## SUPPLEMENTS

### Supplementary Methods

#### *Human primary cells*

All patient material was obtained with the required ethical approval from the NHS Research Ethics Committees (Leeds Teaching Hospitals NHS Trust and Newcastle upon Tyne Hospitals NHS Foundation Trust). Initial analysis of AML patient samples was carried out by the Haematological Malignancy Diagnostic Service (St James's Hospital, Leeds), where cytogenetic abnormalities and sample immunophenotype were determined at the time of disease diagnosis. Several sets of 6–colour FACS staining were performed on a FACSCanto or FACSCanto II flow cytometer using antibodies obtained from BD Bioscience (Oxford, UK) as follows: Anti-CD34 - APC / PerCP:Cy5.5 (clone 8G12), Anti-CD117 – PE / PE:Cy7 (clone 10402), Anti-CD45 – PerCP:Cy5.5 / APC:Cy7 (Clone 2D1), Anti-CD15 – FITC (clone MMA), Anti-CD13 – PE (clone L138), Anti-HLADR – APC:Cy7 (Clone L243 (G46-6)), Anti-CD33 – PE (clone P37.6), Anti-CD7 – FITC (clone M-T701), Anti-CD19 – PerCP:Cy5.5 (clone SJ25C1), Anti-CD56 – APC (clone NCAM16.2), Anti-CD14 – FITC (clone MφP9), Anti-CD38-PE:Cy7 (clone HB7) and Anti-CD64 – PE (clone MD22). The result of flow-cytometry analysis is shown in Supplementary Table 2. The presence of the t(8;21) translocation and expression of mRNA encoding the RUNX1/ETO fusion protein were determined using RT-PCR, with the primers 5'-TCAAAATCACAGTGGATGGGC and 5'-CAGCCTAGATTGCGTCTTCAC A.

Mononuclear cells freshly obtained from patients #1 and #2 and from the patient with the KN-AML were prepared by differential centrifugation using Lymphoprep (Axis-Shield UK, Cambridgeshire, UK), and $CD34^+$ blast cells were then isolated using MACS Micro Beads staining and separation on magnetic columns according to the manufacturer's guidelines. Purity of $CD34^+$ cells was >85%. Leukemic blast cells from t(8;21) patients A and B were purified from white blood cells and neutrophils by Ficoll–Hypaque density-gradient centrifugation. The sample was frozen in liquid nitrogen with 10%

dimethyl sulphoxide in 20% FCS, 10% DMSO, RPMI. After thawing, cells were washed and immediately used for ChIP.

***Real-time RT-PCR***

Reverse transcription was performed was carried out either with oligo-dT or random hexamers using MMLV RNase (Invitrogen) as suggested by the manufacturer. Real time PCRs were performed using Sybr-Green Mix (Applied Biosystems, Warrington, UK), 100 nM primers and 20% v/v diluted RT reaction mix using standard conditions on a 7500 Sequence detection system (Applied Biosystems, Foster City, CA, USA). Primers are listed in **Supplementary Table 3**.

***Western Blotting***

Western blotting was performed as described (1). The following antibodies were applied: ETO (C-20, Santa Cruz Biotechnology, Santa Cruz, CA, USA), RUNX1 (AML1-RHD, PC-285, EMD Chemicals), β-actin (AC-15**,** Sigma-Aldrich UK) and GAPDH (5G4, HyTest, Turku, Finland).

***Chromatin immunoprecipitation***

The cells were harvested and then resuspended to 2 x $10^7$ in 10 ml of growing medium, and cross-linked with 1% formaldehyde (Pierce) for 10 min at RT. The cross-linking reaction was stopped by adding glycine to a final concentration of 0.4 M, followed by two washes with ice–cold PBS. Cells were resuspended in 10 ml of ice-cold ChIP buffer A (10 mM HEPES pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, proteinase inhibitor cocktail (Roche UK, Burgess Hill, UK) and 0.1 mM PMSF), incubated for 10 min at 4°C with rotation, and centrifuged 5 min at 500 x g at 4 °C. The pellet was resuspended in 10 ml of ice–cold ChIP buffer B (10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100, protease inhibitor cocktail and 0.1 mM PMSF), incubated for 10 min at 4 °C with rotation and centrifuged for 5 min at 500 x g at 4 °C. Cells were resuspended in 600 µl of ice-cold ChIP lysis buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.25% SDS, protease inhibitor cocktail and 0.1 mM PMSF), incubated 10 min on ice and sonicated at

5 °C using a Bioruptor™ (Diagenode, Liege, Belgium) to generate fragments an average length of 400-500 bp (10 min with 30 s “ON” and “OFF” cycles, power setting high). The lysates were centrifuged for 5 min at 16,000 x g at 4 °C and the supernatants were diluted with two volumes of ice-cold ChIP dilution buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 7.5% glycerol, protease inhibitor cocktail and 0.1 mM PMSF).

For each IP, 15 µl of Dynabeads® protein G were pre–incubated with 50 µg BSA and 2 µg antibody against ETO (Santa Cruz, sc-9737X), RUNX1 (Abcam, ab23980), RNA-Polymerase II phospho S2 (ab5095) and H3K9Ac (ab4441) for 2 h at 4 °C with rotation. The blocked antibody-bound protein G mix was added to 20–25 µg chromatin in a total volume of 500 µl diluted ChIP lysis buffer and incubated for 2 h at 4°C with rotation. After magnetic separation the beads were washed once with 1 ml wash buffer 1 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), twice with 1 ml wash buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), once with 1 ml LiCl buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Na-deoxycholate) and twice with 1 ml TE/NaCl buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA). For each wash the beads were mixed with ice-cold washing buffers for 10 min at 4 °C. The immunoprecipitated DNA was eluted two times with 50 µl ChIP elution buffer (100 mM NaHCO3, 1% SDS) for 15 min at RT with shaking. At this step the input control (1% of the starting material) was included in the experimental procedure after first adjusting the final volume to 100 µl with ChIP elution buffer. The eluted DNA was incubated overnight at 65 °C in the presence of 50 µg proteinase K. The DNA was finally purified using Agencourt® AMPure® (Beckman Coulter) magnetic beads according to the manufacturer’s instructions, eluted with 50 µl 0.1 x TE and analyzed by qPCR. Primers are listed in **Supplementary Table 3.**

***DNase I hypersensitive site mapping***

DNase I digestions were carried out on permeabilised $CD34^+$ cells as previously described (2). Briefly, cells were suspended at a concentration of 3 x $10^7$ cells/ml in nuclei digestion buffer (60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris pH7.4, 300

mM glucose). Digestions were then performed in 2 – 6 µg/ml DNase I (Worthington, DPPF grade) at 22ºC for 3 minutes, by adding the enzyme in an equal volume of digestion buffer containing 2 mM $CaCl_2$ and 0.4% Nonident P-40.

Genomic DNA was extracted from DNase I-treated cells using phenol/chloroform extraction, and then run out on 0.8% agarose gels. Levels of DNase I digestion were assessed using real time PCR, measuring the ratio of presence of known DNase I hypersensitive regions to more resistant gene free regions. Sequences of real time PCR primers used were, for the 'active' region, TBP promoter: 5'-CTGGCGGAAGTGACAT TATCAA and 5'- GCCAGCGGAAGCGAAGTTA; and for the 'Inactive' region, a gene-free control region of chromosome 18: 5'- ACTCCCCTTTCATGCTTCTG and 5'- AGGT CCCAGGACATATCCATT. Fragments in the range of 100-600 bp in size, from samples with similar low levels of DNase I digestion, were excised and purified from gel slices in preparation for sequencing library generation.

***Library preparation***

Libraries of DNA fragments from chromatin immunoprecipitation or DNase I treatment were prepared from approximately 10 ng of DNA. Firstly, overhangs were repaired by treatment of sample material with T4 DNA polymerase, T4 PNK and Klenow DNA polymerase (all enzymes obtained from New England Biolabs UK) in a reaction also containing 50 mM Tris-HCl,10 mM $MgCl_2$, 10 mM Dithiothreitol, 0.4 mM dNTPs and 1 mM ATP. Samples were purified after each step using Qiagen MinElute columns (according to the manufacturer's guidelines). Adenosine bases were added to 3' ends of fragments using Klenow Fragment (3´- 5´ exo-minus), allowing for subsequent ligation of adapter oligonucleotides (Illumina part #1000521) using Quick T4 DNA ligase. After a further column clean up to remove excess adaptors, fragments were amplified in an 18 cycle PCR reaction using adapter-specific primers (sequences 5'-CAAGCAGAAGACGGCATACGAGCTCTTCCGATC*T and 5'-AATGATACGGCGACCG AGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T). The libraries were purified and adapter dimers removed by running PCR products on 2% agarose gels and excising gel slices corresponding to fragments approximately 200-300 bp in size, which were then extracted using the Qiagen gel extraction kit. Libraries were validated using

quantitative PCR for known targets, and quality assessed by running 1 µl each sample on an Agilent Technologies 2100 Bioanalyser. Libraries from cells from patients A and B and their respective input DNA were preparing using a kit for paired end sequencing. Once prepared, DNA libraries were subject to massively parallel DNA sequencing on an Illumina Genome Analyzer.

***Plasmid construction***

To analyze the cis-regulatory sequences identified as RUNX1 and/or RUNX1/ETO binding sites in ChIP-seq, reporter constructs were generated in the pGL4-Basic plasmid (Promega UK) with promoter sequences or in the pGL4-TK plasmid (Promega) with distal sequences. Regions of interest were amplified from purified genomic DNA from Kasumi-1 cells by PCR with primers containing consensus sequences for *Kpn*I and *Nhe*I restriction enzymes (promoter sequences) or *BamH*I and *Sal*I restriction enzymes (enhancer sequences) using Extensor Hi-Fidelity PCR Master Mix (Thermo Fisher Scientific Inc) under the following conditions: 94°C for 2 min, followed by 35 cycles 94°C for 30 s, 60°C for 30 s, and 68°C for 2 min and 68°C for 7 min. Primers are listed in **Supplementary Table 3**. PCR products were gel purified, cut with the restriction enzymes and ligated into pGL4-Basic vector (promoter sequences) or pGL4-TK vector (enhancer sequences) using Quick T4 DNA ligase obtained from New England Biolabs UK. All constructs were verified by sequencing. Plasmid DNAs were prepared from these constructs using the Plasmid Maxi kit (Qiagen).

***Transient transfections and luciferase gene reporter assay***

Transient transfections of RAW264.7 cells were conducted using the Lipofectamine™ 2000 Reagent (Invitrogen) according to the manufacturer's protocol. The cells were grown in 12 well plates and transfected at 80% confluence. 0.6 µg of each expression vector, 0.2 µg of each studied construct and 0.05 µg of pRL-SV40 control vector (Promega) was added to transfection mixes. Luciferase assays were performed using a dual-luciferase reporter assay system (Promega). Cells were harvested 24 h after transfection, lysed with 100 µl of passive lysis buffer and measurements performed

using G Berthold luminometer (Lumat LB 9507). Firefly luciferase activity was normalized to Renilla activity to account for transfection efficiency. The transfections were always performed in triplicate.

***Identification of DNaseI- and ChIP-sequencing peaks.***

The raw sequence data returned by the Illumina Pipeline was aligned to the hg18 assembly (NCBI Build 36.1) using BWA (3), and data was displayed using the UCSC Genome Browser (4). The raw paired end reads from patient A and patient B have been also aligned using BWA. There are two runs per patient; those have been combined before peak detection. Regions of enrichment (peaks) of DNaseI and ChIP data were identified using MACS software (5). A dynamic local lambda option was calculated by MACS for every peak region, with $1x10^{-5}$ as a "p-value" cutoff, and a default "-mfold" range. A control input dataset was used with ChIP-Seq data and the "-nomodel" flag was used for DNaseI-seq data.

Several of the analyses in the paper require the comparison of peak sets obtained from different experiments and the enumeration of peaks that overlap between sets. Two effects make the comparison of peak sets difficult. First, sequencing experiments produce varying total numbers of reads: with a larger number, peaks of the same relative height are more statistically significant and smaller peaks can move above a chosen threshold. Second, background noise varies between experiments, resulting in lower statistical significance for peaks of the same height when noise is higher. While it is difficult to correct for these effects in a completely principled way, we adopted the pragmatic approach of using a consistent statistical significance cut-off (above) and also using a cut-off on peak height to remove the smallest peaks from data sets with high read count. We measured the height of a peak as the number of reads within it divided by the total number reads in peaks (to account for varying noise), and peaks below threshold values of $8x10^{-6}$ and $2x10^{-5}$ for DNaseI-Seq peaks and ChIP-Seq peaks, respectively, were removed from the analysis.

In all cases two peaks were deemed to overlap if the centre of one lay within the boundaries of the other. In practice this is a stringent definition of overlap, allowing no extension beyond the delineated boundary of each peak, but we found by

experimentation that small relaxations of this criterion made little difference to the answer.

HOMER (6) was used to analyse the genomic regions occupied by peaks, dividing the genome into promoter regions (transcription start site (TSS) +/- 500bp in this case), exons, introns and intragenic regions. HOMER was also used to produce the data for scatter plots, comparing DNaseI-seq read counts (log2) under a defined peak set in different cell types. Custom programs were written to produce distributions of distances between peaks and the nearest TSS. In cases where analysis was carried out at the level of individual genes, each peak was associated to the gene whose TSS was closest, allowing the association of more than one peak to each gene.

Analysis of H3K9Ac profiles was performed using Seqminer (7). RUNX1/ETO high confidence peak regions were used as reference coordinates against all aligned reads for the acetylation and RUNX1/ETO. Mean read density profiles were produced for 2 clusters, generated using the default K-means method.

***Motif analysis***

*De novo* motif analysis was performed using HOMER (6). Motifs of length 8, 10, and 12 bp were identified on the peak sequences of length ± 100 bp from the peak summit. A random background sequence option matching the CpG characteristics of the input sequence was used and the CpG percentage in both the target and background sequences was used for sequence content normalisation. The motif matrices generated by HOMER were scanned against JASPAR with the use of STAMP (8) to identify similarity to known transcription factor binding sites. The top enriched motifs with significant log p value score were recorded. The annotatePeaks function in HOMER was used to find occurrences of motifs in peaks using cut-off values (a log odds score of 6). In this case we used the discovered motif PWM with the most significant log p value, although use of other discovered PWMs or the corresponding PWM from Jasper produced very similar results.

***Microarray gene expression data analysis***

The microarray gene expression data were analysed in GenomeStudio software (Illumina, Little Chesterford, UK) with background subtraction. The raw data output by GenomeStudio was analysed using the Lumi R package (9) with quantile normalisation. The 10% threshold (p value <= 0.1) was applied to all data. Genes with at least two fold-change in expression (either up or down) were selected.

Clustering of gene expression time courses was carried out only for genes associated with at least a two-fold change at least one time point and correlated with RUNX1/ETO genes. Hierarchical clustering was used with Euclidean distance and average linkage clustering. Alternative clustering methodology and parameters were investigated and these were found to give the most biologically interpretable result. In particular, correlation based clustering was less successful, presumably because four time points is too small a number to effectively estimate a correlation. Heatmaps were generated using Mev from TM4 microarray software suite (10).

Gene expression was correlated with both RNA polymerase II occupancy and H3K9 acetylation measured by ChIP-seq. Since RNA polymerase II can bind anywhere in the transcribed region of the gene, changes in occupancy were estimated as ratio of read counts in peaks anywhere in the transcribed region, between control and RUNX1/ETO knockdown conditions. This analysis was carried out only for direct targets of RUNX1/ETO. In the case of H3K9 acetylation changes were estimated as the ratio of read counts under the relevant RUNX1/ETO binding peak. Heatmaps were produced by using clusters from the expression only analysis and displaying changes in acetylation and RNA polymerase II occupancy levels alongside in descending order.

NMF (Non-negative Matrix Factorization) analysis was used to extract metagenes from illumina expression probe values and projected onto other datasets by the pseudo-inverse method as outlined in (11, 12). Briefly, probes below intensity <20 or >10000 were designated out of range and only probes with a range of 100 between minimum and maximum and a fold change of 2 were considered. All probes were column rank normalised. Where required, Illumina probe IDs were mapped to Affymetrix IDs according to manufacturers designated best matches. Two Metagenes derived from our own data were projected across a time course of siRNA-treated SKNO-1 cells, a RUNX1/ETO knock-down study using patient-derived t(8;21)-AML blasts and three

independent expression profiling studies by Corsello *et al.* (13), Mulloy *et al*. (14), and Tonks *et al.* (15) from GEO Datasets GSE15648, GSE8023 and Arrayexpress Dataset mexp583 respectively. All Affymetrix .cel files were processed and normalised using RMA (Robust Multiarray Average). All analysis and figures were generated using R/Bioconductor and by using a modified version of the NMF scripts provided by Brunet et al. (11).

In order to determine resemblance of Kasumi cells to normal hematopoietic sub-populations following *RUNX1/ETO* knock-down we projected the metagenes across the expression profiles from Novashtern *et al* GEO Dataset GSE2475 (16). The expression of the projected F1 metagene was taken to indicate the relative similarity of expression profiles for each hematopoietic sub-population.

GSEA was performed using the stand-alone application (Broad Institute) by ranking genes according to their correlation (Pearson) with the F1 metagene; the metagene highly expressed in cells infected with shRNA for *RUNX1/ETO*. GSEA was performed using the MSigDB gene set databases modified to add our list of CHIP-Seq binding targets as additional gene sets.

David/EASE analysis was performed using the online tool at david.abcc.ncifcrf.gov (17).

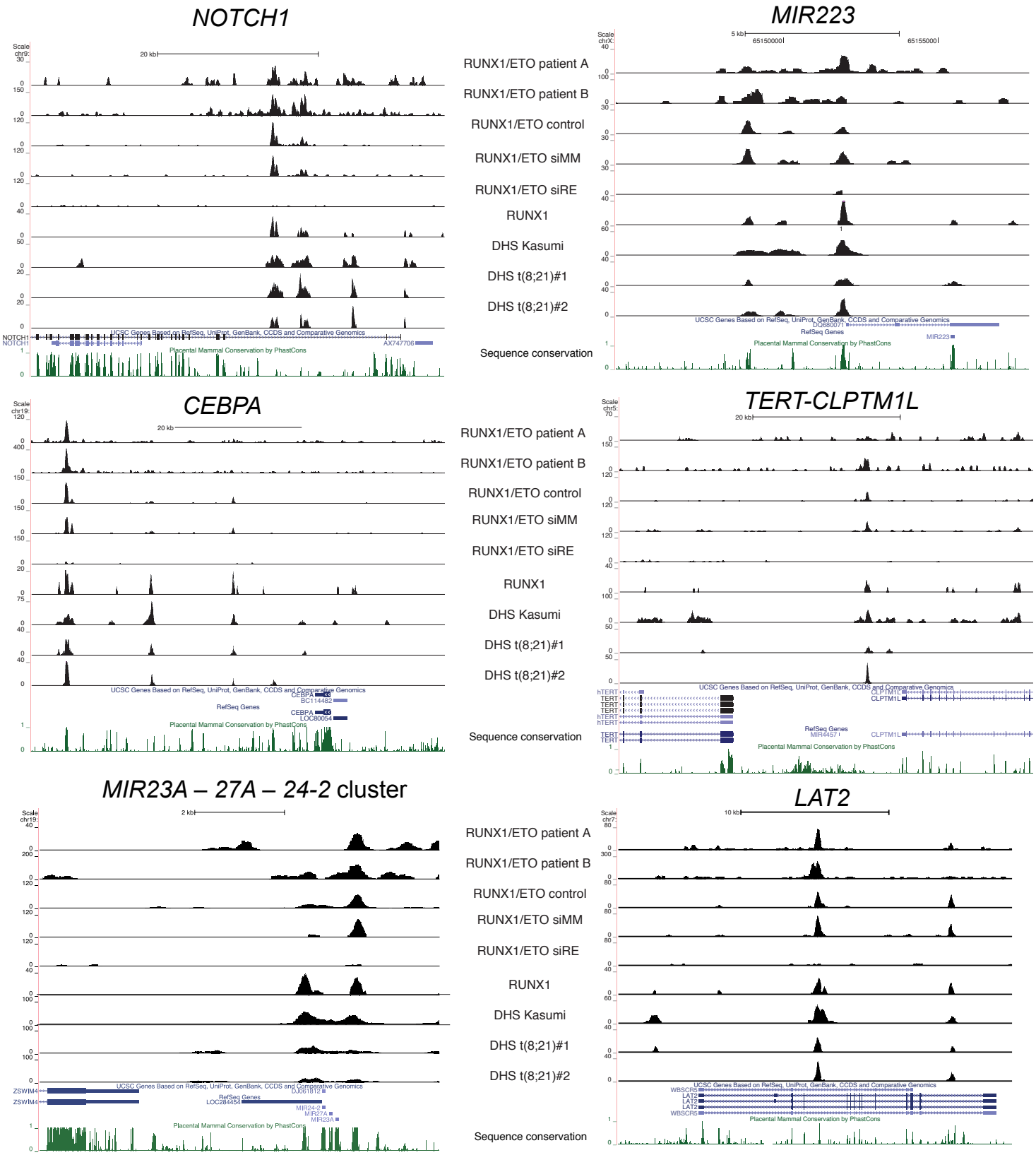
**Supplementary References**

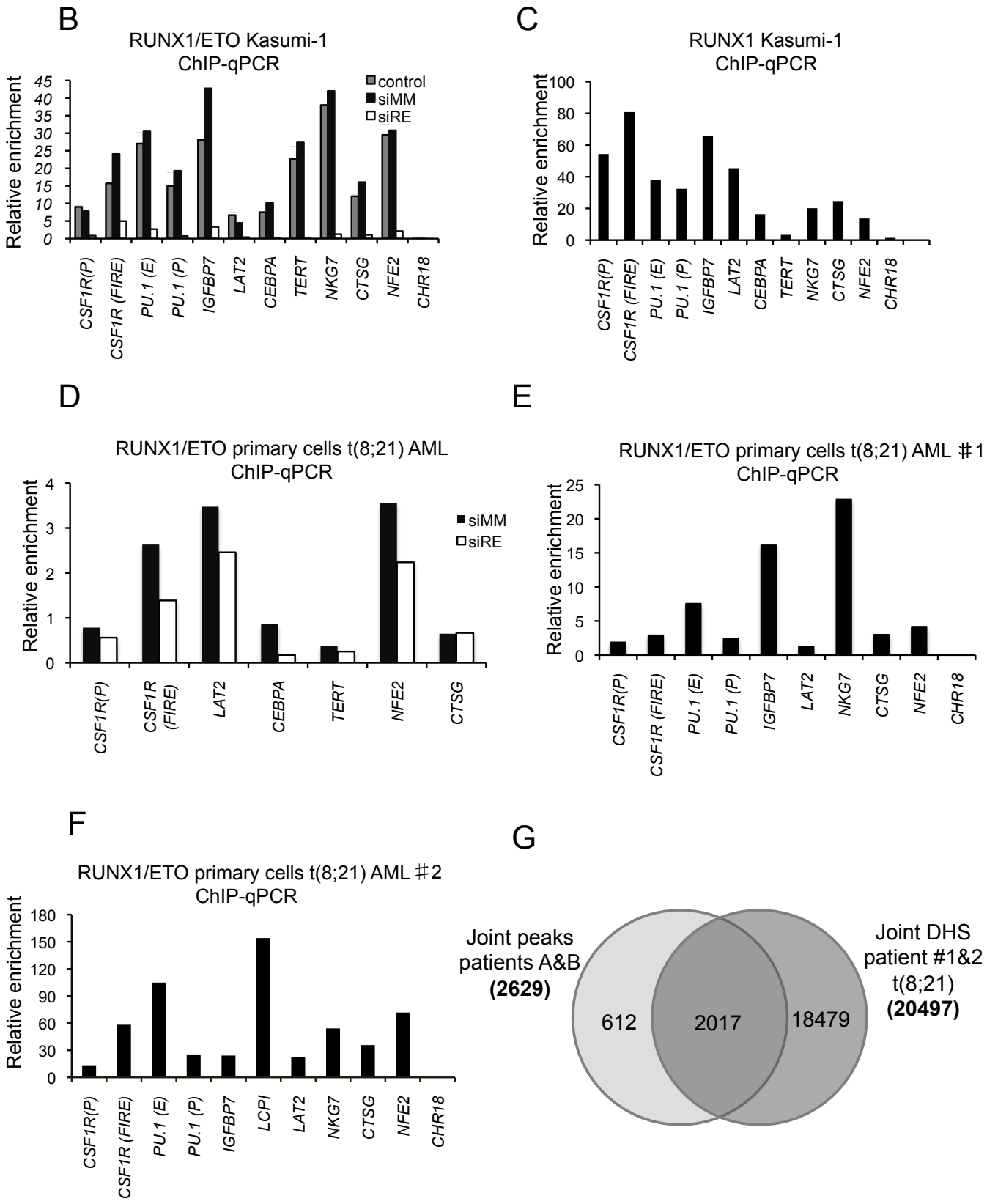

1. Dunne J, Cullmann C, Ritter M, Soria NM, Drescher B, Debernardi S, et al. siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. Oncogene. 2006;25(45):6067-78.
2. Bert AG, Johnson BV, Baxter EW, Cockerill PN. A modular enhancer is differentially regulated by GATA and NFAT elements that direct different tissue-specific patterns of nucleosome positioning and inducible chromatin remodeling. Mol Cell Biol. 2007;27(8):2870-85.
3. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2010;26(5):589-95.
4. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.
5. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology. 2008;9(9):R137.
6. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576-89.

7. Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, et al.. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 2011 Mar; 39(6):e35.
8. Mahony S, Benos PV. STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic acids research. 2007;35(Web Server issue):W253-8.
9. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics (Oxford, England). 2008;24(13):1547-8.
10. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. TM4 microarray software suite. Methods Enzymol. 2006;411:134-93.
11. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(12):4164-9.
12. Tamayo P, Scanfeld D, Ebert BL, Gillette MA, Roberts CW, Mesirov JP. Metagene projection for cross-platform, cross-species characterization of global transcriptional states. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(14):5959-64.
13. Corsello SM, Roti G, Ross KN, Chow KT, Galinsky I, DeAngelo DJ, et al. Identification of AML1-ETO modulators by chemical genomics. Blood. 2009;113(24):6193-205.
14. Mulloy JC, Jankovic V, Wunderlich M, Delwel R, Cammenga J, Krejci O, et al. AML1-ETO fusion protein up-regulates TRKA mRNA expression in human CD34+ cells, allowing nerve growth factor-induced expansion. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(11):4016-21.
15. Tonks A, Pearn L, Musson M, Gilkes A, Mills KI, Burnett AK, et al. Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. Leukemia. 2007;21(12):2495-505.
16. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell. 2011;144(2):296-309.
17. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.

# Supplementary Figure 1

A






**Supplementary Figure 1:**

**(A)** UCSC Genome Browser image of the human *NOTCH1*, *MIR223*, *CEBPA, TERT – CLPTM1L* and *MIR23A – 27A – 24-2* loci depicting ChIP-Seq tags for (top to bottom)

RUNX1/ETO in patient A and B, RUNX1/ETO in mock-treated Kasumi-1 cells (control), Kasumi-1 treated with mismatch control siRNA (siMM) and with RUNX1/ETO siRNA (siRE) for 48 hrs, RUNX1 in Kasumi-1 cells, as well as DHS profiles in Kasumi-1 cells and in CD34-positive blasts from two t(8;21) AML patients**. (B)** Validation of ChIP-seq results in Kasumi-1 cells by ChIP-qPCR**.** Real-time PCR (qPCR) analysis of RUNX1/ETO binding to cis-elements of *CSF1R* (promoter and intronic enhancer (FIRE))*, PU.1* (-14 kb enhancer and promoter)*, IGFBP7, LAT2, CEBPA, TERT, NKG7, CTSG* and *NFE2* in siRNA-treated and untreated Kasumi-1 cells. Primers specific for *CHR18,* a transcriptionally inactive region of chromosome 18, served as negative control (Supplementary Table 3). **(C)** Manual validation of RUNX1 Chip-seq peaks in Kasumi-1 cells. **(D)** Validation of the reduction in RUNX1/ETO binding after knock-down at the indicated loci in 34-positive blasts from a t(8;21) AML patient treated for 48 hrs with mismatch control siRNA (siMM) and with RUNX1/ETO siRNA (siRE) as compared by ChIP-qPCR. Values represent fold enrichment over non-specific binding at a negative control region at Chr 18. (**E, F)** Validation of RUNX1/ETO ChIP peaks in primary t(8;21) AML cells from two additional patients. ChIP-qPCR demonstrating excellent agreement of RUNX1/ETO binding at the examined loci between Kasumi-1 cells and blasts from two t(8;21) AML patients. **(G)** Venn diagram showing the overlap between joint DHS from leukemic blast cells of two different t(8;21) patients (patient #1 and #2) and peaks bound by RUNX1/ETO in both patient A and B.

# Supplementary Figure 2

A

**Motifs associated with high confidence RUNX1 peaks in Kasumi-1**

| motif | match | score (log p-value) |
|---|---|---|
|  | ERG/ FLI-1 | 3135 |
|  | ETS | -2806 |
|  | RUNX1 | -2546 |
|  | ERG/ FLI-1 | -928 |

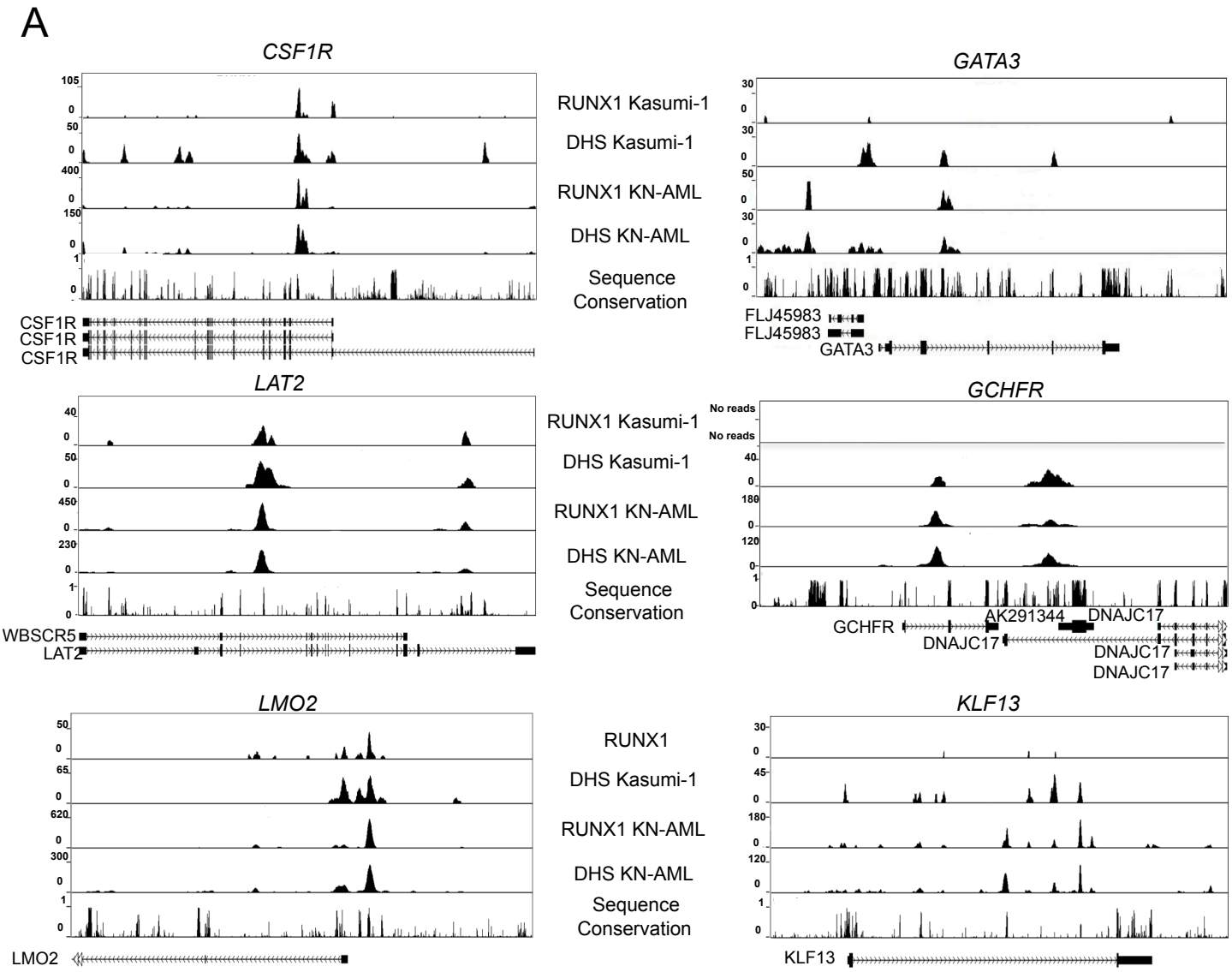
B

**Motifs associated with high confidence RUNX1/ETO peaks in Kasumi-1**

| motif | match | score (log p-value) |
|---|---|---|
|  | RUNX1 | -2266 |
|  | ETS | -1646 |
|  | E-box | -950 |
|  | E-box | -559 |

De novo motif discovery performed on the set of regions bound by RUNX1 **(A)** or RUNX1/ETO **(B).** Both RUNX1 and RUNX1/ETO peaks are enriched for RUNX1 and ETS-family binding sites. E-Box motifs were only enriched in RUNX1/ETO peaks.

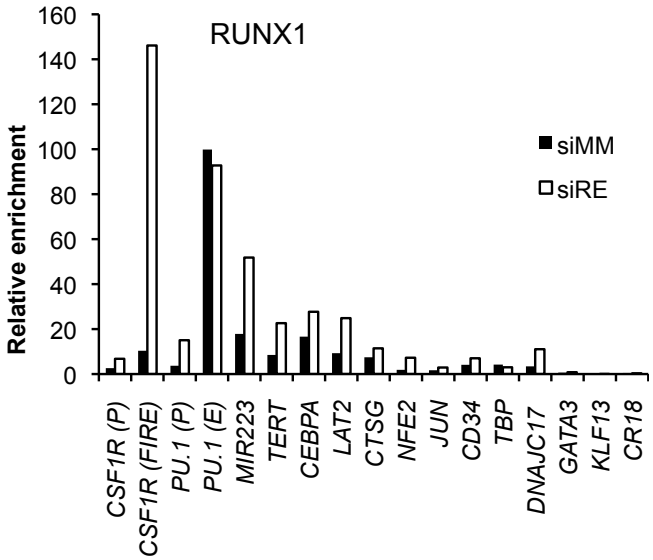
# Supplementary Figure 3




**Supplementary Figure 3:**

UCSC Genome Browser image of the human *CSF1R, GATA3, LAT2, GCHFR/DNAJC17, LMO2* and *KLF13* loci depicting differential RUNX1 binding and DHS profiles in Kasumi-1 cells and KN-AML blasts.

Supplementary Figure 4

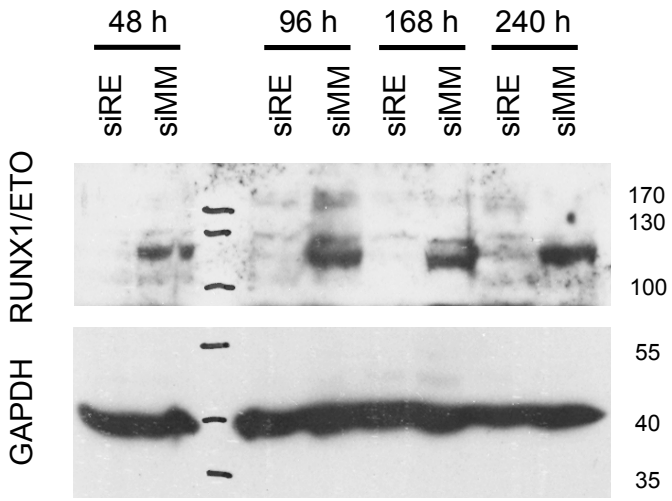



**Supplementary Figure 4:**

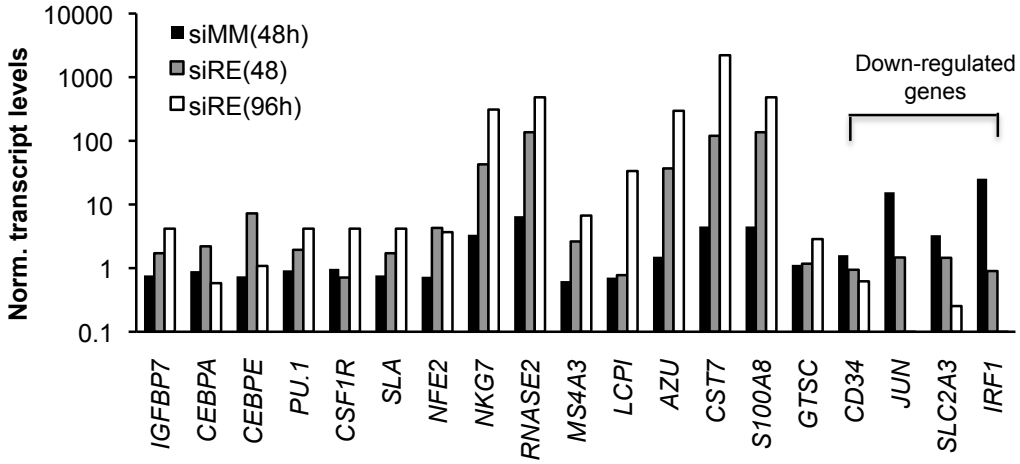
**(A)** Kasumi-1 cells were electroporated with either mismatch control siRNA (siMM) or RUNX1/ETO siRNA (siRE). Two days after siRNA electroporation RUNX1 binding was examined at the indicated loci by ChIP-qPCR. Values represent fold enrichment over non-specific binding at the negative control region.

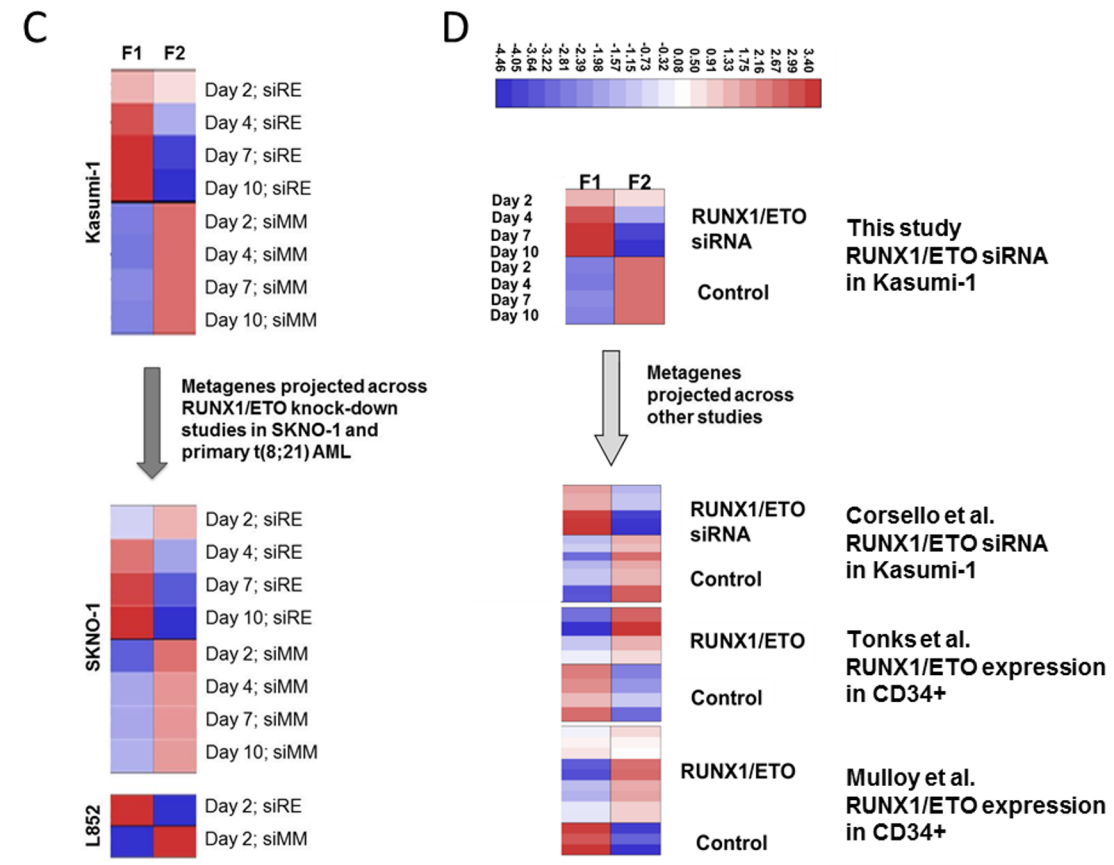
Supplementary Figure 5

A

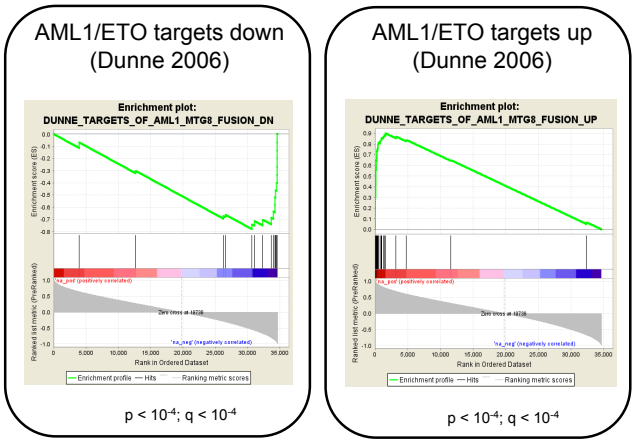



B






E

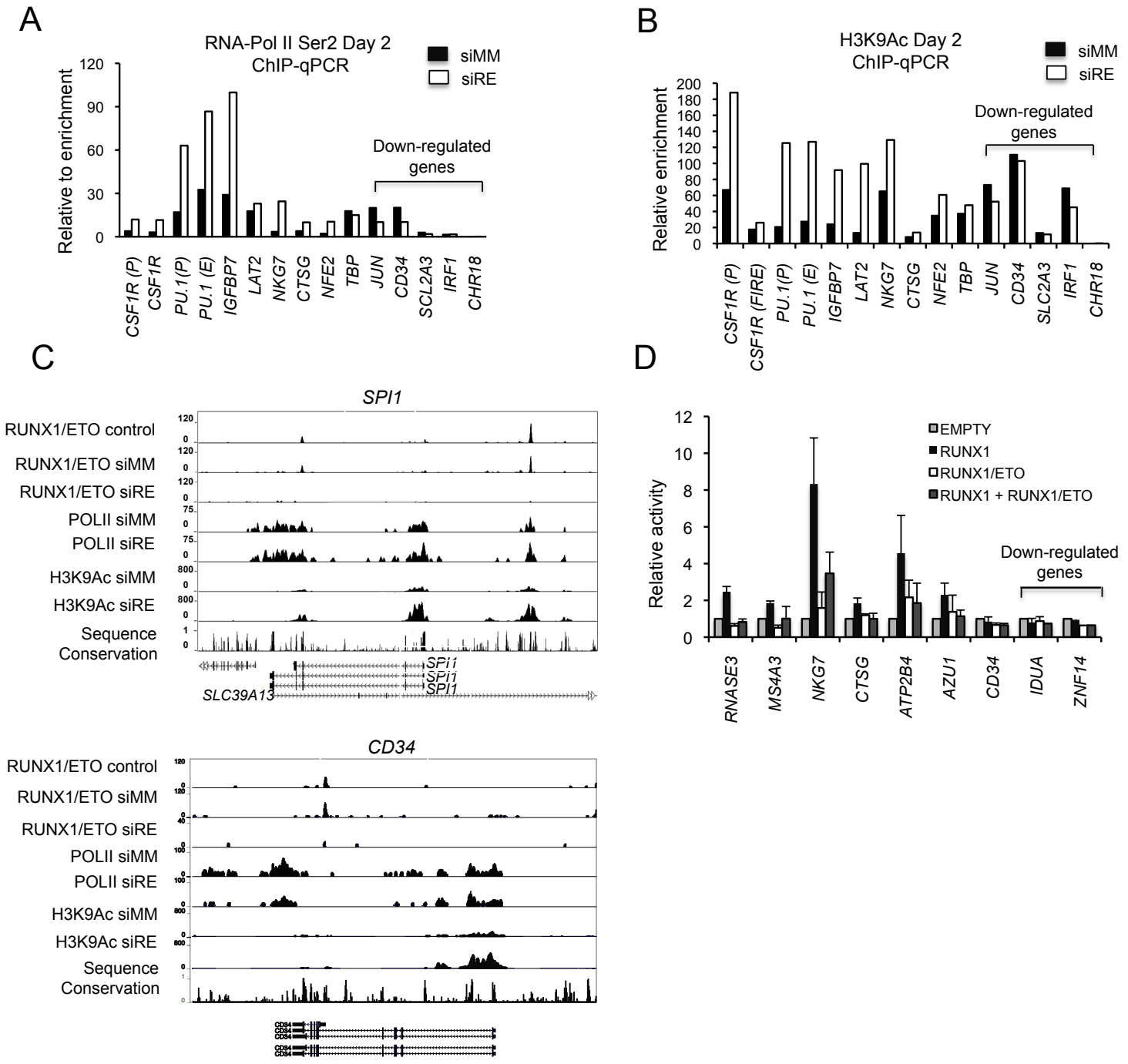



**Supplementary Figure 5:**

**(A)** Western-Blot demonstrating depletion of RUNX1/ETO protein during a 10 day period of knock-down. **(B)** Changes in gene expression caused by RUNX1/ETO knock-

down. Validation of early responding genes in Kasumi-1 cells treated with RUNX1/ETO or mismatch siRNA (siRE and siMM, respectively). Total RNA was isolated at the indicated times after siRNA treatment and analyzed by real-time PCR. **(C)** RUNX1/ETO knock-down yield concordant gene expression signatures in t(8;21)-positive cell lines and primary AML blasts. Two metagenes (single value summarising expression of multiple genes) were extracted from expression profiles of RUNX1/ETO knock-down in Kasumi-1 cells using NMF. Metagene F1 is expressed highly upon RUNX1/ETO knock-down and F2 in the control cells. Projection of these metagenes on a time course of siRNA treatment in SKNO-1 cells and on siRNA-treated RUNX1/ETO-expressing AML blasts demonstrate that RUNX1/ETO knock-down has highly reproducible consequences on gene expression pattern of t(8;21) cell lines and primary AML cells. **(D)** Ectopic expression and depletion of RUNX1/ETO yield concordant gene expression signatures. By projecting the F1 and F2 metagenes across three other studies we were able to validate our own results and show that the F1 Metagene is truly a reflection of RUNX1/ETO dependent genes. For instance, the F1 metagene shows similarly high expression in the knock-down experiments of Corsello et al. and conversely low expression in Tonks' and Mulloy's studies in which RUNX1/ETO is forcibly expressed in CD34+ cells (12-14). **(E)** Gene set enrichment analysis performed using a ranking of genes based on their correlation with the F1 metagene shows excellent concordance between this study and our previous RUNX1/ETO knock-down study (1): (**left**) targets down regulated by RUNX1/ETO shRNA, **(right**) targets up-regulated by RUNX1/ETO shRNA.

# Supplementary Figure 6

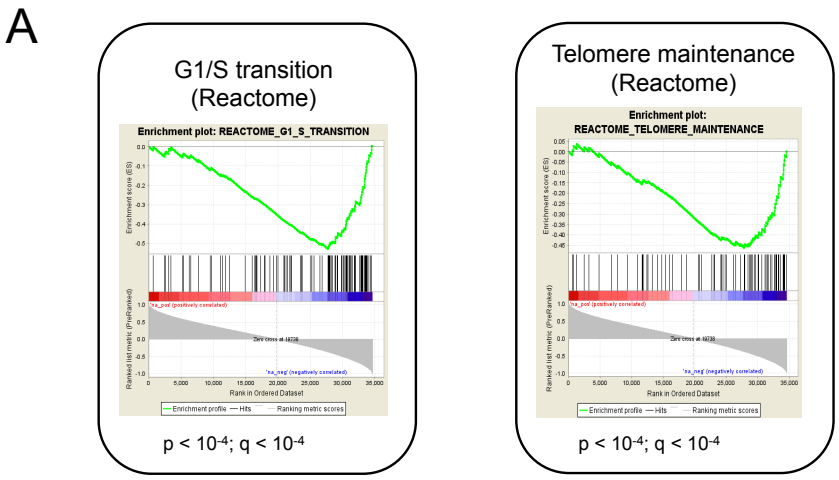


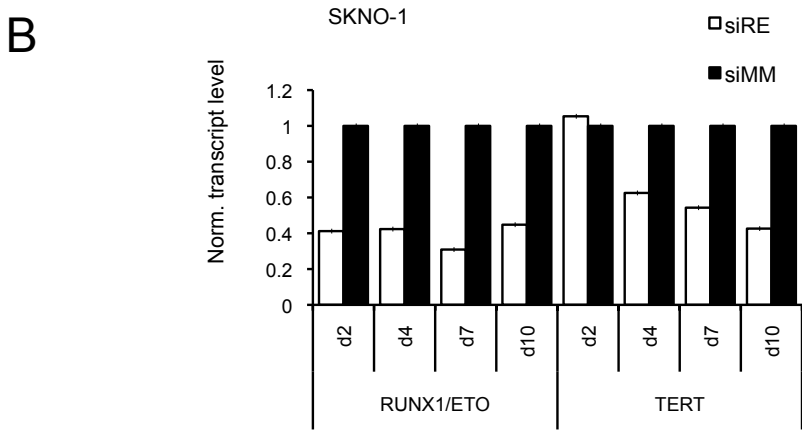

**Supplementary Figure 6:**

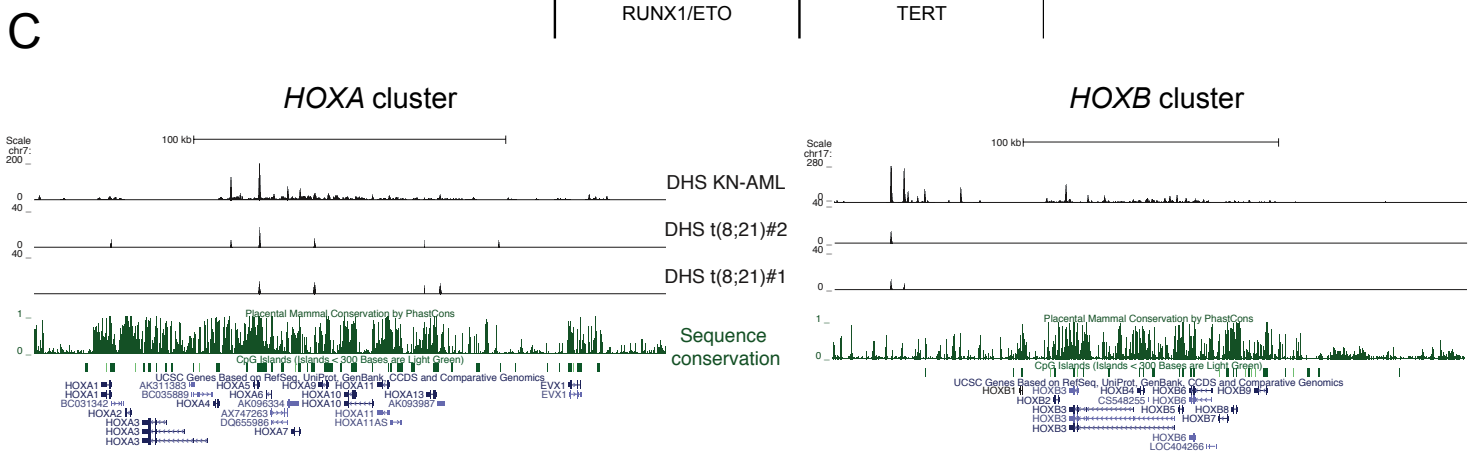
**(A**) Validation of RNA-Pol II ChIP-seq results in Kasumi-1 cells. Real-time PCR analysis of RNA-Pol II binding to *CSF1R*, *PU.1, IGFBP7, LAT2, NKG7, CTSG, NFE2, TBP, JUN, CD34, SCL2A3, IRF1* and *CHR18,* respectively, with and without RUNX1/ETO knock-

down. Values represent fold enrichment over non-specific binding at the Ch18 control region. Primers used for real-time PCR assays are listed in **Supplementary Table 3** and experimental details are described in Supplementary Methods. siRE, RUNX1/ETO siRNA; siMM, mismatch siRNA. Error bars indicate standard deviations. **(B**) Validation of H3K9Ac ChIP-seq results in Kasumi-1 cells**.** Analysis was performed as in (A). **(C)** Browser image of the *PU.1* (*SPI1*) and *CD34* loci showing RUNX1/ETO, RNA Pol II and H3K9Ac occupation in mock (control), mismatch siRNA (siMM) and RUNX1/ETO siRNA-treated (siRE) Kasumi-1 cells. **(D)** Regulatory potential of RUNX1 and RUNX1/ETO bound sequences. Promoters or distal elements identified as RUNX1 or RUNX1/ETO binding sites by ChIP-seq were cloned into pGL4-basic and pGL4-promoter luciferase-reporter vectors, respectively. Constructs were transiently co-transfected with RUN X1 and/or RUNX1/ETO expression vectors in RAW264.7 cells to test for RUNX1/ETO effects on reporter gene activity after 2 days. Results are representatives of three independent experiments. Error bars indicate standard deviations.

# Supplementary Figure 7




**Supplementary Figure 7:**

**(A)** Inverse correlation between RUNX1/ETO knock-down and gene expression signatures involved in cell proliferation and self-renewal. Gene set enrichment analysis using a ranking of genes based on their correlation with the F1 metagene shows inverse enrichment for genes involved in G1-S transition (left) and genes involved in telomere maintenance (right). **(B)** RUNX1/ETO depletion inhibits *TERT* expression. t(8;21)-

positive SKNO-1 were electroporated with either RUNX1/ETO or mismatch siRNA (siRE and siMM, respectively) at day 2, 4, 7 and 10. *RUNX1/ETO* and *TERT* transcript levels were analysed at the indicated time points by real-time PCR. **(C)** UCSC Genome Browser image of the human *HOXA* and *HOXB* loci depicting DHS profiles in CD34-positive blasts from two t(8;21) AML cases and KN-AML blasts.

| Dataset name | Additional information | No of mapped reads | No of Peaks | No of genes |
|---|---|---|---|---|
| RUNX1/ETO control | Kasumi-1cells control (without transfection) ETO antibody | 8,996,460 | 7813 | 4874 |
| RUNX1/ETO siMM | Kasumi-1 cells mismatch control (transfection with mismatch siRNA) ETO antibody | 14,533,444 | 6341 | 4372 |
| RUNX1/ETO siRE | Kasumi-1 cells knock-down (transfection with RUNX1/ETO siRNA) ETO antibody | 6,433,140 | 1595 | 1363 |
| RUNX1 control | Kasumi-1 cell control RUNX1 antibody | 21,199,304 | 8008 | 5462 |
| RUNX1 siMM | Kasumi-1 cells mismatch control (transfection with mismatch siRNA) RUNX1 antibody | 31,107,296 | 8368 | 4955 |
| RUNX1 siRE | Kasumi-1 cells knock-down (transfection with RUNX1/ETO siRNA) ETO antibody | 33,191,671 | 11827 | 6041 |
| RUNX1 non-t(8;21) | Peripheral blood MNC CD34+ from human patient with non-t (8;21)AML patient RUNX1 antibody | 36,396,667 | 14552 | 7619 |
| POL II siRE | Kasumi-1 cells (transfection with RUNX1/ETO siRNA) RNA-Pol II Ser2 antibody | 17,458,661 | 23817 | 10959 |
| POL II siMM | Kasumi-1cells (transfection with mismatch siRNA) RNA-Pol II Ser2 antibody | 14,155,688 | 30561 | 11484 |
| H3K9ac siRE | Kasumi-1cells (transfection with RUNX1/ETO siRNA) H3K9Ac antibody | 29,167,903 | 29356 | 13530 |
| H3K9ac siMM | Kasumi-1 cells (transfection with mismatch siRNA) H3K9Ac antibody | 26,724,237 | 33127 | 14836 |
| DNaseI | Kasumi-1 cells | 36,198,801 | 33832 | 14249 |
| DNaseI t(8;21)#1 | CD34+ cells from human patient with t(8;21) AML | 17,008,402 | 29276 | 13564 |
| DNaseI t(8;21)#2 | CD34+ cells from human patient with t(8;21) AML | 16,237,704 | 28998 | 13212 |
| DNaseI non-t(8;21) | Peripheral blood MNC CD34+ from human patient with non-t (8;21) AML | 31,863,492 | 36033 | 13518 |
| Patient A | Blast cells from patient with t (8;21) AML | 28,900,000 | 8665 | (5534) |
| Patient B | Blast cells from patient with t (8;21) AML | 28,600,000 | 8740 | (6215) |

**Supplementary Table 1**

The number of peaks of RUNX1, RUNX1/ETO, RNA-POL II, H3K9ac and DHS as determined by ChIP-sequencing and DNaseI-sequencing, as well as genes associated with these peaks. Note that the number of genes in patients A and B is likely to be an overestimate due to the number of non-specific peaks. The genomic coordinates of bound by RUNX1, RUNX1/ETO, POL II-RNA and H3K9ac can be found under accession number GSE29225

H12812 (t(8;21) 1): CD34+CD117+CD45+CD15-CD13+HLADR+CD33+CD7-CD19+CD56-CD14-CD38+CD64+/-, NPM WT, FLT3 WT, CKIT WT

H18901/09 (T(8;21)2): CD34+CD117+CD45+CD15+/-CD13+HLADR+CD33+CD7-CD19+/-CD56-CD14-CD38+CD64-, NPM WT, FLT3 WT, CKIT WT

H15668/10 (KN-AML): CD34+CD117+CD45+CD15-CD13+HLADR+/-CD33+CD7+/-CD19-CD56-CD14-CD38+CD64+/-, NPM WT, FLT3-ITD, CKIT WT

Patient A: 18 yr, 45,X,Y,t(8;21)(q22;q22)
Patient B: 34 yr, 45,X,-Y,t(8;21)(q22;q22)

**Supplementary Table 2**

Surface marker profile and mutation status of leukemic blast cells at the point of diagnosis as measured by Flow Cytometry.

# Supplementary Table 3

## Primers used for Real-Time PCR

| cDNA | **Forward** | **Reverse** |
|---|---|---|
| *GAPDH* | CCTGGCCAAGGTCATCCAT | AGGGGCCATCCACAGTCTT |
| *RUNX1/ETO* | TCAAAATCACAGTGGATGGGC | CAGCCTAGATTGCGTCTTCAC A |
| *S100A8* | CGAGCTGGAGAAAGCCTTGA | GACGTCTGCACCCTTTTTCC |
| *CSF1R* | AGCACGAGAACATCGTCAACC | TTCGCAGAAAGTTGAGCAGGT |
| *PU.1* | TCTTGGCCACCAGGTCTCCTA | CGCCCTCCTCCTCATCTGA |
| *IGFBP7* | GAAGTAACTGGCTGGGTGCTG | GCTGATGCTGAAGCCTGTCC |
| *CD34* | CACTGGCTATTTCCTGATGAAT | CCACCGTTTTCCGTGTAAT |
| *SLA* | GGCTCACCTTCCAGTGCCT | TTCTGACGCAAGGTGACAGG |
| *C/EBP* | ATGTCCCACGGGACCTACTACGA | ACAGTGTGCCACTTGGTACTGCAG |
| *AZU1* | AACCTGAACGACCTGATGCTG | ATCGTCACGCTGCTGGTGA |
| *CST7* | CCAACCACACCTTGAAGCAGA | GGGTCAGTGACAACGGAGAAC |
| *CTSG* | TCCTGGTGCGAGAAGACTTTG | GGTGTTTTCCCGTCTCTGGA |
| *SLC2A3* | CCAGGAGATGAAAGATGAGAGTGC | ATGATGATGGGCTGTCGGTAGC |
| *JUN* | TGCTTACCAAAGGATAGTGCGATG | TTGACTTCTCAGTGGGCTGTCC |
| *LCP1* | TGATCCTGATTGTCGGCATG | CAATGCCATCTCCAACAGCA |
| *MS4A3* | CCAAGCCATAAACAACCCCA | TTCTGGTCCCGTCTCACTGC |
| *NFE2* | CCAAGGTGTGTTCAAAGAGGC | GGAGCCGAGTCAGGGAAGAC |
| *NKG7* | CTGATTGCTTTGAGCACCGA | CCTGATATGATGTCCCCATGC |

| | | |
|---|---|---|
| *IRF1* | AACACTTAGCGGGATTC | ACAACAGCCTGATTTCC |
| *RNASE2* | CCCCTGAACCCCAGAACAA | ACCATGTTTCCCAGTCTCCG |
| *C/EBP$\alpha$* | GAGGGACCGGAGTTATGACA | AGACGCGCACATTCACATT |

## Primers used for Real-Time PCR

| **Binding sites** | **Forward** | **Reverse** |
|---|---|---|
| *TBP (P)* | CTGGCGGAAGTGACATTATCAA | GCCAGCGGAAGCGAAGTTA |
| *IVL (P)* | GCCGTGCTTTGGAGTTCTTA | CCTCTGCTGCTGCCACTT |
| *CSF1R FIRE* | GCCTGACGCCAACAATGTG | GGCAAAGGAGGGAAGTGAGAG |
| *PU.1 14 3H* | AACAGGAAGCGCCCAGTCA | TGTGCGGTGCCTGTGGTAAT |
| *PU.1 (P)* | CTGCCGCTGGGAGATAG | CGGCCAGAGACTTCCTGTA |
| *CSF1R (P)* | AGAAGAGGTCAGCCCAAGGA | AGGGATCGGGACACTGGAC |
| *CHR 18* | ACTCCCCTTTCATGCTTCTG | AGGTCCCAGGACATATCCATT |
| *LCP1* | GAACAGAGTGACCTCCTAAG | CCAGCAACTTCTTCTTGG |
| *CTSG* | TCAGTTGCTGCTGTGCTTC | TTCTCAATCCCCTGTCCCCAC |
| *NFE2* | TCTCAGTCCTTCCTCCAAC | ACATCGGGCTCAAGTTTC |
| *NKG7* | CCAAGAAAGAGAGAACAGC | GGGAGGGTATGTGAAAAG |
| *IRF1* | AACACTTAGCGGGATTC | ACAACAGCCTGATTTCC |
| *SLC2A3* | TGTTTATTGACAGTAATCAGC | TGAGCCTACAGATGAACTC |
| *JUN* | TTGGGGTTACTGTAGCCATAAG | CGTGAAGTGACGGACTGTTC |

| | | |
|---|---|---|
| *CD34* | TGTGGTTAGCCAAACTCCAGGTC | TGAGGAATGAAGCAGCAGTGG |
| *IGFBP7* | GTCAAGCACTAAAAGGACAAACCG | TGAATGCCACTGGGAGACAAAG |
| *LAT2* | AAACCCAGAACAACCCAGGC | ATGAGGAAGGATGTGTGTGCGG |
| *C/EBPA* | GCCAGTTTATGGAGGTGTGAGC | ATAGGTGGTGATGATGGTTGCC |
| *MIR223* | TTGGAAAGTTAGTGTCTGTTGAAGG | TGTTGTGAAAGGGTCTGCTACTG |
| *TERT* | AAATGGTCTCAGCCTCACCGTC | TTCCCTCCAATCACACCTTGC |

## Primers used for reporter gene construction

| **Binding sites** | **Forward** | **Reverse** |
|---|---|---|
| *RNASE3* | GTAGGCTGGGTACCTTCACCCAGAGTCCAGATCCCA | GTCTAGAGGCTAGCCAGTTTTGGAACCATGTTTCCTG |
| *MS4A3* | GTAGGCTGGGTACCCTAAGCATGTAAGAATTGAATGTCA | GTCTAGAGGCTAGCAGTCCACCTTGTCGCAAGTAGGA |
| *CTSG* | GTAGGCTGGGTACCGTAGAAACAGGAAGTGGGCAGCT | GTCTAGAGGCTAGCAGGCAATTTGGCTCAAGAATCTAT |
| *NKG7* | GTAGGCTGGGTACCGATTCTGAGCTCCTGCCCGCCTCTT | GTCTAGAGGCTAGCGGTCCTTAGAGCCCAAGAAAGAGAGAAC |
| *ATP2B4* | GTAGGCTGGGTACCAGGAACCCCCTATTCAGACAGGAG | GTCTAGAGGCTAGCTTTGGACGATCTCAACAGCAGAGCT |
| *AZU1* | GTAGGCTGGGTACCCTGCCAGCTCCACTGCCCTGAGC | GTCTAGAGGCTAGCAAGCCGCTTGCCTTCTCTGTGCT |
| *CD34* | GTAGGCTGGGTACCGGCTCGTGTCTCTGTGACCTGGAGT | GTCTAGAGGCTAGCATTTATTCCAGCAGCCTCCAATCA |
| *IDUA* | GTAGGCTGGGTACCGGGTAATTCTCCTTCCTGCTAAAG | GTCTAGAGGCTAGCGCGCGGCGCGGGGGCGCAGGGGA |
| *ZNF14* | GTAGGCTGGGTACCAACGAGACAGATTTCAGTGTTGCA | GTCTAGAGGCTAGCCTGGCCCCGCACACTCACCATTTC |

**Supplementary Table 4**
**Lists of genes bound by RUNX1, RUNX1/ETO, RNA-PolII and carrying H3K9 acetylated histones as well as DNaseI hypersensitive sites in Kasumi-1 cells**

**Supplementary Table 5:**
**Lists of genes jointly bound by RUNX1/ETO in patients A and B and Kasumi-1 cells**

**Supplementary Table 6**
**Genes with ≥ 2 fold changed expression upon RUNX1/ETO knock-down and comparison with RUNX1/ETO, Pol II and H3K9ac peaks**

**Supplementary Table 7**
**GSEA using a F1 metagene**

**Supplementary Table 8**
**Functional annotation clustering of peaks bound by RUNX1/ETO and RUNX1**

**Supplementary Table 9**
**Functional annotation clustering of genes with changed expression upon RUNX1/ETO knock-down**