

# RUNX1-ETO depletion in t(8;21) AML leads to C/EBP $\alpha$ - and AP-1-mediated alterations in enhancer-promoter interaction

Ptasinska, Anetta; Pickin, Anna; Assi, Salam; Chin, Paulynn Suyin; Ames, Luke; Avellino, Roberto; Groeschel, Stephan; Delwel, Ruud; Cockerill, Peter; Osborne, Cameron S; Bonifer, Conny

DOI:

[10.1016/j.celrep.2019.08.040](https://doi.org/10.1016/j.celrep.2019.08.040)

License:

Creative Commons: Attribution (CC BY)

## Document Version

Publisher's PDF, also known as Version of record

## Citation for published version (Harvard):

Ptasinska, A, Pickin, A, Assi, S, Chin, PS, Ames, L, Avellino, R, Groeschel, S, Delwel, R, Cockerill, P, Osborne, CS & Bonifer, C 2019, 'RUNX1-ETO depletion in t(8;21) AML leads to C/EBP $\alpha$ - and AP-1-mediated alterations in enhancer-promoter interaction', *Cell Reports*, vol. 28, no. 12, pp. 3022-3031.e7. <https://doi.org/10.1016/j.celrep.2019.08.040>

[Link to publication on Research at Birmingham portal](#)

## Publisher Rights Statement:

Ptasinska, A, Pickin, A, Assi, S, Chin, PS, Ames, L, Avellino, R, Groeschel, S, Delwel, R, Cockerill, P, Osborne, CS & Bonifer, C 2019, 'RUNX1-ETO depletion in t(8;21) AML leads to C/EBP $\alpha$ - and AP-1-mediated alterations in enhancer-promoter interaction', *Cell Reports*, vol. 28, no. 12, pp. 3022–3031. e1–e7. © 2019 The Authors. <https://doi.org/10.1016/j.celrep.2019.08.040>

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

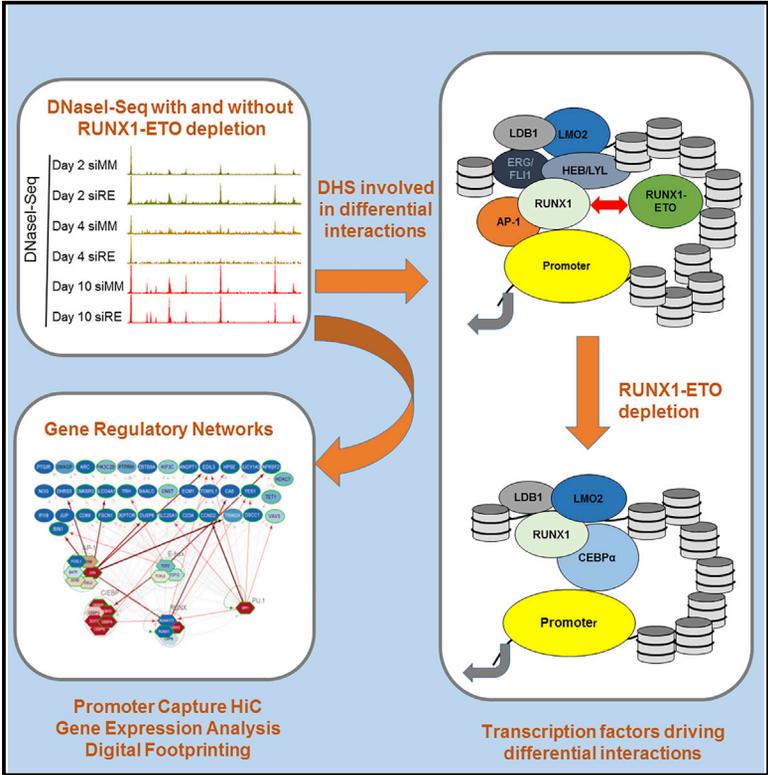
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Cell Reports

## RUNX1-ETO Depletion in t(8;21) AML Leads to C/EBP $\alpha$ - and AP-1-Mediated Alterations in Enhancer-Promoter Interaction

### Graphical Abstract



### Authors

Anetta Ptasinska, Anna Pickin, Salam A. Assi, ..., Peter N. Cockerill, Cameron S. Osborne, Constanze Bonifer

### Correspondence

c.bonifer@bham.ac.uk

### In Brief

Promoter-Capture Hi-C assays, gene expression, and transcription-factor binding data are used to construct a RUNX1-ETO-dependent dynamic gene regulatory network that maintains acute myeloid leukemia (AML). Ptasinska et al. show that RUNX1-ETO participates in cis-regulatory element interactions and that differential interactions after RUNX1-ETO depletion are driven by C/EBP $\alpha$  and AP-1.

### Highlights

- Promoter Capture Hi-C links t(8;21) AML-specific cis-elements to correct promoters
- Interacting cis-elements bound by transcription factors form a gene regulatory network
- The product of the t(8;21) translocation, RUNX1-ETO, participates in interactions
- Differential interactions after RUNX1-ETO depletion are driven by C/EBP $\alpha$  and AP-1



# RUNX1-ETO Depletion in t(8;21) AML Leads to C/EBP $\alpha$ - and AP-1-Mediated Alterations in Enhancer-Promoter Interaction

Anetta Ptasinska,<sup>1,5</sup> Anna Pickin,<sup>1,5</sup> Salam A. Assi,<sup>1,5</sup> Paulynn Suyin Chin,<sup>1</sup> Luke Ames,<sup>1</sup> Roberto Avellino,<sup>3,6</sup> Stephan Gröschel,<sup>3,7</sup> Ruud Delwel,<sup>3,4</sup> Peter N. Cockerill,<sup>1,8</sup> Cameron S. Osborne,<sup>2,8</sup> and Constanze Bonifer<sup>1,9,\*</sup>

<sup>1</sup>Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham B152TT, UK

<sup>2</sup>Department of Medical & Molecular Genetics, King's College London, London SE1 9RT, UK

<sup>3</sup>Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands

<sup>4</sup>Oncode Institute, Erasmus University Medical Center, Rotterdam, the Netherlands

<sup>5</sup>These authors contributed equally

<sup>6</sup>Present address: Department of Immunology, Weizmann Institute of Science, Rehovot 7610001, Israel

<sup>7</sup>Present address: Molecular Leukemogenesis, Deutsches Krebsforschungszentrum, 69120 Heidelberg, Germany; Department of Internal Medicine V, Heidelberg University Hospital, 69120 Heidelberg, Germany

<sup>8</sup>Senior author

<sup>9</sup>Lead Contact

\*Correspondence: [c.bonifer@bham.ac.uk](mailto:c.bonifer@bham.ac.uk)

<https://doi.org/10.1016/j.celrep.2019.08.040>

## SUMMARY

Acute myeloid leukemia (AML) is associated with mutations in transcriptional and epigenetic regulator genes impairing myeloid differentiation. The t(8;21)(q22;q22) translocation generates the RUNX1-ETO fusion protein, which interferes with the hematopoietic master regulator RUNX1. We previously showed that the maintenance of t(8;21) AML is dependent on RUNX1-ETO expression. Its depletion causes extensive changes in transcription factor binding, as well as gene expression, and initiates myeloid differentiation. However, how these processes are connected within a gene regulatory network is unclear. To address this question, we performed Promoter-Capture Hi-C assays, with or without RUNX1-ETO depletion and assigned interacting cis-regulatory elements to their respective genes. To construct a RUNX1-ETO-dependent gene regulatory network maintaining AML, we integrated cis-regulatory element interactions with gene expression and transcription factor binding data. This analysis shows that RUNX1-ETO participates in cis-regulatory element interactions. However, differential interactions following RUNX1-ETO depletion are driven by alterations in the binding of RUNX1-ETO-regulated transcription factors.

## INTRODUCTION

Acute myeloid leukemia (AML) is a hematopoietic malignancy caused by genetic abnormalities in hematopoietic stem cells (HSCs), which restrict their ability to undergo the normal differentiation process (Bonifer and Cockerill, 2015; Kumar, 2011). The

transcription factors (TFs) regulating hematopoiesis have to be expressed in a stage- and lineage-restricted fashion since their mutation or de-regulation impairs differentiation and prolongs the proliferative stage, thus increasing the opportunities for cells to further mutate and progress to AML (Bonifer and Cockerill, 2011; Rosenbauer and Tenen, 2007). One of the best-studied subtypes of AML is the t(8;21)(q22;q22) translocation generating the RUNX1-ETO fusion protein (Erickson et al., 1992; Miyoshi et al., 1991). RUNX1-ETO has a modular structure comprising the RUNX1 DNA-binding domain plus four evolutionary conserved functional domains named neryv homology regions 1 to 4 (NHR1 to NHR4) (Kitabayashi et al., 1998), which recruit transcriptional repressors such as the N-CoR/mSin3/HDAC1 complex (Lutterbach et al., 1998). The expression of this abnormal protein results in a block in differentiation and increased cell survival (Dunne et al., 2006; Heidenreich et al., 2003; Martinez et al., 2004; Ptasinska et al., 2012).

The RUNX1-ETO fusion protein is part of a larger TF complex consisting of RUNX1-ETO; CBF $\beta$ ; the erythroblast transformation-specific (ETS) family of transcription factors (ERG and FLI1); E proteins such as HEB, E2A, and LYL1; and the non-DNA binding factors LDB1 and LMO2 (Martens et al., 2012; Ptasinska et al., 2014; Sun et al., 2013). Each part of the complex is thought to be essential for AML maintenance (Sun et al., 2013). RUNX1-ETO depletion in t(8;21) cells is sufficient to trigger extensive global changes in the transcriptomic and epigenetic profile across hundreds of genes (Ben-Ami et al., 2013; Dunne et al., 2006; Ptasinska et al., 2012; Sun et al., 2013). Depletion upregulates a specific set of RUNX1-regulated genes, such as *CEBPA*, leading to increased recruitment of RUNX1 and C/EBP $\alpha$  to gene regulatory elements throughout the genome, thereby releasing the block on myeloid differentiation and suppressing self-renewal (Loke et al., 2017; Ptasinska et al., 2014; Sun et al., 2013). We have previously used global TF binding and gene expression information to construct a dynamic gene regulatory network linking genes bound by the RUNX1-ETO



complex to dynamic changes of gene expression (Ptasinska et al., 2014). We used such system-wide information to devise a RNAi dropout screen that identified a number of genes associated with AML maintenance (Martinez-Soria et al., 2019). However, to fully explore the power of genome-wide studies, we need to construct gene regulatory networks that enable us to predict the results of perturbation experiments from such data using modeling approaches. Therefore, a number of issues still need to be resolved. RUNX1-ETO mostly binds to distal cis-regulatory elements, and although we can define genes responding to RUNX1-ETO knockdown, we do not know whether this response is direct or indirect, as we do not know which promoter is linked to the sites of fusion protein binding. In addition, we do not know which other TFs participate in the maintenance of the leukemic state and drive the response to RUNX1-ETO knockdown.

To answer these questions, we identified direct cis-element interactions using the Promoter Capture Hi-C (ChI-C) method (Mifsud et al., 2015) in Kasumi-1 cells, a well-known model of t(8;21) AML, with and without small interfering RNA (siRNA)-mediated RUNX1-ETO depletion. RUNX1-ETO knockdown leads to a rewiring of promoter-enhancer interactions, which is driven by increased C/EBP $\alpha$  and loss of AP-1 binding after knockdown. We integrated these results with chromatin immunoprecipitation (ChIP) and digital footprinting data from cell lines and patients to identify regulatory relationships between binding events and gene expression, which will aid further studies aimed at identifying pathways required for t(8;21) AML leukemic maintenance.

## RESULTS AND DISCUSSION

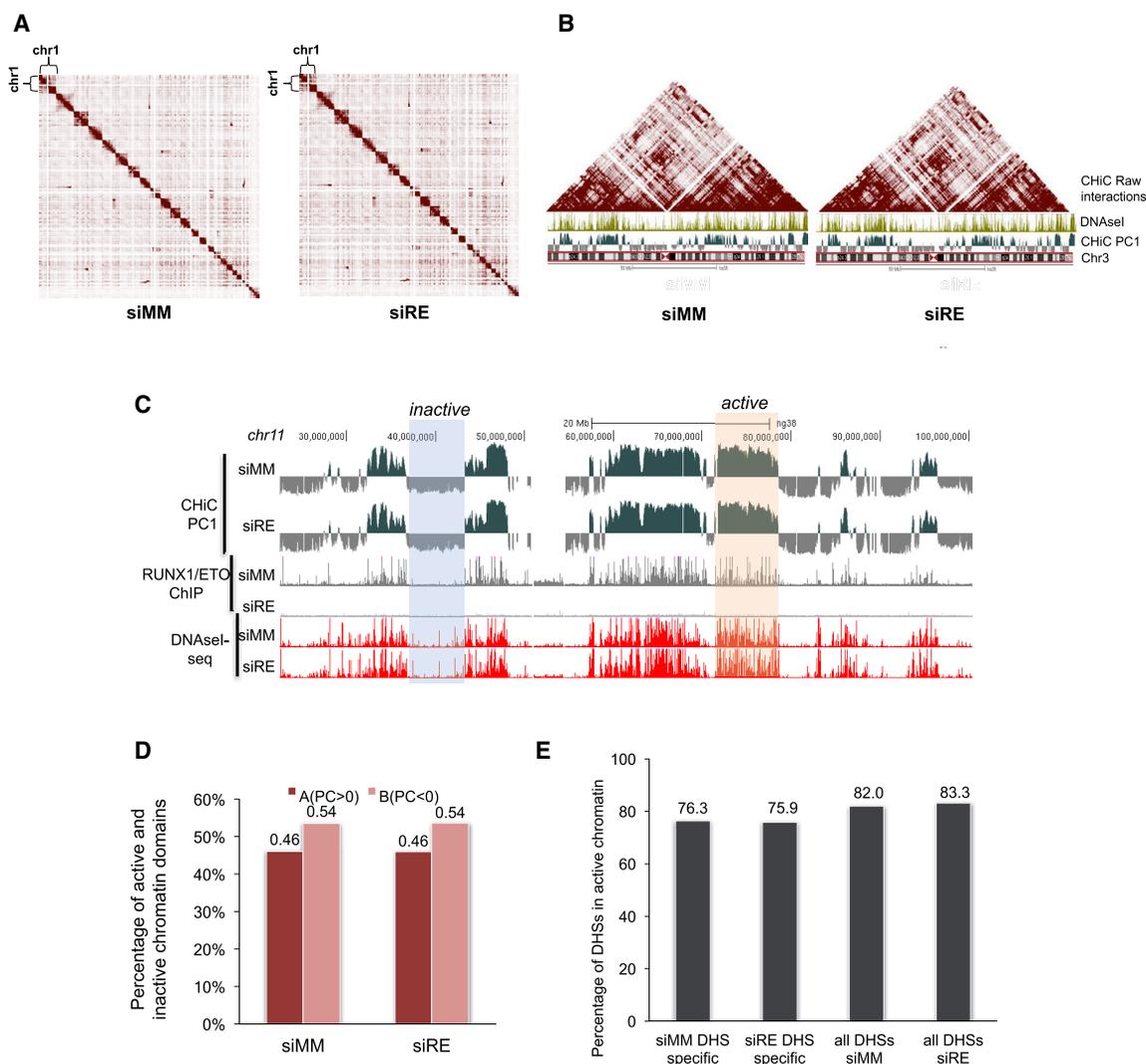
### RUNX1-ETO Depletion Does Not Lead to a Global Reorganization of Chromosome Structure but Changes Promoter-Enhancer Interactions within TADs

The tissue specificity of gene expression is controlled by lineage-restricted TFs binding to distal cis-regulatory elements that need to physically interact with their target promoters in order to activate gene expression (de Laat and Duboule, 2013; Plank and Dean, 2014). To examine whether RUNX1-ETO influences genome-wide cis-element interactions, we generated duplicate ChI-C libraries (Mifsud et al., 2015) from Kasumi-1 cells that were either untreated (mismatch control siRNA [siMM]) or following a 4-day siRNA-mediated treatment to knockdown RUNX1-ETO (siRE) (Figure 1A). Data analysis of the sequenced libraries identified 57,775 significant interactions between promoters and distal elements before and 60,681 significant interactions after RUNX1-ETO depletion. ChI-C libraries were highly reproducible with an average of 70% overlap of significant interactions between replicates (Figures S1A–S1D). To align our ChI-C data with the coordinates of cis-regulatory elements, we mapped Deoxyribonuclease I (DNaseI) hypersensitive sites (DHSs) during a knockdown time course (Figure S1E). The presence (siMM) or absence (siRE) of RUNX1-ETO did not influence global chromosomal organization across all chromosomes (Figure 1A), including the organization of this region into topologically associated domains (TADs; large triangles, projected above the DHS pattern) (Gonzalez-Sandoval and

Gasser, 2016; Figure 1B). Figure 1C shows a University of California, Santa Cruz (UCSC) genome browser screenshot highlighting active and inactive chromatin compartments plotted alongside RUNX1-ETO ChIP data (Ptasinska et al., 2012) and day-10 DNase I hypersensitive sites sequencing (DNaseI-seq) data (this manuscript). These analyses revealed clusters of interactions within active and inactive regions whose ratio was invariant even after an extended period of RUNX1-ETO depletion (Figure 1D).

To investigate whether specific DHS patterns seen following RUNX1-ETO depletion correlated with a specific stage of myeloid differentiation, we compared DNaseI data from control and RUNX1-ETO-depleted Kasumi-1 cells (days 2, 4, and 10) to published assay for transposase-accessible chromatin using sequencing (ATAC-seq) data defining the open chromatin regions of normal stem and progenitor cells representing different developmental stages (Corces et al., 2016; Figure S1F). DHSs specific for control cells (bottom) aligned more closely with HSCs and early progenitors and showed increased AP-1 motif enrichment, whereas DHSs specific for RUNX1-ETO-depleted cells (top) aligned with those of monocytic cells and were enriched for C/EBP motifs. RUNX1-ETO depletion had a profound effect on gene expression with a large number of genes changing expression by day 10 (Figure S1G). Flow cytometry and principal component analyses of DNaseI-seq data revealed that Kasumi-1 cells gradually lose their stem cell markers (CD34 and CD117) while principal component and correlation clustering analyses of the DNaseI-seq data indicated that at day 10, but not yet at day 4, they had differentiated toward monocytic cells (Figures S1J and S1K). Phenotypic changes were accompanied by changes in protein and mRNA expression for a number of TFs visible already at day 2, in particular C/EBP $\alpha$  (Figures S1L and S1M), which is rapidly upregulated after knockdown. *JUN* mRNA was strongly downregulated during the first days of knockdown but then was strongly upregulated in concordance with its important role in regulating monocyte and/or macrophage-specific gene expression (Heinz et al., 2010). The expression of *JUND* protein was upregulated as well, but note that the DNA-binding activity of the AP-1 family of TFs is regulated by signaling-mediated phosphorylation (Bejjani et al., 2019).

Around 80% of all DHSs detected in active regions of control cells or RUNX1-ETO-depleted cells participated in promoter-enhancer interactions (Figure 1E, right panels). To identify differential interactions, we used the ChI-C data to assign the respective DHSs to the promoter for RUNX1-ETO-depleted and control cells (Data S1). Figures 2A and 2B show statistically significant control-specific and RUNX1-ETO-specific interactions at 5-kb resolution involving specific DHSs on chromosome 3, which were not seen with shared DHSs (Figure S2A), indicating that it is the differential DHSs that drive these changes. A total of 1104 DHSs were significantly increased and 1209 were significantly decreased after 10 days of RUNX1-ETO knockdown (Figures 2E and S1F). The majority of these DHS (75% and 76%, respectively) show differential promoter-enhancer interactions already after day 4 of knockdown (Figure 1E), demonstrating that RUNX1-ETO depletion alters the epigenome prior to monocytic differentiation.



**Figure 1. RUNX1-ETO and the Genome Organization in t(8;21) AML**

(A) Contact matrix across the whole genome. Each pixel represents a 10-Mb section of the genome. Color intensity represents interaction frequency. The left-hand plot shows a Capture HiC interaction matrix generated with data from Kasumi-1 cells transfected with mismatch control siRNA (siMM) for 4 days; the right-hand plot shows an interaction matrix from RUNX1-ETO-depleted Kasumi-1 cells transfected with the specific siRNA (siRE).

(B) Contact matrix across chromosome 3 at 10-Mb resolution. The heatmap shows the raw interactions on chromosome 3 using Kasumi-1 cells transfected with siMM (left) and siRE (right); a UCSC track highlighting the DHS pattern is shown below each heatmap together with the Chi-C first principle component (PC1) plot (see below).

(C) UCSC genome browser screenshot shows a first principle component plot for Capture HiC siMM and siRE samples plotted along with RUNX1-ETO ChIP data (Ptasinska et al., 2014) and DNaseI-seq control (siMM) and knockdown (siRE) data from Kasumi-1 cells for a 70-Mb regions on chromosome 11.

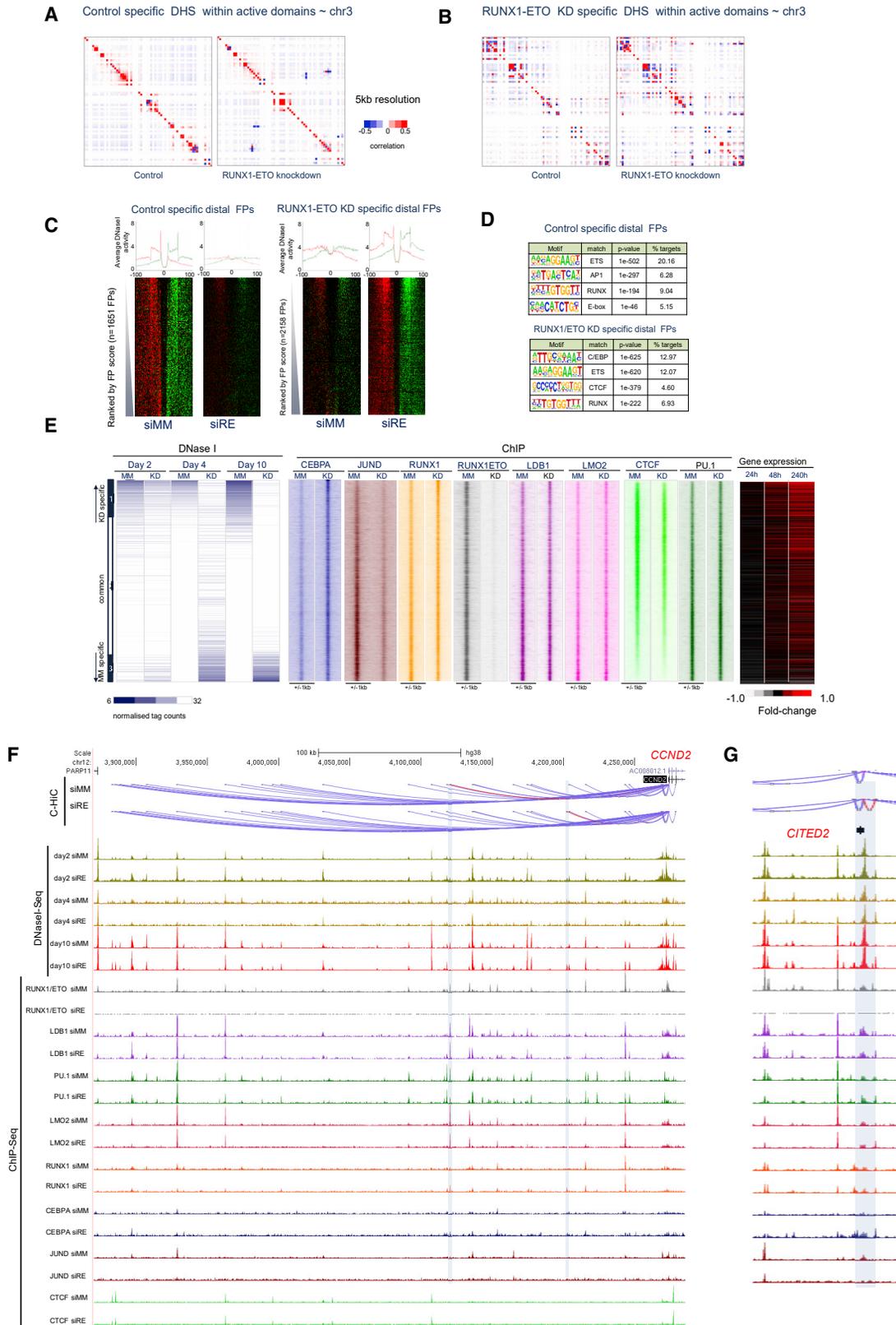
(D) Percentage of DHSs in active and inactive chromatin compartments in Kasumi-1 cells transfected with siMM and siRE.

(E) Percentage of DHSs found at day 10 of knockdown participating in promoter-enhancer interactions (determined at day 4 of knockdown) detected in all active chromatin regions of siMM cells or siRE cells (right two panels), and specific to siMM or siRE cells (left two panels), indicating that the majority of specific DHSs are already present at day 4.

### RUNX1-ETO-Regulated TFs Drive Differential Cis-Regulatory Element Interactions

We next sought to identify the TFs driving the changes in interactions by performing digital footprinting analysis, using our Wellington algorithm (Piper et al., 2013). This approach reveals TF motifs protected from DNaseI digestion and evaluates genome-wide TF occupancy with high accuracy (Figure 2C). Examples of such footprints for day 10 of knockdown are depicted

for ETS and C/EBP motifs at the *IL17RA* locus in Kasumi-1 cells (Figure S2B). Global binding motif analysis confirmed that AP-1 motifs were preferentially occupied in control (siMM) cells whereas C/EBP motifs were occupied in RUNX1-ETO-depleted (siRE) cells (Figure 2D). Motif occupancy was validated by comparing footprinting data with previously generated ChIP-seq data for C/EBP $\alpha$ , RUNX1-ETO, PU.1, JUND, and RUNX1 (Ptasinska et al., 2014; Martinez-Soria et al., 2019) (Figure S2C).



(legend on next page)

Two factors capable of mediating long-range interactions are CTCF and LDB1 (Deng et al., 2012; Splinter et al., 2006). To examine their role, we generated new CTCF and LDB1 ChIP data with and without RUNX1-ETO depletion. We correlated changes in gene expression and TF binding at specific DHSs with differential interactions between DHSs and promoters (Figure 2E). These analyses revealed a global increase in C/EBP $\alpha$  binding after knockdown and a decrease in JUND binding at siMM-specific DHSs (Figure 2E). All other factors showed no or little difference in binding between knockdown and control cells. Changing interactions correlated with differential gene expression (Figure 2E, outermost right panel). Figure 2F shows an example of interactions at the upregulated *CCND2* gene, which shows changes in interactions between its promoter and two upstream distal elements (depicted in red). Two observations are noteworthy. First, the *CCND2* promoter interacts with a large number of distal DHSs. Second, a large number of RUNX1-ETO binding sites are located within these sites, indicating that RUNX1-ETO is part of an extended and mostly invariant chromatin hub. To validate our Chi-C data, we conducted a circularized chromosome conformation capture (4C) experiment that investigated the *SPI1* (PU.1) locus at high resolution, using two different viewpoints (Figure S2D). We detected known interactions between the *SPI1* promoter and an upstream enhancer (URE) (Ebraldize et al., 2008), but also with two upstream promoters. These interactions did not change after RUNX1-ETO depletion. The same result was found using Chi-C (Figure S2E).

We next analyzed the behavior of LDB1 in more detail. LDB1 does not bind to DNA directly but via other TFs such as RUNX1 (Wadman et al., 1997). LDB1 binds to both promoter and distal regions (Figure S2E) and RUNX1-ETO depletion led to a loss of 1,506 LDB1 binding sites and the acquisition of 779 new sites (Figure S2F). De novo motif search of siMM- or siRE-specific LDB1 peaks revealed an enrichment of RUNX1 motifs in both binding site populations, which, however, occurred together with the AP-1 motif in siMM-specific peaks and with C/EBP binding motifs after RUNX1-ETO knockdown (Figure S2G). As expected from it being part of the RUNX1-ETO and RUNX1 complex, LDB1 binding correlated with interactions in control DHSs, which were lost after RUNX1-ETO depletion but also participated in new interactions (Figure S2F,

third panel). To test whether LDB1 was required for RUNX1-ETO binding, we depleted it using siRNA knockdown in Kasumi-1 cells with and without RUNX1-ETO knockdown (Figure S3A). LDB1 depletion led to an increase in cell death both by apoptosis and necrosis, but only in RUNX1-ETO-expressing cells, confirming that it is required for the maintenance of the leukemic phenotype (Sun et al., 2013) (Figure S3B). However, LDB1 was not required for RUNX1-ETO binding to chromatin (Figure S3C). LDB1 was also not required for the upregulation or repression of RUNX1-ETO target genes. As expected, RUNX1-ETO knockdown led to increases in expression of the direct RUNX1-ETO target genes *C/EBPA*, *CTSG*, and *NFE2* and decreases in *CD34* and *ERG* expression. Knockdown of LDB1 alone or together with RUNX1-ETO knockdown had no additive or negative effect on RUNX1-ETO target gene expression changes (Figure S3D). We therefore conclude that other factors besides RUNX1-ETO control LDB1 binding and determine its functional impact.

We next investigated whether the change in interaction was associated with altered TF occupancy. To this end, differential interactions were ranked by fold change in p value (Figure 3A), and associated DHSs were plotted alongside together with C/EBP $\alpha$ , JUND, LDB1, CTCF, RUNX1-ETO, PU.1, and RUNX1 ChIP-seq data. Beneath, we plotted the average profiles of factor binding for control and RUNX1-ETO-depleted cells (blue, ChIP signals associated with lost interactions; red, gained interactions). Differential interactions were associated with the differential binding of some, but not all, TFs. RUNX1-ETO-bound sites were associated with DHSs involved in both decreased and increased interactions, demonstrating that it is not the sole determinant of the interaction pattern. DHS associated with gained interactions showed a strong increase in C/EBP $\alpha$  as well as an increase in RUNX1 binding. Conversely, DHSs associated with decreased interactions after RUNX1-ETO knockdown lost JUND as well as LDB1 binding. An example for a downregulated gene is *CCND2* (Figure S2F), whose expression we have previously shown to be dependent on the presence of AP-1 factors (Martinez-Soria et al., 2019). Increased interactions did not involve an increase in LMO2 or PU.1 binding, and loss of interactions did not involve CTCF. The *CITED2* gene is an example of a gene with a new interaction driven by C/EBP $\alpha$  binding (Figure 2G,

## Figure 2. Differential Promoter-Enhancer Interactions after RUNX1-ETO Depletion Are Driven by Differential TF Binding

(A) Heatmap representing the correlation of normalized interaction ratios across chr3 at 5-kb resolution, showing the correlation of ChIP peaks in regions specific to DHS peaks that are depleted after RUNX1-ETO knockdown. Each pixel represents a 5-kb section of the genome. The left panel shows the interaction heatmap for siMM and the right panel for siRE cells. Positive correlations are shown as red; negative correlation as blue squares. To determine statistically significant interactions, reads from replicates 1 and 2 were merged.

(B) Heatmap representing the correlation of normalized interaction ratios across chr3 at 5-kb resolution and showing the correlation of Chi-C peaks in DHS peaks that are newly formed after RUNX1-ETO (R/E) gene knockdown. For all other features, see (A).

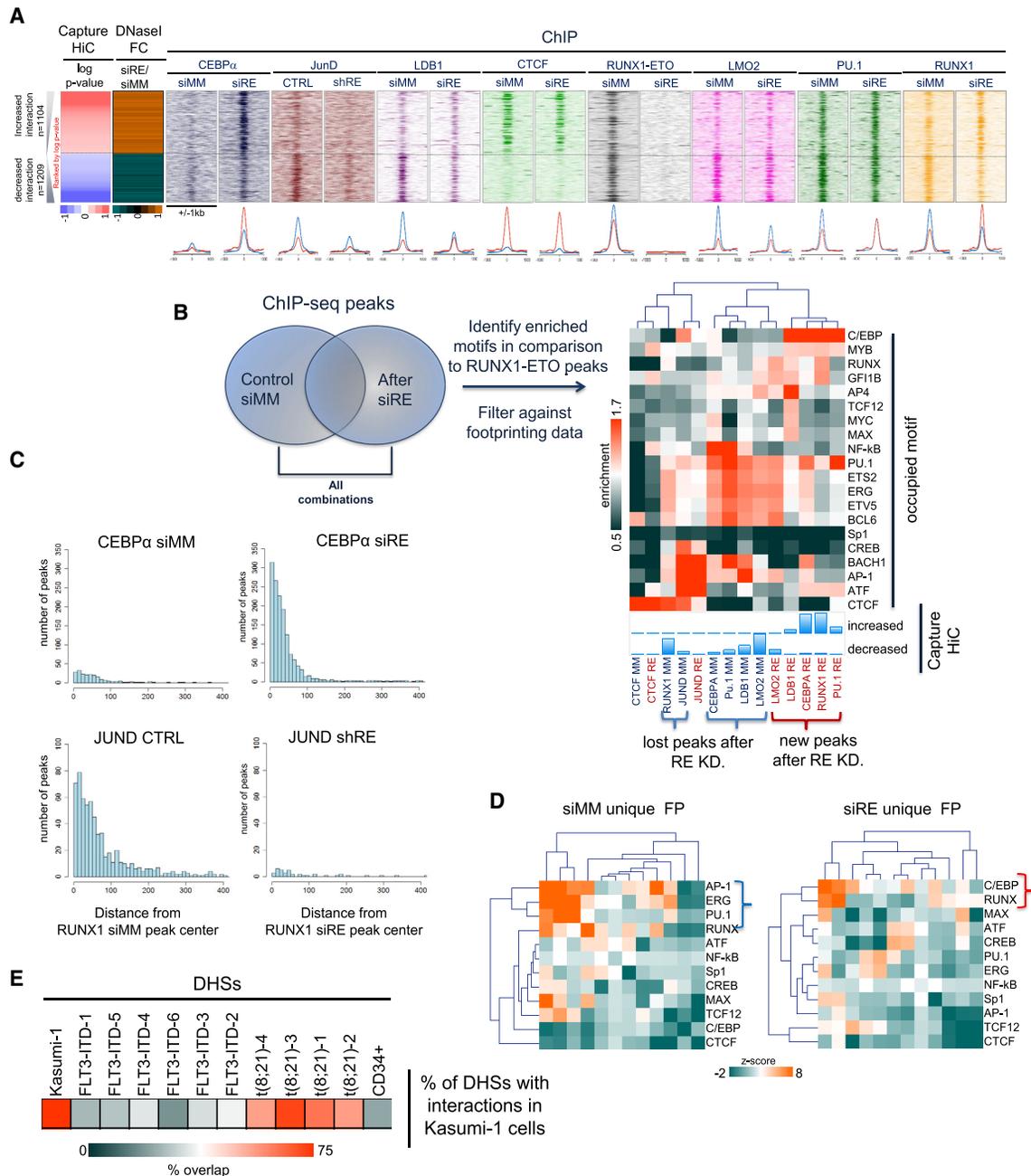
(C) DNaseI cleavage patterns within specific distal footprints predicted by Wellington (Piper et al., 2013). Upper strand cut sites are shown in red and lower strand cut sites in green within a 200-bp window centered on each footprint (gap) for siMM- and siRE-specific footprints.

(D) Analysis of overrepresented binding motifs within each footprint class as defined in (C).

(E) Left panel: time course of DHS development after 2, 4, and 10 days of RUNX1-ETO depletion (see scheme in Figure S1E). Normalized tag counts are ranked alongside day-10 knockdown (KD) and control-specific (bottom) counts; common and siRE-specific DHS are indicated on the left. Alongside the same genomic coordinates, C/EBP $\alpha$ , JUND, LDB1, CTCF, RUNX1-ETO, LMO2, PU.1, and RUNX1 ChIP-seq reads from Kasumi-1 cells with or without RUNX1-ETO depletion are plotted as indicated (middle panel). The right panel shows the expression levels of the genes linked to the associated DNaseI-seq sites (right panel).

(F) UCSC browser screenshot depicting interactions between the *CCND2* promoter and surrounding DHS (shown as arcs) together with the indicated ChIP-seq data before and after RUNX1-ETO knockdown. Changing interactions are shown in red, and their associated DHS/ChIP peaks are highlighted using a vertical shaded bar.

(G) The same analysis as in (F) for the *CITED2* locus.



**Figure 3. The Cooperation of Constitutive and Inducible TFs Is Associated with Differential Interactions**

(A) Log p values of the differential interactions were plotted ranked from high to low for control and RUNX1-ETO-depleted cells. Red represents an increase in interaction strength and blue represents a decrease. Alongside, the DNase-seq fold difference between control and RUNX1-ETO knockdown cells as well as ChIP-seq density profiles for C/EBP $\alpha$ , JUND, LDB1, CTCF, RUNX1-ETO, LMO2, and PU.1 are plotted from Kasumi-1 cells, transfected with either siMM or siRE as indicated. The panels below show the average profiles of the binding of the indicated TFs plotted around the peak summit for control and RUNX1-ETO-depleted cells. Red, ChIP signal specific for peaks with increased interactions; blue, ChIP signal specific for peaks with decreased interactions.

(B) Determination of enriched motifs for other TFs in ChIP-seq peaks specific for control and RUNX1-ETO-depleted cells. Motif enrichment was first identified using HOMER and then filtered against digital footprinting data from day 10 of knockout to ensure that these binding motifs were functional. Enrichment scores were subjected to unsupervised clustering for each of the indicated motifs (on the right). The heatmap depicts the degree of motif enrichment with highly enriched motifs shown in red. Peaks were overlaid with the DHS that show new interactions (red brackets at the bottom) or whose interactions are lost (blue brackets). Enrichment scores were calculated by the level of motif enrichment in the unique peaks, as compared to motif enrichment in RUNX1-ETO peaks. Bottom panels: percentage of peaks showing differential interaction with TFs binding to these sites as determined by ChIP-seq (control cells, blue; RUNX1-ETO-depleted cells, red).

(legend continued on next page)

shaded bar). In summary, our study shows that the main drivers of changes in cis-element interactions are the loss of RUNX1-ETO binding together with the loss of LDB1 and AP-1 binding along with the increased binding of C/EBP $\alpha$  and RUNX1 to new sites.

We next identified additional TFs associated with differential interactions and clustered TF binding motifs enriched in ChIP-seq peaks that either overlapped with new interaction sites or with sites lost after RUNX1-ETO depletion (Figure 3B, left panel and panels below the heatmap). Since the majority of DHS changes participating in differential interactions had already occurred at day 4 of knockdown (Figure 1E), we used our day-10 digital footprinting data to ensure that these motifs were functional and could be occupied. We then calculated the motif enrichment score of such motifs (depicted on the right) (Figure 3B, top-right panel). These analyses showed that the score of enriched motifs for RUNX1 and C/EBP family members increased in differential interactions upregulated after RUNX1-ETO depletion, together with an increase in GF11, MYB, and MYC/MAX binding site occupancy. In contrast, and in concordance with our ChIP-seq data, AP-1 motif enrichment decreased in interactions that were lost after RUNX1-ETO depletion, together with loss of activating TF (ATF) and nuclear factor  $\kappa$ B (NF- $\kappa$ B) motif occupancy. We also detected enrichment of motifs in both gained and lost interactions. This was true for ETS-family factors such as ERG and PU.1, but also for RUNX1, suggesting that factors move to other sites as shown previously (Lichtinger et al., 2012). To confirm this idea, we determined the distribution of distance between RUNX1 binding and other TFs before and after RUNX1-ETO depletion using the ChIP-seq data. This analysis showed a significant co-localization between AP-1 and RUNX1 peaks before, but not after, RUNX1-ETO depletion. In contrast, RUNX1 and C/EBP $\alpha$  show significant co-localization after RUNX1-ETO depletion (Figure 3C). In spite of the appearance of new RUNX1 binding sites after RUNX1-ETO depletion (Ptasinska et al., 2012, 2014), no significant changes were observed in the distribution of distance between RUNX1 and LDB1 and LMO2 and PU.1 peaks (Figures S3F–S3H), indicating no change in this type of factor collaboration. These analyses suggest that RUNX1 cooperates with different factors regulating different biological processes in control and RUNX1-ETO-depleted cells. To examine TF cooperation after the onset of monocytic differentiation, we performed a bootstrapping analysis (Figure 3D) in RUNX1-ETO-depleted and control cells that identified occupied TF binding motifs co-localizing with high significance within 50 base pairs (bp) as compared to the rest of the active genome (highlighted in red). This analysis again confirmed the strong co-association of occupied C/EBP and RUNX1 motifs in differentiated cells and AP-1, ETS, and RUNX motifs co-occur-

ences in control cells. Interestingly, the AP-1 or C/EBP motifs were not preferentially footprinted in the DHSs shared between control and knockout cells and did not co-localize with other motifs (Figure S3I), indicating that co-localizing (RUNX1-AP-1) sites are part of the AML-specific cistrome. In summary, these analyses demonstrated that the establishment of specific RUNX1-ETO-dependent cis-element interactions are mediated by the cooperation of a limited set of constitutive and inducible TFs. The depletion of RUNX1-ETO drives the loss and relocation of TFs and thus the establishment of new interactions via new factor collaborations.

### The Construction of Transcriptional Networks Grounded in Multi-omics Data

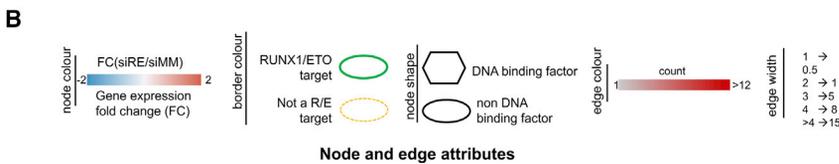
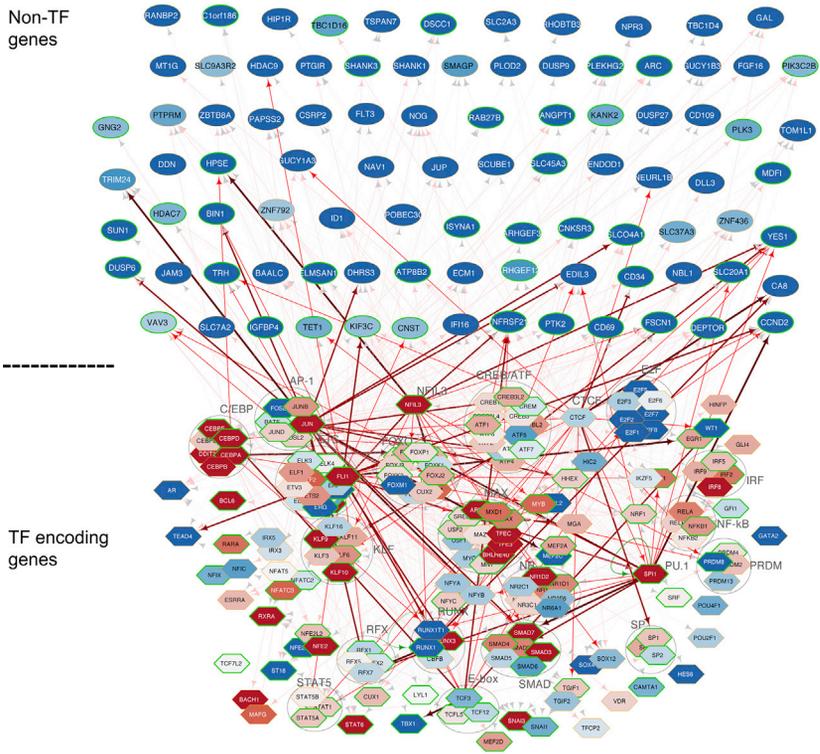
The Kasumi-1 cell line is one of the best-studied human models of t(8;21) AML with numerous multi-omics data available that should be amenable to modeling approaches predicting transcriptional network behavior in response to perturbation. So far we have assigned factor binding site data only to their nearest promoter. However, numerous studies have shown that such assignments were not accurate (Mifsud et al., 2015; Sanyal et al., 2012). In our study, we found that only about 40% of all cis-regulatory elements in control cells interacted with their nearest promoter. Our ChIP-C data enabled us to assign DHSs containing active cis-elements and footprinted regions to promoters (Data S2). More than 70% of all DHSs assigned to their rightful promoter in Kasumi-1 cells were also present in t(8;21) but not in FLT3-ITD patients or normal CD34+ hematopoietic stem cell (HPSCs; Figure 3E; Assi et al., 2019). To construct gene regulatory networks and to examine how these networks shift after RUNX1-ETO depletion and differentiation, we used our footprinting data (control and day-10 siRE) to assign occupied motifs to specific TF families capable of binding to this motif (Table S1, indicated as groups in Figure 4). We then plotted the connections between factors and genes that were downregulated (Figure 4B, blue ovals) or upregulated (Figure S4, red ovals) by at least 2-fold following RUNX1-ETO depletion at day 10, with the former being markers for the leukemic and the latter being markers for differentiated states. We also highlighted which genes were RUNX1-ETO targets (green boundary). This analysis shows a complex web of interactions between effector genes (lined up at the top) and TF encoding genes, many of which are known to respond to RUNX1-ETO depletion, such as C/EBP $\alpha$  or IRF8. The networks highlight the TFs involved in differentiation, again showing that increased C/EBP $\alpha$  activity is the main driver of the changes of the t(8;21) transcriptional network after RUNX1-ETO depletion, with C/EBP family members binding to multiple differentiation-specific cis-regulatory elements and driving the upregulation of their respective genes (Loke et al., 2018;

(C) Bar plots illustrating the distribution of distances between the binding sites of the indicated TFs as determined by ChIP-seq. We measured the changing distance between RUNX1 peaks in siMM and siRE cells and C/EBP $\alpha$  peaks in siMM (top left) and siRE cells (top right), as well as the distance between RUNX1 peaks and JUND control peaks (bottom left) and JUND after R/E KD (bottom right).

(D) Bootstrapping analysis of the significance of co-localizing of footprinted motifs within day-10 DHSs for sites that are either lost (left panel) or gained (right panel) after RUNX1-ETO depletion as compared to the rest of the genome. The heatmap shows the significance of motifs co-localizing within 50 bp as compared to sampling by chance.

(E) Heatmap highlighting the percentage of day-4 Kasumi-1 DHSs with interactions found in different patient groups indicating the similarity between cell-line and primary t(8;21) data. The t(8;21) and FLT3-ITD DHS/ChIP-C patient data were downloaded from GEO: GSE108316 (Assi et al., 2019).

**A** Non-TF genes down regulated by R/E depletion → Link to TF families via footprints → Gene regulatory network of RUNX1-ETO responsive genes



**Figure 4. Differentially Expressed Genes after RUNX1-ETO Knockdown Are Regulated by Different TF Networks**

(A) Top panel: data analysis strategy. Transcriptional network of downregulated (blue) non-TF (effector) genes after RUNX1-ETO knockdown (top rows) connected to genes encoding TF families (bottom rows) as determined by digital footprinting and CHI-C. Arrows going outward can come from any TF family within a group; incoming arrows are specific for each gene.  
(B) Node and edge attributes.

Ptasinska et al., 2014). An example of a downregulated gene specific to the leukemic state includes *UBASH3B*, which has previously been shown to regulate the proliferation of t(8;21) cells (Goyama et al., 2016). Another such example is *YES1*, which, together with another downregulated gene, *MEIS2*, is involved in maintaining leukemic growth (Vegi et al., 2016). AP-1 members are important for maintaining the leukemic growth phenotype, as shown by expressing a dominant-negative FOS protein in t(8;21) cells. Expression of this peptide downregulates the expression of several cell cycle genes, including *CCND2* (Martinez-Soria et al., 2019), and blocks tumor growth *in vivo* (Assi et al., 2019). Such examples of properly annotated RUNX1-ETO-responsive genes with known function show the quality of our analysis with respect to the prediction of important genes required for AML maintenance. Last but not least, our studies serve as paradigm for how high quality multi-omics data can be used to generate in-depth information on the regulatory circuitries of a specific type of AML.

## STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Line Culture
- METHOD DETAILS
  - siRNA Mediated Depletion of RUNX1-ETO or LDB1
  - RNA Extraction
  - RNA Seq Libraries
  - cDNA Synthesis
  - Real-Time Polymerase Chain Reaction
  - Dead Cell Removal and Annexin V/PI Staining for Flow Cytometry
  - DNaseI Hypersensitivity Site Mapping

- Library Production of DNase I Material for High Throughput Sequencing
- ChIP-qPCR and ChIP-Seq Library Preparation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - DNaseI Sequencing Data Analysis
  - ChIP Sequencing Data Analysis
  - Average Tag Density Profile and Heatmap
  - RNA-Seq Data Analysis
  - Promoter Capture HiC Data Analysis
  - 4C-Seq Data Analysis
  - Motif Identification and Clustering
- **DATA AND CODE AVAILABILITY**

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.08.040>.

#### ACKNOWLEDGMENTS

We are grateful to the Genomics Birmingham Sequencing Facility for their expert sequencing service. This work was funded by program grants from Bloodwise to C.B. and P.N.C. (15001) and to C.S.O. (14007), as well as Cancer Research UK studentships to C.B. and A. Pickin. Work in R.D.'s lab is supported by the Dutch Cancer Foundation KWF (grant EMCR 2013-5829), Worldwide Cancer Research (grant 12-1309), and the Tata Memorial Trust Foundation.

#### AUTHOR CONTRIBUTIONS

A. Pickin conducted the CHI-C and 4C experiments; A. Ptasińska designed and performed experiments and wrote the manuscript; P.S.C. performed experiments; S.A.A. analyzed the data; R.D., S.G., and R.A. helped with the 4C methodology; C.S.O. helped in establishing the CHI-C methodology and in writing the paper; P.N.C. helped in designing the study and in writing the paper; and C.B. designed the study and wrote the paper.

#### DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: October 31, 2018

Revised: June 7, 2019

Accepted: August 12, 2019

Published: September 17, 2019

#### REFERENCES

Assi, S.A., Imperato, M.R., Coleman, D.J.L., Pickin, A., Potluri, S., Ptasińska, A., Chin, P.S., Blair, H., Cauchy, P., James, S.R., et al. (2019). Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat. Genet.* *51*, 151–162.

Bejjani, F., Evanno, E., Zibara, K., Piechaczyk, M., and Jariel-Encontre, I. (2019). The AP-1 transcriptional complex: Local switch or remote command? *Biochim. Biophys. Acta Rev. Cancer* *1872*, 11–23.

Ben-Ami, O., Friedman, D., Leshkowitz, D., Goldenberg, D., Orlovsky, K., Pencovich, N., Lotem, J., Tanay, A., and Groner, Y. (2013). Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep.* *4*, 1131–1143.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.

Bonifer, C., and Cockerill, P.N. (2011). Chromatin mechanisms regulating gene expression in health and disease. *Adv. Exp. Med. Biol.* *711*, 12–25.

Bonifer, C., and Cockerill, P.N. (2015). Chromatin Structure Profiling Identifies Crucial Regulators of Tumor Maintenance. *Trends Cancer* *1*, 157–160.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* *48*, 1193–1203.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* *502*, 499–506.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* *149*, 1233–1244.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.

Dunne, J., Cullmann, C., Ritter, M., Soria, N.M., Drescher, B., Debernardi, S., Skoulakis, S., Hartmann, O., Krause, M., Krauter, J., et al. (2006). siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. *Oncogene* *25*, 6067–6078.

Ebralidze, A.K., Guibal, F.C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M.A., Petkova, V., Rosenbauer, F., Huang, G., et al. (2008). PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev.* *22*, 2085–2092.

Erickson, P., Gao, J., Chang, K.S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992). Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood* *80*, 1825–1831.

Gonzalez-Sandoval, A., and Gasser, S.M. (2016). On TADs and LADs: Spatial Control Over Gene Expression. *Trends Genet.* *32*, 485–495.

Goyama, S., Schibler, J., Gasilina, A., Shrestha, M., Lin, S., Link, K.A., Chen, J., Whitman, S.P., Bloomfield, C.D., Nicolet, D., et al. (2016). UBASH3B/Sts-1-CBL axis regulates myeloid proliferation in human preleukemia induced by AML1-ETO. *Leukemia* *30*, 728–739.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* *157*, 369–381.

Heidenreich, O., Krauter, J., Riehle, H., Hadwiger, P., John, M., Heil, G., Vormlocher, H.P., and Nordheim, A. (2003). AML1/MTG8 oncogene suppression by small interfering RNAs supports myeloid differentiation of t(8;21)-positive leukemic cells. *Blood* *101*, 3157–3163.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* *38*, 576–589.

Kitabayashi, I., Ida, K., Morohoshi, F., Yokoyama, A., Mitsunashi, N., Shimizu, K., Nomura, N., Hayashi, Y., and Ohki, M. (1998). The AML1-MTG8 leukemic fusion protein forms a complex with a novel member of the MTG8(ETO/CDR) family, MTGR1. *Mol. Cell. Biol.* *18*, 846–858.

Kumar, C.C. (2011). Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes Cancer* *2*, 95–107.

Kumar, V., Muratani, M., Rayan, N.A., Kraus, P., Lufkin, T., Ng, H.H., and Prabhakar, S. (2013). Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* *31*, 615–622.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289–293.

Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S.A., Lie-A-Ling, M., Noailles, L., Vijayabaskar, M.S., Wu, M., et al. (2012). RUNX1

- reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.* 31, 4318–4333.
- Loke, J., Assi, S.A., Imperato, M.R., Ptasinska, A., Cauchy, P., Grabovska, Y., Soria, N.M., Raghavan, M., Delwel, H.R., Cockerill, P.N., et al. (2017). RUNX1-ETO and RUNX1-EVI1 Differentially Reprogram the Chromatin Landscape in t(8;21) and t(3;21) AML. *Cell Rep.* 19, 1654–1668.
- Loke, J., Chin, P.S., Keane, P., Pickin, A., Assi, S.A., Ptasinska, A., Imperato, M.R., Cockerill, P.N., and Bonifer, C. (2018). C/EBP $\alpha$  overrides epigenetic reprogramming by oncogenic transcription factors in acute myeloid leukemia. *Blood Adv.* 2, 271–284.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Lutterbach, B., Westendorf, J.J., Linggi, B., Patten, A., Moniwa, M., Davie, J.R., Huynh, K.D., Bardwell, V.J., Lavinsky, R.M., Rosenfeld, M.G., et al. (1998). ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol. Cell Biol.* 18, 7176–7184.
- Martens, J.H., Mandoli, A., Simmer, F., Wierenga, B.J., Saeed, S., Singh, A.A., Altucci, L., Vellenga, E., and Stunnenberg, H.G. (2012). ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood* 120, 4038–4048.
- Martinez, N., Drescher, B., Riehle, H., Cullmann, C., Vornlocher, H.P., Ganser, A., Heil, G., Nordheim, A., Krauter, J., and Heidenreich, O. (2004). The oncogenic fusion protein RUNX1-CBFA2T1 supports proliferation and inhibits senescence in t(8;21)-positive leukaemic cells. *BMC Cancer* 4, 44.
- Martinez-Soria, N., McKenzie, L., Draper, J., Ptasinska, A., Issa, H., Potluri, S., Blair, H.J., Pickin, A., Isa, A., Chin, P.S., et al. (2019). The Oncogenic Transcription Factor RUNX1/ETO Corrupts Cell Cycle Regulation to Drive Leukemic Transformation. *Cancer Cell* 35, 705.
- Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606.
- Mifsud, B., Martincorena, I., Darbo, E., Sugar, R., Schoenfelder, S., Fraser, P., and Luscombe, N.M. (2017). GOTHIC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PLoS ONE* 12, e0174744.
- Miyoshi, H., Shimizu, K., Kozu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991). t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc. Natl. Acad. Sci. USA* 88, 10431–10434.
- Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res.* 41, e201.
- Plank, J.L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. *Mol. Cell* 55, 5–14.
- Ptasinska, A., Assi, S.A., Mannari, D., James, S.R., Williamson, D., Dunne, J., Hoogenkamp, M., Wu, M., Care, M., McNeill, H., et al. (2012). 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.
- Ptasinska, A., Assi, S.A., Martinez-Soria, N., Imperato, M.R., Piper, J., Cauchy, P., Pickin, A., James, S.R., Hoogenkamp, M., Williamson, D., et al. (2014). Identification of a dynamic core transcriptional network in t(8;21) AML that regulates differentiation block and self-renewal. *Cell Rep.* 8, 1974–1988.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* 7, 105–117.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109–113.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.
- Sun, X.J., Wang, Z., Wang, L., Jiang, Y., Kost, N., Soong, T.D., Chen, W.Y., Tang, Z., Nakadai, T., Elemento, O., et al. (2013). A stable transcription factor complex nucleated by oligomeric AML1-ETO controls leukaemogenesis. *Nature* 500, 93–97.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53.
- van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* 9, 969–972.
- Vegi, N.M., Klappacher, J., Oswald, F., Mulaw, M.A., Mandoli, A., Thiel, V.N., Bamezai, S., Feder, K., Martens, J.H.A., Rawat, V.P.S., et al. (2016). MEIS2 Is an Oncogenic Partner in AML1-ETO-Positive AML. *Cell Rep.* 16, 498–507.
- Wadman, I.A., Osada, H., Grütz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* 16, 3145–3157.
- Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* 4, 1310.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
ETO	Diagenode	Cat# C15310197
RUNX1 (C-terminal epitope)	Abcam	Cat# 23980; RRID: AB_2184205
C/EBP $\alpha$	Abcam	Cat# 40761; RRID: AB_726792
LDB1	Abcam	Cat# 96799; RRID: AB_10679400
LMO2	R&D	Cat# AF2726; RRID: AB_2249968
CTCF	Abcam	Cat# 70303; RRID: AB_1209546
JUND	Santa Cruz	Cat# sc74; RRID: AB_2130177
GAPDH	Sigma	Cat# F3165; RRID: AB_259529
Anti-Rabbit HRP	Cell signalling	Cat# 7074; RRID: AB_2099233
Anti-Mouse HRP	Jackson	Cat# 115-0350-62; RRID: AB_2338504
CD34 Monoclonal Antibody (4H11), PE	eBioscience	Cat# 12-0349-42; RRID: AB_1548680
c-Kit Monoclonal Antibody (104D2), FITC	eBioscience	Cat# 11-1178-42; RRID: AB_2572472
<b>Oligonucleotides</b>		
Oligonucleotide sequences, see <a href="#">Table S2</a>	This Paper	NA
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
TruSeq Stranded mRNA with Ribo-Zero human assay	Illumina	Cat# 20020596
TruSeq RNA Sample Prep kit	Illumina	Cat#RS-122-2001
KAPA hyper Prep Kit	Kapa Biosystems	KK8500
KAPA Library Quantification kit	Kapa Biosystems	Cat#KK4824
MinElute Gel Extraction Kit	QIAGEN	Cat#28604
DNaseI	Worthington,	DPPF grade
AMPure XP beads	Beckman Coulter	Cat#A63882
NextSeq500 High output 150 cycles	Illumina	Cat#FC-404-2002
NextSeq500 High output 75 cycles	Illumina	Cat#FC-404-2005
Nucleospin RNA column	Machery Nagel,	740955.50
OligoDT primer	Promega	C110A
Murine Moloney Virus reverse transcriptase	Promega	M170A
RNase Inhibitor	Promega	N261A
Sybr Green mix	Applied Biosystems	4309155
Proteinase K	Roche	03115801001
Di(N-succinimidyl) glutarate (DSG)	Sigma-Aldrich	80424
MyOne Streptavidin C1	DynaBeads Invitrogen	65601
MyOne Streptavidin T1	DynaBeads Invitrogen	65001
Formaldehyde	(Pierce, Thermo Scientific, USA	28906
(NEB)2 buffer	NEB	37002
<b>Critical Commercial Assays</b>		
Dead Cell Removal microbeads	Miltenyi Biotec	130090101
Annexin V-APC/PI staining	Ebiosciences	88-8007-74
High Sensitivity DNA Chips	Agilent Technologies	5067-4626
RNA Pico Chips	Agilent Technologies	5067-1513
High Pure PCR Product Purification Kit	Roche	11732676001
SureSelect target enrichment	Agilent Technologies	5190-4393
HindIII	NEB	R0104M

(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
CHIP-seq data	This study	GSE121282
DNaseI-seq data	This study	GSE121282
RNA-seq data	This study	GSE121282
Capture HiC	This study	GSE117108
reprogramming ATAC-Seq	<a href="#">Corces et al., 2016</a>	GSE75384
Published ChIP-Seq	<a href="#">Ptasinska et al., 2014</a>	GSE60131
Experimental Models: Cell Lines		
Kasumi-1 human cell line	DSMZ	ACC220
Software and Algorithms		
Bowtie 2	<a href="#">Langmead and Salzberg, 2012</a>	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
MACS2	<a href="#">Zhang et al., 2008</a>	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
Homer	<a href="#">Heinz et al., 2010</a>	<a href="http://homer.ucsd.edu/homer/motif/">http://homer.ucsd.edu/homer/motif/</a>
Cufflinks v2.2.1	<a href="#">Trapnell et al., 2013</a>	<a href="http://cole-trapnell-lab.github.io/cufflinks/announcements/protocol-paper/">http://cole-trapnell-lab.github.io/cufflinks/announcements/protocol-paper/</a>
Bedtools	<a href="#">Quinlan and Hall, 2010</a>	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
STAR	<a href="#">Dobin et al., 2013</a>	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
Wellington algorithm	<a href="#">Piper et al., 2013</a>	<a href="https://pythonhosted.org/pyDNase/">https://pythonhosted.org/pyDNase/</a>
Illustrator	Adobe System Software Ireland	<a href="https://www.adobe.com/cn/products/cs6/illustrator.html">https://www.adobe.com/cn/products/cs6/illustrator.html</a>
ZEN	Zeiss	<a href="https://www.zeiss.com/microscopy/int/microscope-cameras.html">https://www.zeiss.com/microscopy/int/microscope-cameras.html</a>
GraphPad Prism 6.0	GraphPad Software	<a href="https://www.graphpad.com/scientificsoftware/prism/">https://www.graphpad.com/scientificsoftware/prism/</a>

## LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Constanze Bonifer ([c.bonifer@bham.ac.uk](mailto:c.bonifer@bham.ac.uk)).

This study did not generate new reagents.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cell Line Culture

Cells were maintained in a humidified incubator at 37°C with 5% CO<sub>2</sub>. t(8;21) Kasumi-1 cells were cultured in RPMI with 10% FCS supplemented with 1% glutamine and 1% penicillin/streptomycin.

## METHOD DETAILS

### siRNA Mediated Depletion of RUNX1-ETO or LDB1

1x10<sup>7</sup> cells were electroporated using a EPI 3500 (Fischer, Germany) single 350 V pulse for 10ms. After electroporation, the cells remained in their cuvettes for 10 minutes before being directly added to RPMI-1640 with 10% FCS, supplemented with penicillin/streptomycin and glutamine at a concentration of 0.5 x10<sup>6</sup> cells per ml and returned to an incubator kept at 37°C and 5% CO<sub>2</sub>. siRNA sequences (SIGMA ALDRICH Germany) specific for the translocation breakpoint of RUNX1-ETO were 5' CCUCGAAAUCGUACU GAGAAG -3' (sense) and 5'- UCUCAGUACGAUUUCGAGGUU-3' (antisense). Control siRNA was 5'-CCUCGAAUUCGUUCUGA GAAG-3' (sense) with 5'-UCUCAGAACGAUUUCGAGGUU-3' (antisense). siRNA sequences specific for LDB1 ON-TARGETplus Human LDB1 siRNA SMARTpool (L-016010-00-0005, Dharmacon). siRNA was used at 200 nM.

### RNA Extraction

RNA from Kasumi-1 cells was purified using a Nucleospin RNA column (Machery Nagel, France), according to manufacturer's instructions. The quality of RNA from was assessed using a spectrophotometer, by the ratio of the absorbance at 260 nm and

280 nm wavelengths. RNA has a greater absorbance in the 260 nm wavelength, Eukaryotic Total RNA PICO Bioanalyser chip (Agilent technologies, USA) allows visualization of the size of the RNA molecules and thus, demonstrates whether the sample is degraded or not.

### RNA Seq Libraries

RNA-seq libraries were prepared with a Total RNA Ribo-zero library preparation kit (with ribosomal RNA depletion) (Illumina, USA) according to manufacturer's instructions with the following alterations: 15 cycles of PCR was undertaken to amplify the library and adaptors for multiplexing were used at a 1:4 dilution. Library quality was checked by running the samples on a Bioanalyser and libraries were quantified using a Kapa library quantification kit (Kapa Biosystems, USA) and run in a pool of eight indexed libraries in two lane of a HiSeq 2500 (Illumina, USA) using rapid run chemistry with 100bp paired end reads.

### cDNA Synthesis

1  $\mu\text{g}$  RNA was used to make cDNA with 0.5  $\mu\text{g}$  OligoDT primer, Murine Moloney Virus reverse transcriptase and RNase Inhibitor (Promega, USA) according to the manufacturer's protocol.

### Real-Time Polymerase Chain Reaction

RT-PCR was performed using Sybr Green mix (Applied Biosystems, UK), at 2x dilution. Primers were used at 100 nM final concentration. cDNA was diluted 1:50 depending on expression levels of targets. A 7900HT system (Applied Biosystems, UK) was used to perform qPCR. Primers used in this project are listed in [Table S2](#).

### Dead Cell Removal and Annexin V/PI Staining for Flow Cytometry

Dead cell removal was performed using negative selection on a MS column following incubation with Dead Cell Removal microbeads (Mitenyi Biotech, USA) as per manufacturer's instructions. Dead cell removal was performed on all samples prior to RNA extraction or DHSs mapping. Annexin V-APC/PI staining (Ebiosciences, USA) or was performed according to manufacturer's instructions.

### DNaseI Hypersensitivity Site Mapping

Prior to DNaseI digestion, apoptotic cells were removed using the Dead Cell Removal Kit (Miltenyl Biotech, UK) as per manufacturer's instructions.  $3 \times 10^7$  Kasumi-1 cells were suspended in 1 mL DNase I buffer (0.3M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl<sub>2</sub>, 10 mM Tris pH7.4). Digestion on  $4.5 \times 10^6$  cells was performed with DNase I (Worthington, DPPF grade) at 80 units/ml in DNase I buffer with 0.4% NP-40 and 2 mM CaCl<sub>2</sub> at 22°C for 3 minutes. The reaction was stopped with cell lysis buffer (0.3M NaAcetate, 10mM EDTA pH 7.4, 1% SDS) with 1mg/ml Proteinase K and incubated at 45°C overnight. The digested DNase I material was treated with RNase A (Sigma Aldrich, Germany) at a final concentration of 100  $\mu\text{g}/\text{ml}$  at 37°C for 1 hr. Genomic DNA was extracted using phenol/chloroform method: an equal volume of phenol was added to the reaction and placed on a rotator wheel for 45 minutes. This was centrifuged for 5 minutes at 16000 x g at room temperature. The top layer was transferred to a new tube and the process was repeated sequentially with phenol/chloroform and chloroform. After purification by chloroform extraction, genomic DNA was precipitated with ethanol. This was pelleted by centrifugation for 5 minutes, at 16000 x g at 4°C. The pellet was resuspended with 70% ethanol and centrifugation for 5 minutes, at 16000 x g at 4°C. The pellet was air-dried and dissolved by Tris-EDTA (40 mM Tris Acetate 1 mM EDTA). Digestion was checked visually by running the samples on a 0.7% agarose gel and by RT-PCR evaluating the ratio of open (TBP promoter) to closed regions of DNA (chromosome 18) and active gene body (beta-actin) to prevent selection of over digested samples. Primers used in this project are listed in [Table S2](#). Subsequently, between 2 to 10  $\mu\text{g}$  of DNase I-digested DNA (depending on material available) were run on a 1.5% agarose gel for selection of shorter fragments to increase the fraction of fragments captured from DHSs. Prior to loading on gel, the purified DNA was treated again with RNase A (Sigma Aldrich, USA) at a final concentration of 100  $\mu\text{g}/\text{ml}$  at 37°C for 1 hr. 50-300 bp fragments were isolated and purified from the gel using a MinElute gel extraction kit (QIAGEN, USA) as per manufacturer's instructions and validated by qPCR. Following this, the size selected sample was validated again by RT-PCR, this time using shorter amplicons to enable detection of the shorter fragments enriched by the size selection process.

### Library Production of DNase I Material for High Throughput Sequencing

After size selection, a library was prepared using KAPA Hyper Prep Kit sample preparation kit (Kapa Biosystems, USA) as per manufacturer's protocol. After PCR a final size selection step was performed by running the library on 2% TAE gel, followed by excision of 190-250 bp sized gel fragment. The library was purified from the gel using a MinElute gel extraction kit (QIAGEN, USA). The quality of the libraries was assessed on an Agilent 2100 Bioanalyser. Libraries were subsequently run on two lanes of an Illumina HiSeq 2500 flow-cell for transcription factor footprinting, or as part of 12 indexed libraries in one lane of a NextSeq500 (Illumina, USA) for DHS mapping alone.

### ChIP-qPCR and ChIP-Seq Library Preparation

#### Double Cross-Linking

A double cross-linking technique was used to optimize the efficiency of transcription factor chromatin immunoprecipitation (ChIP).  $2 \times 10^7$  cells were washed thrice in PBS. Di(N-succinimidyl) glutarate (DSG) (Sigma-Aldrich, Germany) at 850  $\mu\text{g}/\text{ml}$  was added to

$2 \times 10^7$  cells per ml and were incubated for forty-five minutes. Cells were washed four times and fixed with 1% formaldehyde (Pierce, Thermo Scientific, USA) for ten minutes. Glycine to produce a final concentration of 100mM was added to stop the reaction. The pellet was washed again with PBS. Buffer A (HEPES pH 7.9 10 mM, EDTA 10 mM, EGTA 0.5 mM, Triton x100 0.25%, complete mini protease inhibitor cocktail (PIC) 1x (Sigma-Aldrich, Germany) was added for 10 mins at 4°C and removed by centrifugation at 500 g for 5 minutes. This was repeated with buffer B (HEPES pH 7.9 10 mM, EDTA 1 mM, EGTA 0.5 mM, Triton x100 0.01%, PIC 1x). The residual nuclei were then spun down at 16000 x g at 4°C for 5 minutes and aliquoted at  $2 \times 10^7$  cells for 4 immunoprecipitations.

### **Chromatin Immunoprecipitation (ChIP)**

Each aliquot of  $2 \times 10^7$  cells was re-suspended in 600  $\mu$ L of sonication buffer (Tris-HCL pH 8 25 mM, NaCL 150 mM, EDTA 2 mM, Triton 100x 1%, SDS 0.25%, Protease inhibitor cocktail (PIC) 1x). 300  $\mu$ L of nuclei in sonication buffer was placed in each polystyrene tube and sonicated at 75% amplitude, 26 cycles: 30 s on and 30 s off per cycle (Q800, Active Motif, USA). Subsequently, 1.2ml of dilution buffer (Tris-HCL pH8 25 mM, NaCL 150 mM, EDTA 2 mM, Triton 100x 1%, glycerol 7.5%, PIC 1x) was added to the pooled post sonication material. This was divided equally between four immunoprecipitations (with 5% of input taken for validation). 20  $\mu$ L protein G beads (Diagenode, Belgium) were washed twice with 500  $\mu$ L of 50 mM citrate phosphate buffer and once with 100 mM sodium phosphate. 4  $\mu$ g antibody ETO (Santa Cruz) or 4  $\mu$ g antibody AML1-ETO (15310197, Diagenode), or RUNX1 (Ab23980, Abcam) or 4 $\mu$ g antibody C/EBP $\alpha$  (A2814, Santa Cruz) or 2 $\mu$ g antibody LBD1 (96799, Abcam) or 2 $\mu$ g antibody LMO2 (AF2726, R&D) or 2 $\mu$ g antibody CTCF (70303, Abcam) or 2 $\mu$ g JUND (sc74, Santa Cruz) was added to 10  $\mu$ L 100 mM sodium phosphate, 0.5% BSA and incubated with protein G beads at 4°C for 1 hour. Chromatin was then added to the protein G beads with antibody and returned to 4°C for 4 hours. Unbound chromatin was separated from the beads by magnet and the attached beads were washed by buffer 1 (Tris HCL 20 mM, NaCl 150 mM, EDTA 2 mM, Triton x100 1%, SDS 0.1%), twice with buffer 2 (Tris HCL 20 mM, NaCl 500 mM, EDTA 2 mM, Triton x100 1%, SDS 0.1%), LiCL buffer (Tris HCL 10 mM, LiCl 250 mM, EDTA 1 mM, NP40 0.5%, sodium deoxycholate 0.5%) and finally twice with wash buffer 4 (Tris HCL pH8, 10 mM, NaCl 50 mM, EDTA 1mM). The column was eluted twice with 50  $\mu$ L buffer (NaHCO<sub>3</sub> 100 mM and SDS 1%) and the eluant containing the chromatin was pooled. Crosslinks were reversed by incubating the samples at 65°C overnight in 500 mM NaCl, 500  $\mu$ g/ml proteinase K. DNA was purified by Ampure beads (Beckman Coulter, USA), as above, with the DNA eluted with 50  $\mu$ L water. Validation of the ChIP was performed by qPCR using a standard curve of genomic DNA from untreated Kasumi-1 cells (10ng/  $\mu$ l followed by serial 1:5 dilutions). The input material was diluted 1:5 with water. Primers used in this project are listed in [Table S2](#). Validation was analyzed as a ratio of the qPCR signal from the ChIP material over the input.

### **Library Production of ChIP Material for High Throughput Sequencing**

Libraries for high throughput sequencing were prepared using the Tru-seq DNA sample preparation kit (Illumina, USA) or Kapa HyperPrep kit (Kapa Biosystems, USA), as per manufacturer's protocol. 18 cycles of PCR was performed and 200-350bp fragments were size selected by running the samples in an agarose gel. Libraries were purified from the gel using a MinElute Gel extraction kit (QIAGEN, USA). Libraries were validated by qPCR, with an analysis of the ChIP signal of a positive control region (e.g., PU.1 3H enhancer) over a negative control region (e.g., *IVL*). Finally, libraries were quantified by Kapa library quantification kit (Kapa Biosystems, USA) and run in a pool of four indexed libraries in one lane of a HiSeq 2500 (Illumina, USA) or 12 indexed libraries in one lane of a NextSeq 500 (Illumina, USA) using 50 cycle single-end reads.

### **Circularized Chromosome Conformation Capture (4C-seq)**

4C analysis was performed exactly as described in [Gröschel et al. \(2014\)](#).  $1 \times 10^7$  Kasumi-1 cells, transfected with mismatch siRNA (siMM) or siRNA specific to siRUNX1-ETO (siRNA), were fixed with 2% formaldehyde and incubated for 10 minutes at room temperature. 1.425 mL of 1M glycine was added to quench the cross-linking reaction. Fixed cells were immediately centrifuged for 8 minutes at 4°C, 500 xg. Supernatant was removed and the pellet resuspended in 1ml lysis buffer (500 $\mu$ l 1M TRIS pH 7.5, 300 $\mu$ l 5M NaCl, 100 $\mu$ l 0.5M EDTA, 250 $\mu$ l 20% NP-40 and 100 $\mu$ l Triton X-100 made up to 10ml with H<sub>2</sub>O) and incubated at room temperature for 5 minutes, followed by 5 minutes at 65°C. Cells were then kept on ice while complete cell lysis was determined via Trypan blue (GIBCO) staining. Cells were centrifuged at 800 xg for 5 minutes and the pellet was taken up in 440  $\mu$ l H<sub>2</sub>O and 60  $\mu$ l 10X RE buffer 2 (NEB). 15  $\mu$ l of SDS was added and the tube placed at 37°C for 1 hour. 75  $\mu$ l of 20% Triton X-100 was added and the tube incubated at 37°C for 1 hour. A 5  $\mu$ l aliquot was removed as an 'undigested control' sample before 200 units of the restriction enzyme DpnII was added. The tube was incubated for 4 hours at 37°C, and then another 200 units of DpnII was added, followed by an overnight 37°C incubation. The following day 200 units of DpnII was added for 4 hr at 37°C. A 5  $\mu$ l aliquot was removed as a 'digested control' sample. To this, along with the 'undigested' sample, 90  $\mu$ l of 10mM Tris pH 7.5 and 5 $\mu$ l Proteinase K (10 mg/ml) was added to reverse the cross links. These control samples were run on a 0.6% agarose gel to assess the digestion efficiency. All 37°C incubations were conducted in a heated block, shaking at 900 RPM. DpnII was selected as the restriction enzyme as it functions in SDS, and combined with the second restriction enzyme (Csp6I) it generates restriction fragments near the target loci, with a suitable size for efficient ligation and PCR amplification. Both of these enzymes are 4bp cutters, so will cut the genome into 256 bp fragments, on average. This allows for a high resolution assay. The DpnII was inactivated by incubation at 65°C for 20 minutes. On ice, 700  $\mu$ l of 10X ligation buffer, 7 mL of milli-Q H<sub>2</sub>O and 10  $\mu$ l T4 Ligase (Roche 5U/ $\mu$ l) were added then samples were incubated overnight at 16°C. The following day, to assess ligation efficiency, a 100  $\mu$ l aliquot of the sample was taken as the 'ligated control'. The crosslinks were reversed as above and the sample run on a 0.6% agarose gel. To reverse the crosslinks, 30  $\mu$ l Prot K (10mg/ml) was added and samples were left overnight at 65°C. The next day, 30  $\mu$ l RNase A (10mg/ml) was added and samples were incubated for 45 minutes at 37°C. DNA was extracted by adding 7 mL phenol-chloroform. Samples were mixed thoroughly then centrifuged at 3000 xg at room temperature.

The water phase was transferred to a new 50 mL tube to which 7 mL of milli-Q H<sub>2</sub>O, 7  $\mu$ L of glycogen, 1.5 mL 2M NaAC pH 5.7 and 35 mL ethanol was added. Samples were placed at  $-80^{\circ}\text{C}$  overnight. The next day samples were centrifuged at  $4^{\circ}\text{C}$  for 30 min, 3000 xg. The supernatant was removed and 10 mL of cold 70% ethanol was added. Samples were centrifuged again for 15 min, 3000 xg at  $4^{\circ}\text{C}$ . The supernatant was removed and the pellet left to dry at room temperature. The pellet was dissolved in 150  $\mu$ L 10mM Tris pH 7.5. Each sample was transferred to a 1.7 mL tube, 50  $\mu$ L 10X restriction buffer and 50 units of the restriction enzyme Csp6I (Fermentas # ER0211) was added and the volume made up to 500  $\mu$ L with milli-Q H<sub>2</sub>O. After an overnight incubation, 500 RPM shaking, at  $37^{\circ}\text{C}$ , a 5  $\mu$ L aliquot of the sample was taken. This 'digestion control' was run on a 0.6% agarose gel. The enzyme was inactivated as previously describe and the samples transferred to a 50 mL tube. 1.4 mL of 10X ligation buffer and 20  $\mu$ L of ligase (100 U) (Roche Catalog # 10799009001) was added, then the reaction made up to 14ml with milli-Q H<sub>2</sub>O. After an overnight ligation at  $16^{\circ}\text{C}$ , 1.4ml 2M NaAC pH 5.6, 14  $\mu$ L glycogen and 35ml of 100% ethanol were added. Samples were stored at  $-80^{\circ}\text{C}$  overnight. The next day samples were centrifuged at  $4^{\circ}\text{C}$  for 45 minutes, at 3750 RPM. The supernatant was removed and 15 mL of cold 70% ethanol was added. The samples were then centrifuged again for 15 minutes, at  $20^{\circ}\text{C}$  and 3750 RPM. Again, the supernatant was removed and the pellet then left to dry at room temperature. Once dry the pellet was dissolved in 150  $\mu$ L 10mM Tris pH 7.5 at  $37^{\circ}\text{C}$ . Samples were then purified using a QIAquick PCR purification kit, according to the manufacturer's protocol. Samples were eluted in 50  $\mu$ L 10mM Tris pH 7.5 and pool samples. DNA concentration of each 4C template was determined via analysis with a NanoDrop 2000 (Thermo Scientific). Restriction fragments greater than 350 bp and within 2kb of the target genomic region were selected as viewpoint fragments, dependent on the ability to design specific primers. A 5' Illumina adaptor sequence was added so the inverse-PCR products did not need further processing prior to sequencing. Reading primers were designed as close to the primary restriction site as possible, to reduce reads from the known viewpoint sequence. Non-reading primers were designed to regions less than 120kb from the secondary restriction site. 200 ng of 4C template was used per PCR reaction. For each viewpoint and template, 16 PCR reactions were conducted using an Expand Long Template system (ROCHE # 11681834001) (see Table S2 for primer sequences). The pooled PCR products (total volume 800  $\mu$ L) were then purified using the High Pure PCR Product Purification Kit (Roche cat. no. 11732676001), to remove any adaptor containing primers (< 120 bp). Samples were centrifuged to pellet any beads that escaped the column. The supernatant was taken, then the concentration and purity of this 4C template was assessed by a NanoDrop 2000 (Thermo Scientific) (260/280 ratio >2 and 260/230 ratio >1.8 was required). The libraries were then visualized on a 1.5% agarose gel. All 4 of the 4C libraries were pooled, and then multiplexed sequencing was performed on the HiSeq 2500 platform. Individual fragment counts were calculated for every 1kb bin. A median was calculated, with a 3kb sliding window, and data from both biological replicates was merged. The R package DESeq2 was used to calculate the log<sub>2</sub> fold change (RUNX1/ETO knockdown versus control) at the local genomic coordinates. Viewpoint specific 4C-seq PCR primers used in this project are listed in Table S2.

### **Hi-C Library Generation**

Hi-C library generation was carried out as described previously (Mifsud et al., 2015; Lieberman-Aiden et al., 2009), with the following modifications which were detailed with the following modifications. After fixation in 2% formaldehyde for 5 min, 50 million Kasumi-1 cells were homogenized in 10 mL of ice-cold lysis buffer ten times on ice with a tight pestle, incubated on ice for 15 min and then homogenized a further ten times. After overnight digestion with HindIII at  $37^{\circ}\text{C}$ , DNA ends were labeled with biotin-14-dATP (Life Technologies) in a Klenow end-filling reaction. After phenol-chloroform purification, the DNA concentration was measured using Quant-iT PicoGreen (Life Technologies), and 40  $\mu$ g of DNA was sheared to an average size of 400 bp, using the manufacturer's instructions (Covaris). The sheared DNA was end repaired, adenine tailed, and double size selected using AMPure XP beads to isolate DNA ranging from 250 to 550 bp in size. Ligation fragments marked by biotin were immobilized using MyOne Streptavidin C1 DynaBeads (Invitrogen) and ligated to paired-end adaptors (Illumina). The immobilized Hi-C libraries were amplified using PE PCR 1.0 and PE PCR 2.0 primers (Illumina) with 8 PCR amplification cycles.

### **Biotinylated RNA Bait Library Design**

Biotinylated 120-mer RNA baits were designed to target both ends of HindIII restriction fragments that overlap Ensembl promoters of protein-coding, noncoding, antisense, snRNA, miRNA and snoRNA transcripts. A target sequence was valid if its GC content ranged between 25 and 65% and the sequence contained no more than two consecutive Ns and was within 330 bp of the HindIII restriction fragment terminus.

### **Promoter Capture Hi-C**

Capture HiC of promoters was carried out with SureSelect target enrichment, using the custom-designed biotinylated RNA bait library and custom paired-end blockers according to the manufacturer's instructions (Agilent Technologies). After library enrichment, a post-capture PCR amplification step was carried out using PE PCR 1.0 and PE PCR 2.0 primers with 4 PCR amplification cycles. Hi-C libraries were sequenced on the Illumina HiSeq 1000 platform.

### **Western Blotting**

Protein extracts were prepared using a co-immunoprecipitation kit (Active Motif, USA). Protein extracts were quantified using Bradford protein reagent (Bio-Rad, USA) and 595nM absorbance quantified by spectrophotometry. Absolute concentrations were determined using a standard curve from a known concentration of BSA (Pierce, USA). Protein extracts was run on an acrylamide gel and transferred to nitrocellulose membrane. The antibodies used in this project are listed in Table S2. Enhanced chemiluminescence by SuperSignal PICO (Thermos Scientific, USA) was used to develop the membrane. Chemiluminescence was detected using either developer or Chemidoc XRS system (BioRad, USA).

### Antibody Staining for Flow Cytometry

$15 \times 10^4$  were centrifuged at 300xg and washed with MACS buffer. The cell pellet was re-suspended in 50  $\mu$ l MACS buffer and 2  $\mu$ l of antibody was added and incubated for 15 minutes at 4°C in the dark. After incubation, the cells were washed once with MACS buffer before resuspension in 300  $\mu$ l MACS buffer and analyzed on Cyan ADP (Beckman Coulter, USA). Data were analyzed on Summit 4.3 (Beckman Coulter, USA). Antibodies used in this project are listed below. CD34 Monoclonal Antibody (4H11), PE, eBioscience Cat #12-0349-42; CD117 (c-Kit) Monoclonal Antibody (104D2), FITC, eBioscience Cat #11-1178-42

## QUANTIFICATION AND STATISTICAL ANALYSIS

### DNaseI Sequencing Data Analysis

DNaseI sequences from all experiments were mapped onto the reference human genome (hg38), with Bowtie version 2.3.1 (Langmead and Salzberg, 2012) using default parameters. Low quality reads were trimmed using Trimmomatic-0.36 (Bolger et al., 2014) and quality control (QC) statistics were obtained using FastQC tools. Unique aligned reads were used for downstream analysis. DNaseI Hypersensitive Sites (DHSs) were called with MACS2 using callpeak function (nomodel, call-summits and q = 0.005 parameters) (Zhang et al., 2008). Clustering of DNaseI-seq samples was carried out using the merged DHSs. The number of reads that mapped to these DHSs was counted in a 400bp window centered on the DHS summit, and subsequently normalized to total sample size using DESeq2 (Love et al., 2014). Pearson correlation coefficients were then calculated for each pair of samples using the log<sub>2</sub> of the normalized read counts, and then hierarchically clustered using Euclidean distance and complete linkage clustering of the correlation matrix in R.

### ChIP Sequencing Data Analysis

ChIP sequencing reads were aligned to the human genome version hg38 with Bowtie version 2.3.1 (Langmead and Salzberg, 2012). Reads that mapped uniquely to the genome were retained and duplicated reads were removed using the MarkDuplicates function in Picard tools (<http://broadinstitute.github.io/picard/>). Peaks were identified with MACS version 1.4.2 (Zhang et al., 2008) and DFilter software (Kumar et al., 2013) with recommended parameters (-bs = 100 -ks = 50 -refine). Peaks common to both peak calling methods were considered for further analysis.

### Average Tag Density Profile and Heatmap

The tag density and average profiles for Figures S1D and S2E were generated by calculating the tag density normalized as coverage per million within 400bp windows of the DNaseI peak summit. The read counts for all union peaks were computed. Coverages were calculated for all union peaks and ranked by log<sub>2</sub> fold change. Heatmap images were generated via Java TreeView (<http://jtreeview.sourceforge.net/>) and average profiles were plotted using R.

### RNA-Seq Data Analysis

RNA-seq reads downloaded from GSE54478 were aligned to the human genome hg38 build with STAR (Dobin et al., 2013) using ENCODE recommend parameters. Separate density profiles for the positive and negative strand were generated using bedtools. Cufflinks (Trapnell et al., 2013) was used to calculate the expression values as Fragments Per Kilobase per Million aligned reads (FPKM) from the aligned RNA-seq data and differentially expressed genes were extracted using the limma R package (Ritchie et al., 2015). All genes with p value  $\leq 0.01$  were considered with at least 2-fold changes between before and after RUNX1-ETO knock down.

### Promoter Capture HiC Data Analysis

The Chi-C paired-end sequencing reads were aligned to the human genome hg38 build using HiCUP pipeline (Wingett et al., 2015). The raw sequencing reads were initially separated and mapped against the reference genome. The reads were then filtered for experimental artifacts and duplicate reads, and then re-paired. Statistically significant interactions were called using GOTHiC package (Mifsud et al., 2017) and HOMER software (Heinz et al., 2010). This uses a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance, by using a background model of random interactions. This analysis assigns each interaction with a p value, which represents its significance. Differential interactions between control and after RUNX1-ETO knock-down were determined with HOMER (Heinz et al., 2010), p value cutoff with at least 0.01 was considered.

### 4C-Seq Data Analysis

4C-seq data analysis was performed using 4Cseqpipe, as described in van de Werken et al. (2012). Sequence extraction, mapping, normalization, and plotting of cis-contact profiles around PU.1 promoter and enhancer viewpoints were done using packages called by 4Cseqpipe tools. Custom restriction site tracks were built using the -build\_re\_db option of 4Cseqpipe for the hg19 human genomic version with HindIII as first and second restriction cutters. 4C reads were mapped to the custom hg19 tracks with the in-built 4Cseqpipe mapper. Near-cis domainograms were generated for PU.1 viewpoints using the median stat type and plotting the 20th and 80th quantile of the distribution of normalized contact intensities for 5kb sliding windows.

### Motif Identification and Clustering

De novo motif analysis was performed on peaks using HOMER (Heinz et al., 2010). The annotatePeaks function in HOMER was used to find occurrences of motifs in peaks. In this case we used known motif position weight matrices (PWM) from HOMER database.

Motif clustering: Digital footprinting of DNaseI high-depth sequencing data was performed using the Wellington algorithm (Piper et al., 2013) with FDR = 0.01. For the heatmap that shows hierarchical clustering of motif occurrences within specific and common footprints (Figure 3D). The distance between the centers of each motif pairs was calculated and the motif frequency was counted if the first motif was within 50bps distance from the second motif. Z-scores were calculated from the mean and standard deviation of motif frequencies observed in random sets using bootstrapping analysis. For bootstrapping, peak sets with a population equal to that of the footprinted peaks were randomly obtained from the union of DNase-Seq footprints.

### DATA AND CODE AVAILABILITY

The CHIP-seq data, DNaseI-seq data, and RNA-seq data generated during this study are available at GEO: GSE121282. The Capture HiC data generated during this study are available at GEO: GSE117108. The published article includes reprogramming ATAC-Seq (Corces et al., 2016) GEO: GSE75384 analyzed during this study. The published article includes Published ChIP-Seq (Ptasinska et al., 2014) GEO: GSE60131 analyzed during this study.