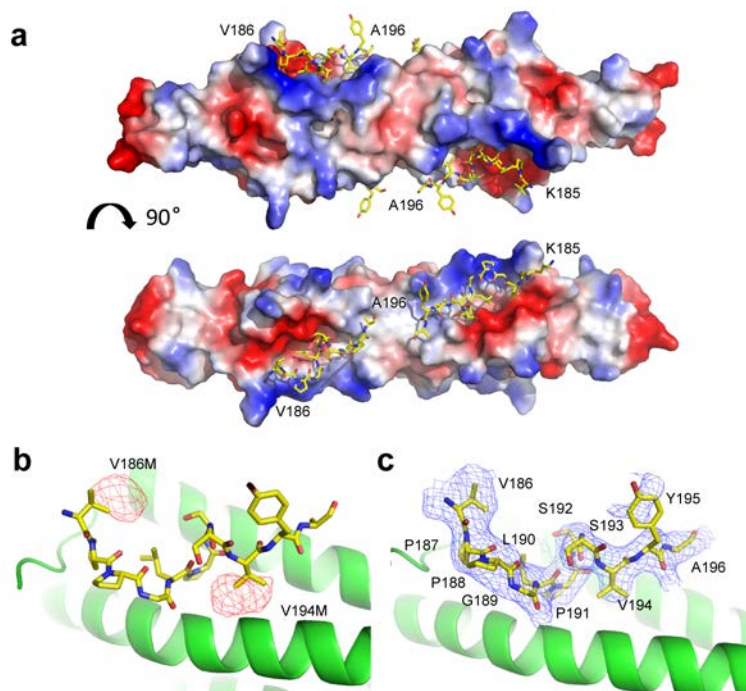
Supplementary Figure 10. Structural details and validation of the directionality and register of the N2B(177-200) peptide of HEB bound to the NHR2 domains of AML1-ETO. **a**, Side (top) and look-down (bottom) surface views of the structure of the complex between the N2B(177-200) peptide and the tetrameric α-helical bundle of the NHR2 domain. **b**, Anomalous difference map contoured at 4σ level of the selenomethionine-labeled double mutation V186M/V194M (red density) of the N2B peptide in the complex with NHR2 domains. Note that the anomalous signal verified the positions of both Val186 and Val194 on the N2B peptide and hence defined both the directionality and register of the bound peptide. **c**, Electron density map of the N2B(177-200) peptide fragment bound to the NHR2 domains. The 2Fo-Fc electron density map was contoured at 1σ level.

a

mNHR1
```
                 A
                 ↑
ARQLSKLKRFLTTLQQFGNDISPEIGERVRTLVLGLVNSTLTIEEFHSK
LQEATNFPLRPFVIPFLKANLPLLQRELLHCARLAKQNPAQYLAQHEQ
                 ↓
                 A
```

mNHR2
```
              AA                  AAA
              ↑↑                  ↑↑↑
QEEMIDHRLTDREWAEEWKHLDHLLNCIMDMVEKTRRSLTVLRRCQEAD
REELNYWIRRYSDAEDLKK
```

b

c
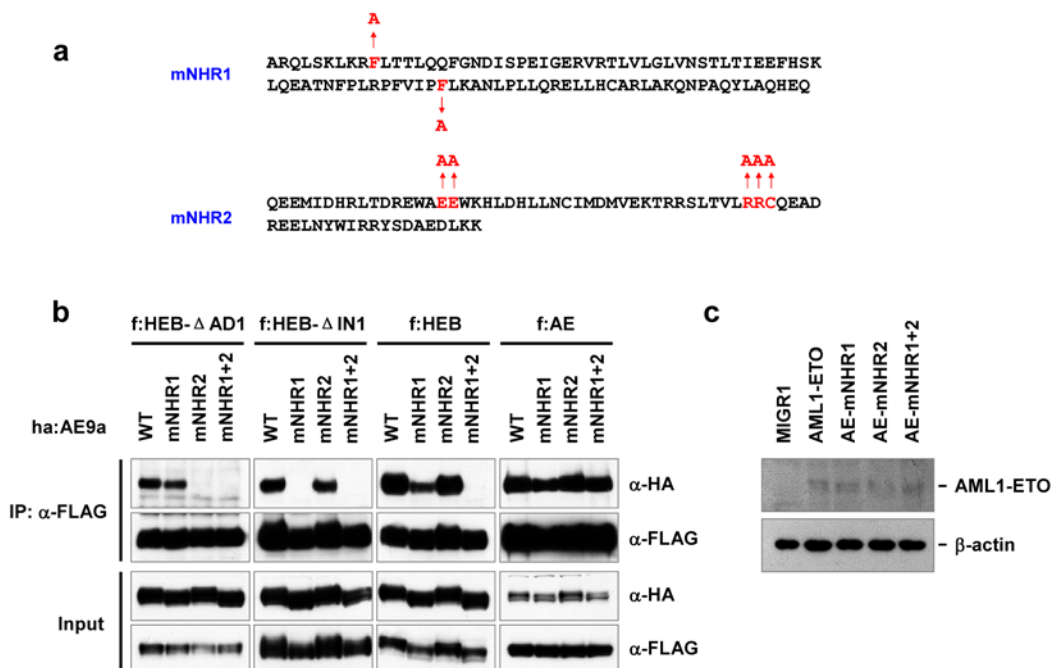
Supplementary Figure 11. Sequences and validations of AML1-ETO mutations. **a**, Sequences of the mNHR1 and mNHR2 mutations, in which the indicated amino acids (red) were mutated to alanine. **b**, Point mutations in AML1-ETO specifically disrupt the NHR1-AD1 interaction (mNHR1), the NHR2-N2B interaction (mNHR2) or both (mNHR1+2). None of these mutations affect AML1-ETO oligomerization (right panels). Co-IP and immunoblot assays were performed following co-expression of the indicated FLAG (f:)- and HA (ha:)-tagged proteins in 293T cells. The data shown were generated with AML1-ETO9a; the same results were observed with full-length AML1-ETO. **c**, Expression levels of AML1-ETO and derived mutants in the human CD34$^+$ cells are comparable. Human CD34$^+$ cells were transduced with MIGR1-derived retroviruses expressing AML1-ETO and derived mutants, and cells were harvested after 1 week of *in vitro* culture and subjected to immunoblot analysis. β-actin was used as a loading control.
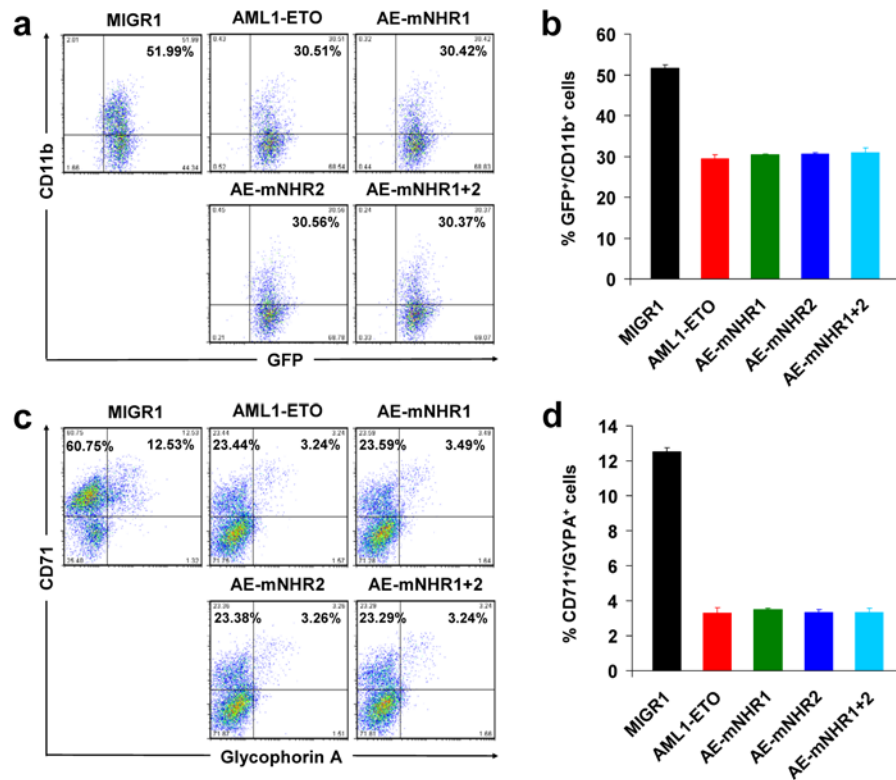
Supplementary Figure 12. NHR1/NHR2 mutations do not affect the ability of AML1-ETO to inhibit differentiation. **a**, Flow cytometry analysis of the percentage of CD11b$^+$ myeloid cells upon transduction with vector (MIGR1), wild-type and mutated AML1-ETO. **b**, Average percentages of CD11b$^+$ cells calculated from three separate assays for each construct. **c**, Flow cytometry analysis of the percentage of CD71$^+$/Glycophorin A (GYPA)$^+$ erythrocytes upon transduction with vector, wild-type and mutated AML1-ETO. **d**, Average percentages of CD71$^+$/GYPA$^+$ cells calculated from three separate assays for each construct.
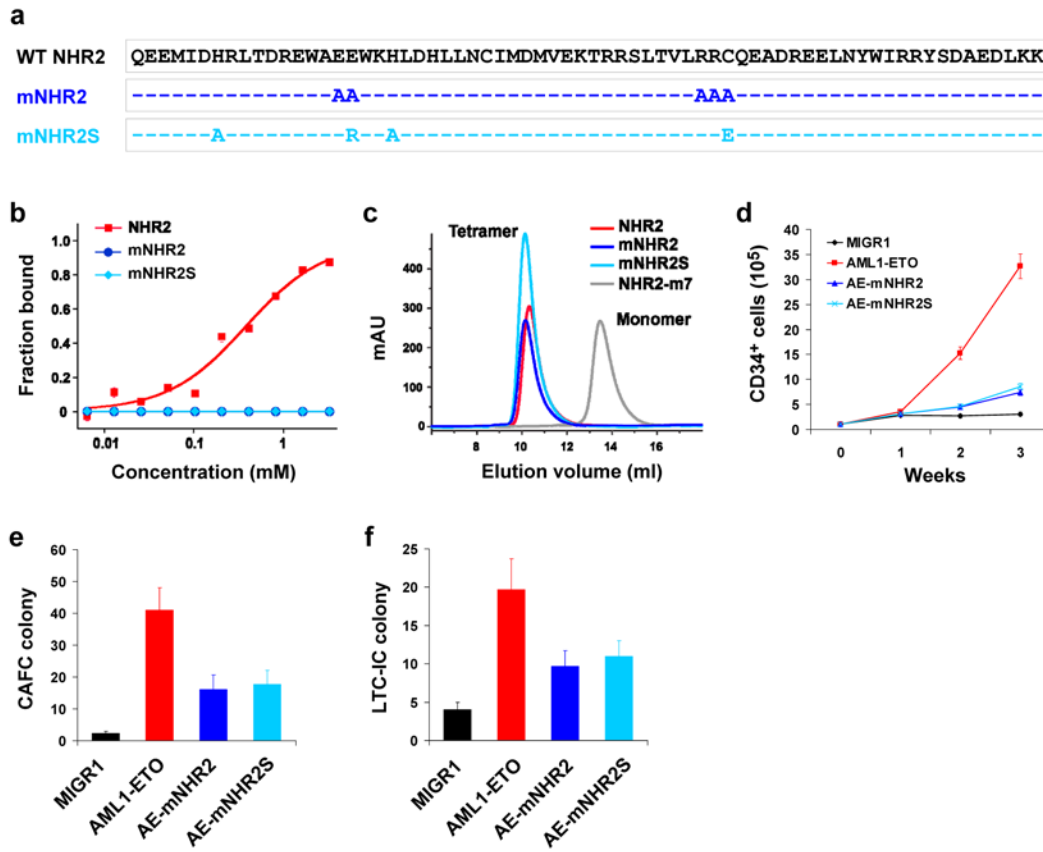
Supplementary Figure 13. Two different NHR2 mutants show similar effects in haematopoietic stem cell self-renewal. **a**, Sequences of the wild-type (WT) and mutated NHR2 domains. Dashes denote identical amino acids with WT, whereas the letters denote mutated amino acids. **b**, Fluorescence polarization measurement of the binding of WT and mutant NHR2 domains with a short HEB(182-196) peptide. **c**, Superdex 75 gel-filtration profiles of WT and mutant NHR2 domains. NHR2-m7 was used as a monomer control (see Fig. 2d). Note that, although their sequences are different, both mNHR2 and mNHR2S disrupt the NHR2-N2B interaction (**b**), but not the NHR2 oligomerization (**c**). **d-f**, Similar effects of mNHR2 and mNHR2S in CD34$^+$ HSC maintenance (**d**), CAFC assay (**e**) and LTC-IC assay (**f**). Averages and standard deviations were calculated from three independent experiments.
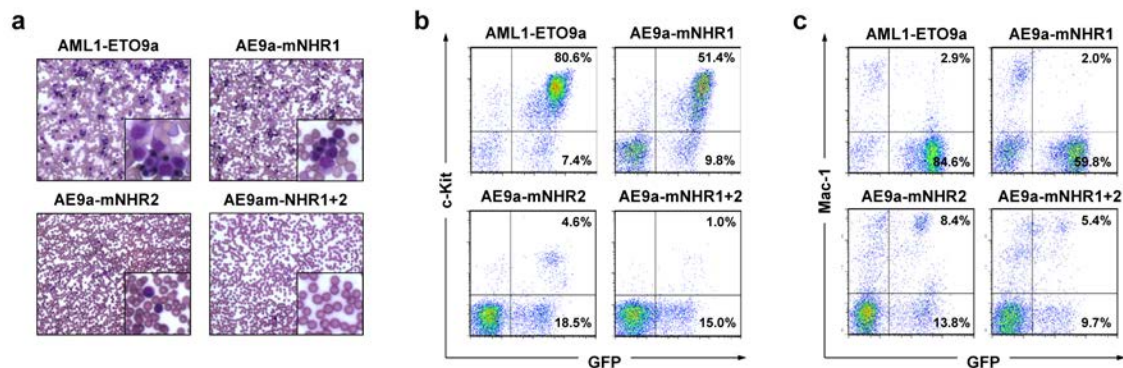
Supplementary Figure 14. Effects of AML1-ETO9a (AE9a) and derived mutants in mouse leukemogenesis. **a**, Morphological analysis of peripheral blood. The insets highlight the enriched, darkly stained leukemic cells relative to the erythrocytes. **b**, **c**, Flow cytometry analysis of peripheral blood. The c-Kit$^+$/Mac-1$^-$/GFP$^+$ cells represent leukemic blast cells that express AML1-ETO9a or indicated mutants.
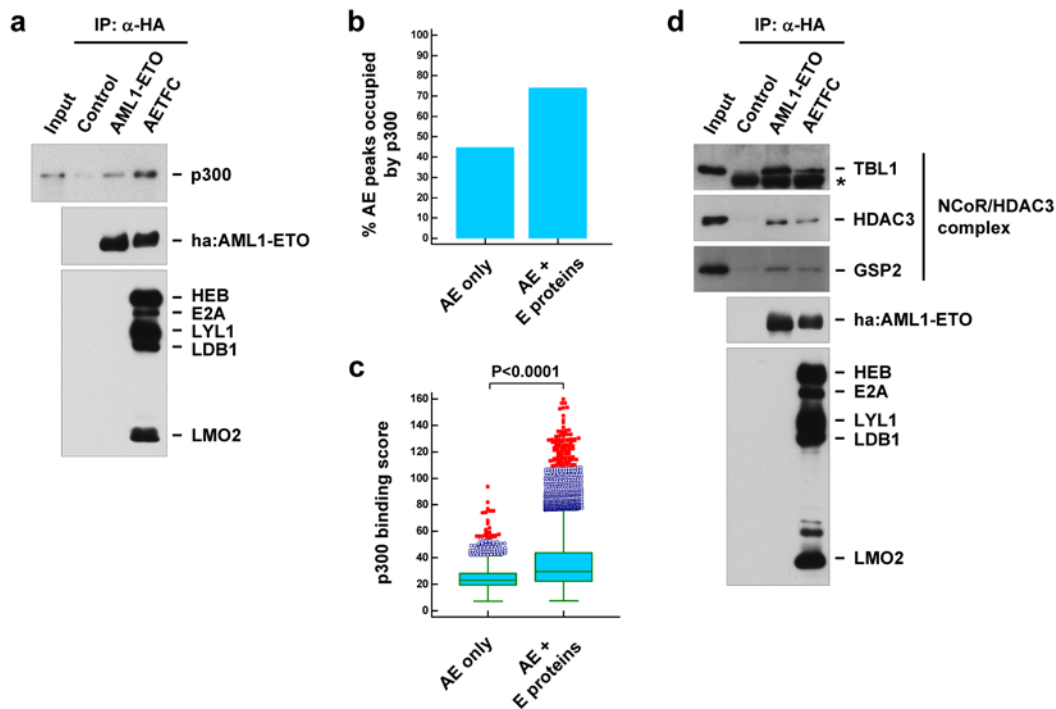
Supplementary Figure 15. Role of AETFC in recruitment of p300 and HDACs. **a**, In vitro interaction assay with purified AML1-ETO versus AETFC, showing that the formation of the AETFC complex enhances p300 binding to AML1-ETO. **b**, **c**, ChIP-seq analysis of p300 binding on the sites only bound by AML1-ETO versus those co-bound by AML1-ETO and E proteins. Note that p300 occupies a much higher percentage of the co-bound sites relative to the AML1-ETO–only sites (**b**), and that the p300 binding score is significantly higher on the co-bound sites (**c**). **d**, Formation of the AETFC complex does not enhance the AML1-ETO interaction with the NCoR/HDAC3 complex. The binding-immunoprecipitation (IP) analyses in (**a**) and (**d**) used HA-tagged AML1-ETO and an AETFC complex reconstituted with HA-tagged AML1-ETO and other subunits, purified p300 (**a**) and purified NCoR.HDAC3 complex (**d**).

Supplementary Tables

Supplementary Table 1. Genes Upregulated by the AETFC Complex.

Gene symbol	Full name	Expression in Kasumi-1 cells (RPKM)	Fold change (log2)	
			shAE	shLDB1
EGR1	early growth response 1	42.81	-0.27	-2.06
MTSS1	metastasis suppressor 1	40.18	-0.97	-0.68
BAALC	brain and acute leukemia, cytoplasmic	36.18	-1.34	-0.62
THSD1	thrombospondin, type I, domain containing 1	35.79	-1.80	-1.23
ETV5	ets variant 5	31.90	-0.89	-0.70
TSHZ3	teashirt zinc finger homeobox 3	29.97	-1.88	-1.35
ADAM28	ADAM metallopeptidase domain 28	29.30	-3.49	-1.49
ADAMTS3	ADAM metallopeptidase with thrombospondin type 1 motif, 3	27.94	-1.04	-0.53
ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3	27.56	-1.15	-0.81
PADI3	peptidyl arginine deiminase, type III	27.48	-3.57	-1.18
ITPKA	inositol-trisphosphate 3-kinase A	26.86	-1.17	-0.54
FOS	FBJ murine osteosarcoma viral oncogene homolog	26.77	-0.73	-2.93
CD226	CD226 molecule	25.39	-2.25	-0.93
NPTX1	neuronal pentraxin I	25.04	-3.24	-0.35
SULT1C2	sulfotransferase family, cytosolic, 1C, member 2	24.71	-2.87	-1.19
TAL1	T-cell acute lymphocytic leukemia 1	23.68	-0.64	-0.23
GAD1	glutamate decarboxylase 1	23.40	-2.09	-1.26
ZNF521	zinc finger protein 521	23.09	-2.06	-0.56
NR5A2	nuclear receptor subfamily 5, group A, member 2	22.74	-1.91	-0.57
HPSE	heparanase	22.36	-3.30	-1.16
CD48	CD48 molecule	22.32	-1.61	-0.65
ST18	suppression of tumorigenicity 18 (breast carcinoma) (zinc finger protein)	22.28	-1.90	-0.97
RAB27B	RAB27B, member RAS oncogene family	22.01	-1.77	-1.28
TSPAN18	tetraspanin 18	21.79	-1.84	-0.89
MEF2C	myocyte enhancer factor 2C	21.31	-0.84	-0.77
SHANK3	SH3 and multiple ankyrin repeat domains 3	21.22	-2.15	-0.77
IL17RE	interleukin 17 receptor E	20.76	-1.24	-0.50
DLG5	discs, large homolog 5 (Drosophila)	19.57	-1.52	-0.54

ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein	19.51	-0.43	-1.04
SMAGP	small cell adhesion glycoprotein	19.11	-1.71	-1.10
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	18.74	-2.21	-0.90
PLCH1	phospholipase C, eta 1	17.92	-1.14	-0.85
MFSD2A	major facilitator superfamily domain containing 2A	17.89	-0.88	-0.88
ARID5B	AT rich interactive domain 5B (MRF1-like)	17.43	-2.03	-1.07
ZBTB8A	zinc finger and BTB domain containing 8A	17.12	-1.01	-1.05
STYK1	serine/threonine/tyrosine kinase 1	14.26	-1.52	-0.70
LRP4	low density lipoprotein receptor-related protein 4	14.20	-1.52	-0.57
TRAF5	TNF receptor-associated factor 5	13.91	-1.53	-1.05
SMAD6	SMAD family member 6	13.73	-0.77	-0.53
CD109	CD109 molecule	13.72	-1.23	-1.70
PSD3	pleckstrin and Sec7 domain containing 3	13.31	-0.77	-0.65
ST8SIA6	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 6	12.73	-3.18	-1.12
PLK3	polo-like kinase 3	12.40	-1.17	-0.58
PAG1	phosphoprotein associated with glycosphingolipid microdomains 1	11.13	-2.38	-1.68
MIR155	microRNA 155	10.57	-0.70	-0.99
ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide	10.50	-0.93	-1.23
SCN1B	sodium channel, voltage-gated, type I, beta subunit	10.40	-3.24	-1.16
ADAMTS15	ADAM metallopeptidase with thrombospondin type 1 motif, 15	10.26	-3.17	-0.88
OLFML2A	olfactomedin-like 2A	8.72	-3.88	-1.82
FOSB	FBJ murine osteosarcoma viral oncogene homolog B	8.38	-0.55	-1.79
IL18R1	interleukin 18 receptor 1	7.53	-1.90	-1.78
ARAP2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2	7.44	-0.64	-1.05
SH2D2A	SH2 domain containing 2A	7.42	-1.23	-0.93
FAM65C	family with sequence similarity 65, member C	7.35	-3.42	-1.51
IL18RAP	interleukin 18 receptor accessory protein	7.33	-2.08	-1.74
ETV4	ets variant 4	7.24	-2.33	-2.53
ZFPM1	zinc finger protein, multitype 1	7.08	-1.34	-0.85
CCR4	chemokine (C-C motif) receptor 4	6.73	-4.45	-2.89

GAS1	growth arrest-specific 1	6.38	-1.30	-1.38
CDKN1C	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	6.35	-3.45	-2.30
TMCC2	transmembrane and coiled-coil domain family 2	6.10	-1.65	-1.26
ADAMTS19	ADAM metallopeptidase with thrombospondin type 1 motif, 19	6.04	-1.85	-0.74
FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)	5.96	-5.70	-2.35
AHR	aryl hydrocarbon receptor	5.92	-1.06	-1.10
CXCL2	chemokine (C-X-C motif) ligand 2	5.91	-2.93	-0.72
TIGD4	tigger transposable element derived 4	5.41	-1.59	-1.03
AKAP5	A kinase (PRKA) anchor protein 5	4.99	-1.88	-0.95
MMP28	matrix metallopeptidase 28	4.80	-4.00	-1.88
P2RX7	purinergic receptor P2X, ligand-gated ion channel, 7	4.72	-1.90	-1.02
HEMGN	hemogen	4.69	-5.52	-3.46
CXCL3	chemokine (C-X-C motif) ligand 3	4.21	-2.29	-1.60
F3	coagulation factor III (thromboplastin, tissue factor)	4.02	-1.18	-1.19
RGS9	regulator of G-protein signaling 9	3.19	-3.54	-1.43

Supplementary Table 2. Genes Downregulated by the AETFC Complex.

Gene symbol	Full name	Expression in Kasumi-1 cells (RPKM)	Fold change (log2)	
			shAE	shLDB1
RASGRP2	RAS guanyl releasing protein 2 (calcium and DAG-regulated)	25.96	1.05	0.67
LAPTM5	lysosomal protein transmembrane 5	24.14	1.25	0.94
SYTL1	synaptotagmin-like 1	22.07	1.24	0.60
VAV1	vav 1 guanine nucleotide exchange factor	20.65	1.15	0.70
BPI	bactericidal/permeability-increasing protein	18.00	1.79	0.75
AP5B1	adaptor-related protein complex 5, beta 1 subunit	17.88	1.48	1.36
NDRG1	N-myc downstream regulated 1	16.84	0.74	1.11
HCST	hematopoietic cell signal transducer	16.05	1.23	0.93
NCF2	neutrophil cytosolic factor 2	15.63	1.00	0.99
CST7	cystatin F (leukocystatin)	14.42	1.89	1.18
PSD4	pleckstrin and Sec7 domain containing 4	14.06	1.37	0.77
PTPN22	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	13.27	1.00	1.16
CCDC88B	coiled-coil domain containing 88B	12.57	1.35	0.86
CHST12	carbohydrate (chondroitin 4) sulfotransferase 12	12.42	1.07	0.68
SLA	Src-like-adaptor	12.35	1.54	1.11
SMPD1	sphingomyelin phosphodiesterase 1, acid lysosomal	12.20	0.87	1.19
ANXA1	annexin A1	12.01	1.85	1.76
PLEC	plectin	11.99	1.01	0.95
RASAL3	RAS protein activator like 3	11.91	1.25	1.16
ALDH3B1	aldehyde dehydrogenase 3 family, member B1	11.73	1.60	1.27
TP53TG1	TP53 target 1 (non-protein coding)	11.03	1.35	1.08
CD82	CD82 molecule	9.19	2.16	1.28
H6PD	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)	9.08	0.97	1.03
NUDT18	nudix (nucleoside diphosphate linked moiety X)-type motif 18	9.07	1.04	0.82
SLC48A1	solute carrier family 48 (heme transporter), member 1	8.85	1.26	0.63
PPM1M	protein phosphatase, Mg2+/Mn2+ dependent, 1M	7.98	1.40	1.07
LDHD	lactate dehydrogenase D	7.81	1.30	1.52

ADAMTS20	ADAM metallopeptidase with thrombospondin type 1 motif, 20	7.73	1.05	1.07
PTPN12	protein tyrosine phosphatase, non-receptor type 12	7.53	1.13	1.07
C1orf162	chromosome 1 open reading frame 162	7.44	2.10	1.46
GPR77	G protein-coupled receptor 77	7.13	1.48	1.54
ATP10D	ATPase, class V, type 10D	7.06	1.04	1.06
ARHGEF17	Rho guanine nucleotide exchange factor (GEF) 17	7.00	1.15	1.03
SPNS3	spinster homolog 3 (Drosophila)	6.94	2.24	0.68
TMEM53	transmembrane protein 53	6.57	1.34	0.77
ZC3H12D	zinc finger CCCH-type containing 12D	6.45	1.14	0.82
OSCAR	osteoclast associated, immunoglobulin-like receptor	6.19	0.91	1.84
ERN1	endoplasmic reticulum to nucleus signaling 1	6.00	1.39	0.87
OPRL1	opiate receptor-like 1	5.99	1.10	1.35
FLJ20021	uncharacterized LOC90024	5.96	1.03	1.02
SIRPB1	signal-regulatory protein beta 1	5.87	0.98	1.63
BHLHE40	basic helix-loop-helix family, member e40	5.82	1.01	0.66
GHRL	ghrelin/obestatin prepropeptide	5.79	1.22	1.44
TMEM187	transmembrane protein 187	5.71	1.35	1.34
PCSK4	proprotein convertase subtilisin/kexin type 4	5.67	1.27	1.12
RHPN1	rhophilin, Rho GTPase binding protein 1	5.36	1.12	0.97
EPX	eosinophil peroxidase	5.10	1.90	1.09
RIPK3	receptor-interacting serine-threonine kinase 3	4.96	1.36	1.08
DNAJC28	DnaJ (Hsp40) homolog, subfamily C, member 28	4.79	1.01	1.07
PVRL1	poliovirus receptor-related 1 (herpesvirus entry mediator C	4.79	1.17	1.18
NIPAL2	NIPA-like domain containing 2	4.69	1.58	0.69
FAM151B	family with sequence similarity 151, member B	4.57	1.10	0.84
GBGT1	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1	4.51	1.66	0.47
DENND2D	DENN/MADD domain containing 2D	4.51	1.63	0.99
HIST1H2AC	histone cluster 1, H2ac	4.49	1.64	1.96
FBLN5	fibulin 5	4.00	2.08	0.42
AOAH	acyloxyacyl hydrolase (neutrophil)	3.77	1.28	0.75
SLC25A45	solute carrier family 25, member 45	3.58	2.19	0.97
TFAP2E	transcription factor AP-2 epsilon (activating enhancer binding protein 2 epsilon)	3.53	1.09	1.34
NLRP3	NLR family, pyrin domain containing 3	3.53	1.18	1.96

KIAA1407	KIAA1407	3.42	1.19	1.17
LRRC23	leucine rich repeat containing 23	3.38	1.10	0.98
SLC16A3	solute carrier family 16, member 3 (monocarboxylic acid transporter 4)	3.30	1.87	1.61
ZCWPW1	zinc finger, CW type with PWWP domain 1	3.19	1.13	0.60
ZMYND12	zinc finger, MYND-type containing 12	3.08	1.06	0.85
DLX4	distal-less homeobox 4	3.06	1.39	0.81
CYTH4	cytohesin 4	3.05	1.71	0.60
SPAG4	sperm associated antigen 4	3.02	1.65	0.86
FAM116B	family with sequence similarity 116, member B	3.01	1.19	1.41
FLJ39639	uncharacterized protein FLJ39639	2.91	1.15	0.97
CLEC12A	C-type lectin domain family 12, member A	2.76	3.14	3.56
C1QTNF6	C1q and tumor necrosis factor related protein 6	2.74	1.13	1.21
RNASE3	ribonuclease, RNase A family, 3	2.65	3.34	1.83
NPAS1	neuronal PAS domain protein 1	2.60	2.91	1.18
AVPR2	arginine vasopressin receptor 2	2.54	2.30	1.42
SELPLG	selectin P ligand	2.53	3.77	2.01
KLF7	Kruppel-like factor 7 (ubiquitous)	2.51	1.86	1.52
ANXA3	annexin A3	2.50	2.50	1.96
LRRC25	leucine rich repeat containing 25	2.48	2.07	2.23
S100A8	calcium ion binding	2.38	3.00	3.04
UPP1	uridine phosphorylase 1	2.36	1.57	0.69
TNFRSF14	tumor necrosis factor receptor superfamily, member 14	2.34	1.83	1.32
HIST2H2BE	histone cluster 2, H2be	2.33	1.14	1.52
MGLL	monoglyceride lipase	2.32	2.59	0.99
BANK1	B-cell scaffold protein with ankyrin repeats 1	2.25	2.62	0.93
ADHFE1	alcohol dehydrogenase, iron containing, 1	2.03	1.28	2.01
NEK11	NIMA (never in mitosis gene a)- related kinase 11	2.00	1.15	0.97
MXRA7	matrix-remodelling associated 7	1.98	1.26	1.33
CAPN5	calpain 5	1.95	1.00	1.81
SNAI3	snail homolog 3 (Drosophila)	1.89	1.21	0.59
FBN1	fibrillin 1	1.88	3.45	2.37
OGG1	8-oxoguanine DNA glycosylase	1.86	2.45	1.33
WNT5B	wingless-type MMTV integration site family, member 5B	1.85	1.05	1.21
PRAM1	PML-RARA regulated adaptor molecule 1	1.80	3.65	2.09

Supplementary Table 3. X-ray Statistics of AML1-ETO/HEB Complex.

Data collection and refinement statistics				
Crystal	AML-ETO(482-548) /HEB(177-200)	AML-ETO /HEB(V186M)	AML-ETO /HEB(V194M)	AML-ETO /HEB(V186M/V194M)
Beam line	APS-24ID-E	APS-24ID-C	APS-24ID-C	BNL-X29
Wavelength	0.97919	0.97919	0.97919	0.97919
Space group	$P3_2$	$P3_2$	$P3_2$	$P3_2$
Unit cell				
a, b, c (Å)	139.5,139.5,43.2	139.7,139.7,43.4	139.4,139.4,42.8	139.0,139.0,42.8
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	30-2.9 (3.0-2.9)[a]	30-3.2 (3.27-3.21)	30-3.36 (3.42-3.36)	30-3.25(3.31-3.25)
R_{sym}	0.072 (0.722)	0.068 (0.528)	0.097 (0.609)	0.060 (0.603)
I/σ (I)	47.9 (3.4)	25.4 (2.2)	17.1 (1.9)	27.9 (2.3)
Completeness (%)	99.0 (100.0)	99.9 (99.4)	99.8 (99.9)	99.9 (99.4)
Redundancy	9.3 (9.5)	3.1 (3.2)	3.0 (3.0)	2.8 (2.8)
Number of unique reflections	20470	31172	26580	29025
R_{work}/R_{free} (%)	20.7/25.2			
Number of non-H atoms				
Protein	2425			
Other ligands	0			
Average B factors (Å^2)				
Protein	102.9			
Other ligands	none			
R.m.s. deviations				
Bond lengths (Å)	0.009			
Bond angles (°)	1.40			

[a] Highest resolution shell (in Å) shown in parentheses.

Supplementary Discussion

Apart from revealing new mechanisms by which AML1-ETO functions through AETFC in leukemia, as described in the main text, this study also has addressed two specific questions in the field. The first question concerns the factors/events that lie directly downstream of the functionally critical AML1-ETO NHR2 dimerization. Although earlier studies had suggested that the NHR2 domain of the ETO protein might contribute to the association of ETO with corepressors (e.g., NCoR/SMRT, mSin3A, HDACs)[38,39], a recent study showed that disruption of NHR2 dimerization does not affect the ability of AML1-ETO to interact with the corepressors[4]. In contrast, we have provided definitive evidence that NHR2 dimerization is essential for formation of a binding interface for the N2B motif of E proteins and, further, that this N2B interaction is required for AML1-ETO-induced haematopoietic stem cell self-renewal and leukemogenesis. These data thus suggest that the NHR2dimer-N2B interaction occurs as a consequence of the AML1-ETO dimerization and contributes to leukemogenesis. The second question is whether the interaction between AML1-ETO and E proteins is important for leukemogenesis. We previously found that the NHR1 domain of the ETO protein interacts with the AD1 domain of E proteins and that this interaction allows ETO to inhibit E protein-mediated transcription[14]. However, subsequent functional studies showed that disruption of this NHR1-AD1 interaction or deletion of NHR1 does not significantly affect the leukemogenic activity of AML1-ETO in the assays employed[5,6,26]. While leaving open the likely possibility that the NHR1-AD1 interaction could be relevant to the regulation of E protein target genes in cells that normally express ETO, these results raised a question concerning the contribution of the AML1-ETO–E protein interaction to leukemogenesis. Here, the identification of the AETFC complex provides a new perspective for understanding the role of the E proteins in AML1-ETO–associated leukemia. Thus, our data show unequivocally that E proteins play a critical role in stabilizing AML1-ETO/AETFC and in regulating gene expression and leukemogenesis. Notably, among the two pairs of the AML1-ETO–E protein interactions, the NHR2-N2B interaction, but not the NHR1-AD1interaction, requires NHR2 dimerization and may act as a conformational switch within AETFC and

contribute to AML1-ETO–mediated leukemogenesis. Thus, these results are in agreement with previous studies and, from a new perspective, address the unanswered questions and provide new insights into the mechanism(s) of action of AML1-ETO in leukemogenesis.

Regarding the diverse functions of AML1-ETO in both gene activation and repression, there seems little doubt that there must be a context-dependent mechanism(s) to appropriately govern the balance between coactivator (e.g., p300) and corepressor (e.g., HDACs) interactions and that it is of great importance to understand this mechanism[17,18,40,41]. The p300 and HDAC interactions with AML1-ETO are much weaker than the interactions between AETFC components. Our AETFC complex purification employed high stringency conditions (buffer containing detergent and 300 mM KCl) under which the p300 and HDAC entities do not associate with the AETFC complex. In contrast, by using a lower salt concentration (150 mM KCl), both p300 and HDACs were detected in the AML1-ETO immunoprecipitate (data not shown). In this regard, it was important to determine whether the formation of AETFC plays a role in the recruitment of p300 and HDACs. To address this question, we performed *in vitro* interaction assays with purified/recombinant AML1-ETO, AETFC, p300 and NCoR/HDAC3 complex. The results showed that, compared with AML1-ETO alone, the AETFC complex exhibits significantly stronger binding to p300 (Supplementary Fig. 14a). Furthermore, and related, our ChIP-seq analysis showed that the p300 binding scores are higher on the genomic sites that are co-bound by AML1-ETO and E proteins relative to the AML1-ETO–alone sites (Supplementary Fig. 14b, c). In contrast to the situation with p300 binding, AML1-ETO and AETFC showed a similar ability to interact with the NCoR/HDAC3 complex (Supplementary Fig. 14d). These observations suggest that AETFC components may differentially contribute to p300 and HDAC recruitment (for example, that p300 is recruited by the complete AETFC complex whereas HDACs may be recruited by AML1-ETO *per se*) and, thus, that the AETFC complex may serve as a key determinant in regulating the balance between p300 and HDAC function on AML1-ETO target genes. While the mechanisms need to be further verified by detailed studies,

our identification and characterization of the AETFC complex provides a completely new

perspective on this complex problem and will pave the way for further studies toward a better

understanding of AML1-ETO–mediated gene regulation.

References

38. Zhang, J. *et al.* Oligomerization of ETO is obligatory for corepressor interaction. *Mol. Cell Biol.* **21**, 156-163 (2001).
39. Amann, J. M. *et al.* ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol. Cell Biol.* **21**, 6470-6483 (2001).
40. Ptasinska, A. *et al.* Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* **26**, 1829-1841 (2012).
41. Saeed, S. *et al.* Chromatin accessibility, p300, and histone acetylation define PML-RARalpha and AML1-ETO binding sites in acute myeloid leukemia. *Blood* **120**, 3058-3068 (2012).