Super-Enhancers and Broad H3K4me3 Domains Form Complex Gene Regulatory Circuits Involving Chromatin Interactions

Fan Cao\textsuperscript{1*}, Yiwen Fang\textsuperscript{1*}, Hong Kee Tan\textsuperscript{1*}, Yufen Goh\textsuperscript{1}, Jocelyn Yeen Hui Choy\textsuperscript{1}, Bryan Koh\textsuperscript{2}, Jiong Hao Tan\textsuperscript{2}, Nicolas Bertin\textsuperscript{1,\#}, Aroul Ramadass\textsuperscript{3}, Ewan Hunter\textsuperscript{3}, Jayne Green\textsuperscript{3}, Matthew Salter\textsuperscript{1}, Alexandre Akoulitchev\textsuperscript{3}, Wilson Wang\textsuperscript{2}, Wee Joo Chng\textsuperscript{1,2,7}, Daniel G. Tenen\textsuperscript{1}, Melissa J. Fullwood\textsuperscript{1,4,5,6†}

\textsuperscript{1}Cancer Science Institute of Singapore, National University of Singapore.
\textsuperscript{2}National University Hospital, Singapore
\textsuperscript{3}Oxford BioDynamics Plc, Oxford, England.
\textsuperscript{4}School of Biological Sciences, Nanyang Technological University, Singapore.
\textsuperscript{5}Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore.
\textsuperscript{6}Yale-NUS Liberal Arts College, Singapore.
\textsuperscript{7}National University Cancer Institute, Singapore, National University Health System, Singapore

*These authors made equal and critical contributions
\#Current address: Human Longevity Singapore Pte. Ltd., Singapore
†Correspondence should be sent to:
Melissa J. Fullwood
Cancer Science Institute Singapore (CSI),
National University of Singapore (NUS),
Centre for Translational Medicine (MD6),
14 Medical Drive, #12-01,
Singapore 117599.
Contents:

1. Supplementary Text
2. Supplementary Table Legends
3. Supplementary Figure Legends
4. Supplementary Tables
5. Supplementary Figures
Supplementary Text

EpiSwitch™ validation in K562

The EpiSwitch™ method was tested in K562 cells to check its efficacy and whether it could validate ChIA-PET data from K562 cells. Two technical repeats were screened for each region and 0/2 indicates no interaction observed while 1/2 and 2/2 indicates 1 and 2 interactions observed respectively in each individual K562 sample. For bait-hit2, there were 2 different primers designed at slightly different loci, indicated as bait-hit2-1 and bait-hit2-2. While there are a few differences, the results are mostly similar between bait-hit2-1 and bait-hit2-2. Negative control region 1, which has no reported interaction, was used for this analysis.

Individual heterogeneity in patients

We note that we also detected examples of individual heterogeneity in chromatin interactions in patients. While certain individuals had similar interaction patterns, others had different patterns – for example, MYC Bait-Hit1 was seen in 9 out of 13 peripheral blood samples examined while MYC Bait-Hit2-1 was seen in 12 out of 13 peripheral blood samples examined (Figure S9a). Interestingly, the MYC Bait-Hit1 interaction was observed in the peripheral blood but not bone marrow samples of patients AD454 and AD548 (Figure S9a). This particular interaction was also detected in only 4 out of 10 CML patient samples in contrast to 7 out of 8 non-CML patient samples that were screened (Figure S9a).

Having observed heterogeneity between individuals in terms of chromatin interactions, we asked whether there was individual heterogeneity in enhancer activity levels. Due to limitations in the quantities of the patient samples, we were only able to examine two enhancers through the anti-H3K27ac ChIP-qPCR assay. We examined an enhancer at MYC-335, which is not present in K562 cells but is important in solid cancers, and the super-enhancer at miR1205 which loops over to the MYC promoter in K562 cells. We found that the enhancer activity levels vary in individual samples (Figure S9c). We did not see any clear correlation between enhancer activity levels and chromatin interaction levels (Table S5).

Next, we investigated whether there were any genetic or transcriptomic differences between the patient samples, and if so, what might be the impact of genetic differences between samples on chromatin interactions and enhancer activity levels. We performed PCR amplification followed by capillary sequencing to determine the genotype of Single Nucleotide Polymorphisms as well as whether there were any novel somatic mutations in the region and at two cancer mutation hotspot regions proximal of the miR1205 super-enhancer in patient samples (Table S4). Although several SNPs varied between samples and several SNPs occurred in CTCF binding regions as measured by ChIP-Seq, none showed any correlation with chromatin interaction levels or enhancer levels. It should be noted, however, that no SNPs were found at the CTCF motif. We did not detect any different SNPs or mutations in the two mutation hotspot regions proximal of the miR1205 super-enhancer. We did not detect by Sanger sequencing any novel somatic mutations in neither CML nor non-CML samples at the miR-1205 super-enhancer (Table S4). Taken together, we did not observe any clear correlations between different genetic status or the presence of novel DNA mutations and enhancer status or chromatin interaction levels or expression of MYC in the miR1205 region examined (Table S5) in these clinical samples.
Supplementary Table Legends

Supplementary Table 1. Enhancers, Chromatin Interactions and Associated Genes (Excel file). Sheet 1 & 3, “All enhancers” indicates the entire set of super-enhancers and typical enhancers and whether they associate with any chromatin interactions, as well as their associated genes. Sheet 2 & 4, “Enhancers of both TSG and OG” indicates the entire set of enhancers that associate with tumor suppressor genes and oncogenes, either by proximity or through chromatin interactions. Column A-C: genomic coordinates of the enhancer element (A: chromosome; B: start; C: end). D: name of the element. E: signal level. F: strand. G: type of enhancer element - whether proximal or distal; whether a typical enhancer (TE) or super-enhancer (SE). H: gene transcripts which overlap with the enhancer element. The GENCODE transcript is reported, with the common gene name reported in brackets. Multiple transcripts are separated by a semicolon. Oncogenes are indicated by <OG>, tumor suppressors are indicated by <TSG>, and census cancer genes are indicated by <CCG>. If there is no gene, a dot is indicated in the cell instead. I: the CAGE expression levels of the genes indicated by the transcripts in H are reported here. J: the RNA-Seq expression levels of the genes indicated by the transcripts in H are reported here. K: number of interactions associated with the enhancer element. L: number of interacting distal super-enhancers (DSE). M: number of interacting proximal super-enhancers (PSE). N: number of interacting distal typical enhancers (DTE). O: number of interacting proximal typical enhancers (PTE). P: genes at proximal super-enhancers (PSE) that interact with this enhancer element. Q: the CAGE expression levels of the genes indicated by the transcripts in P are reported here. R: the RNA-Seq expression levels of the genes indicated by the transcripts in P are reported here. S: Genes at proximal typical enhancers (PTE) that interact with this enhancer element. T: the CAGE expression levels of the genes indicated by the transcripts in S are reported here. U: the RNA-Seq expression levels of the genes indicated by the transcripts in S are reported here. V: The maximum specificity score of the CAGE clusters at each transcript.

Supplementary Table 2. Listing of tumor suppressor genes, oncogenes and census cancer genes associated with super-enhancers and broad H3k4me3 domains (Excel file). This file lists all the tumor suppressor genes, oncogenes, and COSMIC census cancer genes that are targeted by super-enhancers or broad H3k4me3 domains by proximity or looping in K562 (Sheet 1) and MCF-7 (Sheet 2). Row 1 describes whether the genes are targeted by proximal super-enhancers (PSE)/broad H3k4me3 domains (PBD) or distal super-enhancers (DSE)/broad H3k4me3 domains (DBD) and whether they are targeted by proximity or by chromatin interactions.

Supplementary Table 3. H3k4me3 domains, Chromatin Interactions and Associated Genes (Excel file). Sheet 1 and 2 are data for K562 and MCF-7, respectively. Column A-C: genomic coordinates of the enhancer element (A: chromosome; B: start; C: end). D: name of the element. E: signal level. F: strand. G: type of enhancer element - whether proximal or distal; whether a typical domain (TD) or broad domain (BD). H: gene transcripts which overlap with the enhancer element. The GENCODE transcript is reported, with the common gene name reported in brackets. Multiple transcripts are separated by a semicolon. Oncogenes are indicated by <OG>, tumor suppressors are indicated by <TSG>, and census cancer genes are indicated by <CCG>. If there is no gene, a dot is indicated in the cell instead. I: the CAGE expression levels of the genes indicated by the transcripts in H are reported here. J: the RNA-Seq expression levels of the genes indicated by the transcripts in H are reported here. K: number of interactions associated with the enhancer element. L: number of interacting distal broad domains (DBD).
M: number of interacting proximal broad domains (PBD). N: number of interacting distal typical domains (DTD). O: number of interacting proximal typical domains (PTD). P: genes at proximal broad domains (PBD) that interact with this enhancer element. Q: the CAGE expression levels of the genes indicated by the transcripts in P are reported here. R: the RNA-Seq expression levels of the genes indicated by the transcripts in P are reported here. S: Genes at proximal typical domains (PTD) that interact with this enhancer element. T: the CAGE expression levels of the genes indicated by the transcripts in S are reported here. U: the RNA-Seq expression levels of the genes indicated by the transcripts in S are reported here. V: The maximum specificity score of the CAGE clusters at each transcript.

**Supplementary Table 4.** Listing of ChIP-Seq libraries used in the study.

**Supplementary Table 5.** Listing of patient samples examined (in this document). These samples include both peripheral blood and bone marrow from patients with chronic myelogenous leukemia as well as other diseases, and include knee bone aspirates from normal individuals.

**Supplementary Table 6.** Genetic variations in clinical samples (in this document). 6 SNPs were found in miR1205 region. SNPs that fall within the CTCF regions are annotated with *. A>C indicates a change of base from A to C while A/C indicates two peaks identified from sequencing. No mutations and SNPs were found in the two DNA hotspot regions that were sequenced. The genomic location for the miR1205 sequenced region is chr8:128978404-128981253 while chr8:128,973,357-128,973,382 and chr8:129,066,969-129,067,005 are the genomic locations for DNA hotspot region 1 and 2 from Weinhold et al. respectively ¹.

**Supplementary Table 7.** Cross-comparison between chromatin interactions, enhancers, gene expression data and sequencing at the miR1205 super-enhancer region (in this document). Enhancer activity is indicated by ChIP-qPCR of H3K27ac (Figure S9C). N.D. indicates “not done”, due to low cell numbers in the clinical samples. Chromatin interactions are indicated as present (Yes) or absent (No) on the basis that 2 or 3 of the 3 tested interactions must be positive in order to conclude that the interaction is present. The source data is shown in Figure S9A. The MYC expression data is indicated by RT-qPCR (Figure S9B). N.D. indicates “not done”, due to low cell numbers in the clinical samples. The source data of genetic signatures is from Table S6.

**Supplementary Table 8.** Listing of primers used in experiments (in this document). This is a listing of all primers used for ChIP-qPCR, RT-qPCR, DNA sequencing, and EpiSwitch™.
Supplementary Figure Legends

**Figure S1.** Schematic of analyses performed in this paper. The input data are indicated in light green, and the analyses performed are shown in light beige.

**Figure S2.** Characterizing proximal and distal super-enhancers. a. The signal profile of the super-enhancer calling and some oncogenes (red), tumor suppressor genes (blue), and census cancer genes (underlined) they target by proximity or by looping. b-c. Screenshots of super-enhancers at FOXA1, GATA3 and ESR1, CDKN1B. d. Fraction of proximal elements found at leukaemia associated genes. e. Histone modifications (H3K27ac, H3K4me1, H3K4me3) at the four types of regulatory elements, including proximal super-enhancers (PSE), proximal typical enhancers (PTE), distal super-enhancers (DSE) and distal typical enhancers (DTE). f-g. The specificities (f) and expression levels (g) of CAGE clusters at proximal super-enhancers and proximal typical-enhancers. All boxplots presented were prepared in the following manner: the black horizontal line indicates the median, the top and bottom of the box indicates the third and first quartile respectively, and the whiskers indicate 1.5*the interquartile range. Widths of boxes are in proportion to the square root of the number of data points in each category and the statistics testing was done using Dunn’s Test. h-i. The cell-type specificities and expressions of enhancer RNAs at distal super-enhancers (DSE) and distal typical enhancers (DTE). j. The number of distal enhancers, including distal super-enhancers (DSE) and distal typical enhancers (DTE) associated with enhancer transcription. TPM indicates tags per million sequences.

**Figure S3.** The overlap of enhancers with chromatin interactions and their effects on the transcription of the remote target genes a. Fractions of the four types of elements associated with at least one chromatin interaction. b. Boxplot for the number of interactions each type of element has (only interacting elements were included). c. Distribution of distances of the nearest TSS (blue), nearest active TSS (green), and TSSes through chromatin interactions (red) to the center of distal super-enhancers. d-e. The expression levels (d) and cell-type specificity scores (e) of CAGE clusters at proximal elements that are connected to the four types of elements as indicated on the x-axis. “None” indicates the set of CAGE clusters located at non-interacting proximal elements. f The expression levels of tumor suppressor genes (tsg), oncogenes (og), and census cancer genes (ccg) targeted by broad domains through proximity (pbd_p) and looping (pbd_d and dse_d) measured by RNA-Seq. Data shown is for MCF-7.

**Figure S4.** Analysis of broad H3K4me3 peaks and chromatin interactions. a. Rank of H3k4me3 peaks by size and the cancer associated genes targeted by some broad domains. b. Some histone modifications at the four types of H3k4me3 domains. c. The fraction of proximal broad and typical domains located near leukemia associated genes. p-value is produced Fisher’s Exact Test. d. The fraction of each type of elements by H3k4me3 involved in chromatin interactions. e-f. The cell-type specificity scores (e) and expression levels (f) of CAGE clusters at proximal H3k4me3 elements that are connected to the four types of H3k4me3 elements as indicated on the x-axis. “None” indicates the set of CAGE clusters located at non-interacting proximal elements. g. The expression levels of tumor suppressor genes (tsg), oncogenes (og), and census cancer genes (ccg) targeted by broad domains through proximity (pbd_p) and looping (pbd_d and dbd_d) measured by RNA-Seq. Data shown is for MCF-7.
Figure S5. Normalization of the association between enhancers and chromatin interactions. a-b. Fraction of extended elements with chromatin interactions (a) and boxplot of number of chromatin interactions each type of elements has (b). c. Boxplot of number of chromatin interaction the four types of enhancer have after normalization against the total Pol2 signal present in the ChIA-PET dataset. d-e. Fraction of elements with chromatin interactions from HiC data (d) and boxplot of number of chromatin interactions from HiC data each type of elements has (e). f-g. Fraction of extended elements with chromatin interactions from HiC data (f) and boxplot of number of chromatin interactions from HiC data each type of elements has (g). The data shown is based on K562 cells.

Figure S6. Comparison of expressions (a) and specificities (b) of CAGE clusters at proximal super-enhancers/proximal typical enhancers involved in chromatin interactions and CAGE clusters at those not involved in chromatin interactions in K562 cells.

Figure S7. Overlap between super-enhancers and broad H3k4me3 domains in MCF-7 cells. a. Histone modifications at proximal regions with both super-enhancers and broad H3k4me3 domains (P\_SE\_BD), only super-enhancers (P\_SE\_O), only broad H3k4me3 domains (P\_BD\_O), and neither (P\_O\_O). b. Histone modifications at distal regions with both super-enhancers and broad H3k4me3 domains (D\_SE\_BD), super-enhancers only (D\_SE\_O), broad H3k4me3 domains only (D\_BD\_O), and neither (D\_O\_O). c. The fraction of different regions involved in chromatin interactions. d. Boxplot of number of chromatin interactions each type of regions is involved in. e. Pairwise comparison of enrichment at oncogenes (OG), tumor suppressor genes (TSG), COSMIC census cancer genes (CCG), and housekeeping genes (HKG) between different distal regions through looping (left), and different proximal regions through looping (middle) and proximity (right). Comparison was performed using Fisher’s Exact Test followed by Holm-Sidak correction for multiple testing. The color indicates the product of negative log-transformed p-values and sign of the log-transformed corresponding odds ratio.

Figure S8. Characterization of chromatin interactions in human samples by EpiSwitch™. a. EpiSwitch™ tested regions were selected from analyzing K562 RNA Polymerase II ChIA-PET and super enhancer locations published in Hsniz et al. The TP53 Bait region is located near the TP53 promoter. The TP53 Hit 1 region is located near the MPDU1 gene promoter, and spans over CD68. The TP53 Hit 2 region is located near the KDM6B gene promoter. All regions include proximal super-enhancers and proximal broad domains. The MYC Bait region is located at a proximal typical-enhancer, a proximal broad domain, and near the MYC promoter. The MYC Hit 1 region is located at a distal typical-enhancer; in a PVT1 intron, and near miR-1205. MYC Hit 2 region is located at a distal super-enhancer, a distal typical domain; and in a PVT1 intron. b. EpiSwitch™ results at the MYC locus. Results are shown at different input levels of DNA for semi-quantitative measurements, and with primers designed at each locus. Certain loci have two or more primer pairs used which indicate different locations within each locus. Results are shown from peripheral blood and bone marrow from Chronic Myelogenous Leukemia patients and control patients respectively. “1” indicates the interaction was detected, while “0” indicates the interaction was not detected. c. EpiSwitch™ results at the TP53 locus. Results are shown at different input levels of DNA for semi-quantitative measurements, and with primers designed at each locus. Certain loci have two or more primer pairs used which indicate different locations within each locus. Results are shown from peripheral blood and bone marrow from Chronic Myelogenous Leukemia.
patients and control patients respectively. “1” indicates the interaction was detected, while “0” indicates the interaction was not detected.

**Figure S9.** Characterization of chromatin interactions at **MYC** locus in human samples by EpiSwitch™, RT-qPCR and ChIP-qPCR. **a.** The interactions at the **MYC** locus were screened in triplicate and at multiple using 8 CML peripheral blood samples and 6 other patients (non-CML) peripheral blood. Two CML and two non-CML bone marrow samples were run as a pilot test of any possible differences in bone marrow (AD454 and AD548). “PB” indicates peripheral blood and “BM” indicates bone marrow. Three technical repeats were screened for each region and the number of interactions detected were indicated in numerical values out of the total three replicates for each of the individual patient samples. 2 or 3 of the 3 tested interactions must be positive to conclude that the interaction is detected. “*” indicates patient in blast crisis. It should be noted that this set of data was obtained in earlier experiments, before different input levels of DNA were used for semi-quantitative results. **b.** RT-qPCR comparison of **MYC** mRNA expression levels between CML and normal cells. **MYC** mRNA was quantified by qPCR and normalized against the expression of the β-actin housekeeping gene. The fold difference in **MYC** mRNA expression was then calculated relative to K562. Error bars indicate the mean + standard deviation. **c.** ChIP-qPCR of H3K27ac at **MYC-335**, a super-enhancer near miR-1205, **MYC** promoter and a negative control region. IgG showed low enrichment levels comparable to the negative control (results not shown). Error bars indicate the mean + standard deviation.

**Figure S10.** Analysis of regions with different super-enhancer and broad H3k4me3 domain coverage and their effects on transcription of target genes in K562. **a-b.** Counts of overlaps between H3k4me3 domains and H3k27ac enhancers. **c-d.** Specificities and expressions of CAGE clusters at different types proximal regions. **e-h.** The cell-type specificity scores (e, g) and expression levels (f, h) of CAGE clusters at proximal elements that are connected to regulatory regions as indicated on the x-axis. “None” indicates the set of CAGE clusters located at non-interacting proximal elements.

**Figure S11.** K562 RNA-Seq expression levels of oncogenes (OG), tumor suppressor genes (TSG), and census cancer genes (CCG) regulated by **a.** different types of distal regulatory regions through looping; **b.** different types of proximal regulatory regions through looping. **c.** different types of proximal regulatory regions by proximity.

**Figure S12.** Analysis of regions with different super-enhancer and broad H3k4me3 domain coverage and their effects on transcription of target genes in MCF-7. **a-b.** Counts of overlaps between H3k4me3 domains and H3k27ac enhancers. **c-d.** Specificities and expressions of CAGE clusters at different types proximal regions. **e-h.** The cell-type specificity scores (e, g) and expression levels (f, h) of CAGE clusters at proximal elements that are connected to regulatory regions as indicated on the x-axis. “None” indicates the set of CAGE clusters located at non-interacting proximal elements.

**Figure S13.** MCF-7 RNA-Seq expression levels of oncogenes (OG), tumor suppressor genes (TSG), and census cancer genes (CCG) regulated by **a.** different types of distal regulatory regions through looping; **b.** different types of proximal regulatory regions through looping. **c.** different types of proximal regulatory regions by proximity.
Figure S14. Enhancer sizes with enhancers ordered by the H3k27ac signals for K562 (a) and MCF-7 (b). A smooth window of 20 was applied. The vertical dotted line indicates the split between typical enhancers and super-enhancers.
Supplementary Tables

| Sample ID | Sample type | Race | Gender | Age | Drug treatment | Description provided by clinic |
|-----------|-------------|------|--------|-----|----------------|-------------------------------|
| AD548     | PB/BM       | Asian| Female | 40  | Glivec/Hydroxyurea | CML – chronic                 |
| AD454     | PB/BM       | Asian| Male   | 38  | Hydroxyurea     | CML – accelerated             |
| AD35      | PB          | Asian| Male   | 32  | Hydroxyurea and IFN-alpha Glivec/Anagrelide | CML with blast transformation |
| AD35a     | PB          | Asian| Male   | 32  | DAUNO/ARA-C 3 + 7 Glivec/Ara-C Anagrelide/Hydroxyurea | CML with blast transformation |
| AD96      | PB          | Asian| Female | 34  | Hydroxyurea/Glivec | CML – chronic                 |
| AD256     | PB          | Non-Asian| Male | 43  | Glivec         | CML                           |
| AD294     | PB          | N/A  | N/A    | N/A | N/A            | N/A                           |
| AD301a    | PB          | Asian| Female | 41  | Glivec         | CML – chronic                 |
| AD035     | PB          | N/A  | N/A    | N/A | N/A            | N/A                           |
| AD096     | PB          | Asian| Female | 34  | Imatini/Nilotinib | CML – chronic                 |
| AD256     | PB          | N/A  | N/A    | N/A | N/A            | N/A                           |
| AD264     | PB          | Asian| Male   | 25  | Imatinib       | CML – chronic                 |
| AD277     | PB          | Asian| Male   | 29  | Imatinib/Nilotinib/Dasatinib | CML – chronic |
| AD290a    | PB          | Asian| Female | 57  | Imatinib       | CML – chronic                 |
| AD315     | PB          | Asian| Male   | 64  | Imatinib       | CML – chronic                 |
| AD60      | BM          | Asian| Male   | 45  | Hydroxyurea/Glivec/HSC | CML – chronic                 |
| AD61      | BM          | N/A  | N/A    | N/A | N/A            | N/A                           |
| AD166     | BM          | N/A  | N/A    | N/A | N/A            | N/A                           |
| AD171     | BM          | N/A  | N/A    | N/A | N/A            | N/A                           |
| AD173     | BM          | Asian| Female | 21  | Hydroxyurea/Glivec/HSC | CML – chronic                 |
| AD202     | BM          | Asian| Male   | 19  | Nilotinib      | CML – chronic                 |
| AD206     | BM          | Asian| Female | 33  | Nilotinib      | CML – chronic                 |
| AD212     | BM          | Asian| Female | 38  | Glivec/Nilotinib | CML – chronic                 |
| AD241     | BM          | Asian| Male   | 31  | Nilotinib      | CML – chronic                 |

Non-Chronic Myelogenous Leukemia (non-CML) samples

| Sample ID | Sample type | Race | Gender | Age | Drug treatment | Description provided by clinic |
|-----------|-------------|------|--------|-----|----------------|-------------------------------|
| AD278     | PB          | Asian| Male   | 42  | N/A            | Waldenstrom Macroglogulinemia |
| AD283     | PB          | Asian| Male   | 60  | N/A            | Splenic Marginal zone lymphomas |
| AD289     | BM          | Asian| Female | 31  | N/A            | Reactive marrow               |
| Sample ID | Type  | Age | Gender | Disease  |
|-----------|-------|-----|--------|----------|
| AD339     | BM    | 60  | Male   | Clonal small B cell lymphoproliferative disease Mantle cell lymphoma excluded possibly atypical CLL. |
| AD394     | PB    | 55  | Male   | No conclusive evidence of lymphoproliferative disease. |
| AD412     | PB    | 53  | Female | End stage renal failure |
| AD419     | PB    | N/A | N/A    | N/A |
| AD446     | PB    | 33  | Female | Hodgkin Lymphoma |
| AD272     | PB    | 56  | Male   | CEL with FIP1L1-PDGFRa positive |
| AD276     | PB    | 61  | Female | Hodgkin's lymphoma (IIIb) |
| AD289     | PB    | 32  | Female | Reactive marrow |
| AD316     | PB    | 69  | Male   | Stage 4B T cell lymphoma |
| AD334     | PB    | 44  | Female | DLBCL |
| AD346     | PB    | 31  | Female | Hodgkin's disease |
| Knee 8    | BM    | 71  | Male   | N/A |
| Knee 9    | BM    | 77  | Female | N/A |
| Knee 10   | BM    | 65  | Female | N/A |
| Knee 11   | BM    | 53  | Female | N/A |
| Knee 12   | BM    | 70  | Female | N/A |
| Knee 13   | BM    | 84  | Male   | N/A |
| Knee 14   | BM    | 67  | Female | N/A |
| Knee 15   | BM    | 68  | Male   | N/A |

*PB indicates Peripheral Blood; BM indicates Bone Marrow

| Patient Type | Sample ID | SNPs ID in miR1205 |
|--------------|-----------|--------------------|
|              | rs755519  | rs752235 | rs890443 | rs752427* | rs752429* | rs4733584* |
| CML          | AD256PB   | A/G       | A/C      | T>C      | C/T      | T/C       |
|              | AD454PB   | A>G       | A>C      | T>C      | C>T      | T>C       |
|              | AD301aPB  | A/G       | A/C      | T/C      | C/T      | T/C       |
| Sample ID  | Enhancer Activity | Chromatin Interaction | MYC Gene Expression Level | Genetic Signatures                      |
|------------|-------------------|-----------------------|---------------------------|-----------------------------------------|
| CML        |                   |                       |                           |                                          |
| AD35 PB    | 17.13             | No                    | 0.37                      | rs755519 A>G                            |
|            |                   |                       |                           | rs752235 A>C                            |
|            |                   |                       |                           | rs890443 T>C                            |
|            |                   |                       |                           | rs752427 C>T                            |
|            |                   |                       |                           | rs752429 T>C                            |
|            |                   |                       |                           | rs4733584 T>C                            |
| AD35a PB   | 3.53              | No                    | 0.09                      | rs755519 A>G                            |
|            |                   |                       |                           | rs752235 A>C                            |
|            |                   |                       |                           | rs890443 T>C                            |
|            |                   |                       |                           | rs752427 C>T                            |
|            |                   |                       |                           | rs752429 T>C                            |
|            |                   |                       |                           | rs4733584 T>C                            |
| AD96 PB    | 2.98              | Yes                   | N.D.                      | rs755519 A/G                            |
|            |                   |                       |                           | rs752235 A/C                            |
|            |                   |                       |                           | rs890443 T/C                            |
|            |                   |                       |                           | rs752427 C/T                            |
|            |                   |                       |                           | rs752429 T/C                            |
|            |                   |                       |                           | rs4733584 T/C                            |
| AD256 PB   | N.D.              | No                    | N.D.                      | rs755519 A/G                            |
|            |                   |                       |                           | rs752235 A/C                            |
|            |                   |                       |                           | rs890443 T>C                            |

Table S6. Single nucleotide polymorphisms detected in the miR1205 super-enhancer region.
|     |     |     |                  | rs752427 C/T | rs752429 T/C | rs4733584 T>C |
|-----|-----|-----|------------------|-------------|-------------|---------------|
| AD294 PB | N.D. | Yes | N.D.            | rs755519 A>G | rs752235 A>C | rs890443 T>C |
|       |     |     |                  | rs752427 C>T | rs752429 T>C | rs4733584 T>C |
| AD301a PB | N.D. | No  | N.D.            | rs755519 A/G | rs752235 A>C | rs890443 T>C |
|       |     |     |                  | rs752427 C>T | rs752429 T>C | rs4733584 T>C |
| AD454 PB | N.D. | Yes | N.D.            | rs755519 A>G | rs752235 A>C | rs890443 T>C |
|       |     |     |                  | rs752427 C>T | rs752429 T>C | rs4733584 T>C |
| AD548 PB | 5.04 | Yes | N.D.            | rs755519 A>G | rs752235 A>C | rs890443 T>C |
|       |     |     |                  | rs752427 C>T | rs752429 T>C | rs4733584 T>C |
| AD454 BM | 0.90 | No  | N.D.            | rs755519 A>G | rs752235 A>C | rs890443 T>C |
|       |     |     |                  | rs752427 C>T | rs752429 T>C | rs4733584 T>C |
| AD548 BM | 3.85 | No  | 1.18            | rs755519 A>G | rs752235 A>C | rs890443 T>C |
|       |     |     |                  | rs752427 C>T | rs752429 T>C | rs4733584 T>C |
| Non-CML |     |     |                  | rs755519 A>G | rs752235 A>C | rs890443 T>C |
| AD278 PB | N.D. | Yes | N.D.            | rs755519 A/G | rs752235 A/C | rs890443 T>C |
|       |     |     |                  | rs752427 C/T | rs752429 T/C | rs4733584 T>C |
| AD283 PB | 2.64 | Yes | 0.33            | rs755519 A/G | rs752235 A/C | rs890443 T>C |
|       |     |     |                  | rs752427 C/T | rs752429 T/C | rs4733584 T>C |
| AD419 PB | N.D. | Yes | N.D.            | rs755519 A/G | rs752235 A/C | rs890443 T>C |
|       |     |     |                  | rs752427 C/T | rs752429 T/C | rs4733584 T>C |
| Sample | Treatment | Expression | p-value | Alleles |
|--------|-----------|------------|---------|---------|
| AD446  | PB        | N.D.       | Yes     | N.D.    |
| AD394  | PB        | N.D.       | No      | N.D.    |
| AD412  | PB        | 10.0       | Yes     | N.D.    | rs755519 A/G
|         |           |            |         |         | rs752235 A/C
|         |           |            |         |         | rs890443 T>C
|         |           |            |         |         | rs752427 C/T
|         |           |            |         |         | rs752429 T/C
|         |           |            |         |         | rs4733584 T>C
| AD289  | BM        | N.D.       | Yes     | 0.03    | rs755519 A>G
|         |           |            |         |         | rs752235 A>C
|         |           |            |         |         | rs890443 T>C
|         |           |            |         |         | rs752427 C>T
|         |           |            |         |         | rs752429 T>C
|         |           |            |         |         | rs4733584 T>C
| AD339  | BM        | 7.68       | Yes     | N.D.    | rs755519 A>G
|         |           |            |         |         | rs752235 A>C
|         |           |            |         |         | rs890443 T>C
|         |           |            |         |         | rs752427 C>T
|         |           |            |         |         | rs752429 T>C
|         |           |            |         |         | rs4733584 T>C

**Table S7.** Cross-comparison between chromatin interactions, enhancers, gene expression data and sequencing.
| Primer name                        | Sequence (5’ to 3’)              |
|-----------------------------------|----------------------------------|
| **ChIP-qPCR primers**             |                                  |
| MYC promoter Forward              | AGAAGGGCAGGGCTTTCTC              |
| MYC promoter Reverse              | CCGAAAACCGGCTTTTATAC             |
| MYC-335 Forward                   | CAAAAGGCCAGGAACCCAGCA            |
| MYC-335 Reverse                   | GGCATGCACAAACCTTCCC             |
| miR-1205 superenhancer Forward    | GCTCCTGCAAAAGACGGA              |
| miR-1205 superenhancer Reverse    | AAGGGAGCCTGTGTGTGTGCCT           |
| Negative control Forward          | GGCCAGAATCAGATTTTCCA            |
| Negative Control Reverse          | GCCTTAACGTGACGGCTGT              |
| **RT-qPCR primers**               |                                  |
| MYC-Forward                       | TTCGGGTAGTGGAAAACCAG            |
| MYC-Reverse                       | CAGCAGCTCGAATTTTCTCC            |
| β-actin-Forward                   | ACCCTGAAGTACCCCATCGA            |
| β-actin-Reverse                   | CTCAAAACATGATCTGGGTCATCT        |
| **Genomic DNA amplification and Capillary Sequencing primers** | |
| PVT1-1 Forward                    | TGCTGTTGGATCAATGTAGC            |
| PVT1-1 Reverse                    | CACTGGCTACATAGGGAGGG            |
| PVT1-2 Forward                    | GTGGGGCTGTGGAAAACCAG            |
| PVT1-2 Reverse                    | AGAAAGGCCCTGCTTTCTCT            |
| PVT1-3 Forward                    | CGCTGATGACAAAGGGGACAG           |
| PVT1-3 Reverse                    | GGACCACATCGGTCATCTTC            |
| PVT1-4 Forward                    | GCCCAAAAGACGGAAAAAGGA           |
| PVT1-4 Reverse                    | AACAGACATCCCCAGACCTG           |
| PVT1-5 Forward                    | AACTGGCTGGCTTCTACT              |
| PVT1-5 Reverse                    | TCCCTTGACTTTGCTGCTGA            |
| PVT1-6 Forward                    | TCTTGGGATAGGTGAGCTC             |
| PVT1-6 Reverse                    | TCAGGGTAAATGTGGTGCA             |
| Hotspot 1 Forward                 | GGCACGAATAGGGGAGGAAT            |
| Hotspot 1 Reverse                 | GGACCACATCTGCTGATCTTG          |
| Hotspot 2 Forward                 | CCACATGATTTGGGCACTCC            |
| Hotspot 2 Reverse                 | AGTGACATCTGGGACTGTTG           |
**Table S8a.** Primer sequences used for ChIP-qPCR, RT-qPCR, Sanger Sequencing and 4C.

| Primer Name | Sequence | Description |
|-------------|----------|-------------|
| FULL2-19    | CTCTCTCGCTAATCTCCGCC | Bait Forward Inner |
| FULL2-20    | TCATGCGGCTCTCTTACTCTG | Bait Forward Outer |
| FULL2-23    | ACTGGATCGGGGTAAAGTGAC | Bait Forward Inner |
| FULL2-24    | GCTGAGATGAGTCGAATGCC | Bait Forward Outer |
| FULL2-25    | TGTCCCTTTGTGTCAGCGG | Hit 1 Reverse Inner |
| FULL2-26    | AACACACACTTGTGTCAGGG | Hit 1 Reverse Outer |
| FULL2-29    | CGATGTCTTGAAGGACACAGTG | Hit 2 Reverse Inner |
| FULL2-30    | AGCTCCGTCAGTGATCAGTC | Hit 2 Reverse Outer |
| FULL2-33    | GCATTCTCTGGCTCTCTCTG | Hit 2 Reverse Inner |
| FULL2-34    | GCATTCTCTGGCTCTCTGCGAC | Hit 2 Reverse Outer |
| FULL2-37    | CGTACAGAGCTGCTGTTGGG | Control Reverse Inner |
| FULL2-38    | TGCTACCTATTTTGTGCACACC | Control Reverse Outer |
| FULL2-3    | CGTGCTAAGGAACACAGTG | Bait Forward Inner |
| FULL2-4    | CGTGCTAAGGAACACAGTG | Bait Forward Outer |
| FULL2-9    | CAGGGAGGCTCATAACCTTG | Hit 2 Reverse Inner (Fig S7) |
| FULL2-10   | GGGTACTGAGTGTCAGTGAGG | Hit 2 Reverse Outer (Fig S7) |
| FULL2-13   | GCTTGGTGGACGTTCAGAG | Hit 1 Reverse Inner |
| FULL2-14   | CTGTCTGACGTACCCACCC | Hit 1 Reverse Outer |
| Full2-41   | TTATCCCTCTGGCCAGGTTC | Control Reverse Inner (Fig S7) |
| Full2-42   | GTTTACAGGTCCCATGTTG | Control Reverse Outer (Fig S7) |
| Primer ID | Sequence               | Function                                  |
|-----------|------------------------|-------------------------------------------|
| FULL2-5   | TCCGGGTTGATTCTTTGCC    | Control Reverse Inner (Fig S8)            |
| FULL2-6   | ACCCATGAGCATCAGTCACC   | Control Reverse Outer (Fig S8)            |
| FULL2-11  | AAGAGGGATCACCGGTTTGG   | Hit 2 Forward Inner (Fig S8)              |
| FULL2-12  | AGAGGGTGGGGTGGACTGTC   | Hit 2 Forward Outer (Fig S8)              |

**Table S8b.** Primer sequences used for EpiSwitch™.
Supplementary Figures
Identify super-enhancers/broad H3k4me3 domains from H3K27ac/H3k4me3 ChIP-Seq data (ENCODE) (Hnisz et al., Cell, 2013) (Chen et al., Nat Genet., 2015) (Fig. 1a-d, Fig. 4a-c, Table S1)

Classify super-enhancers/broad H3k4me3 domains into proximal and distal (Table S1)

Identify high-confidence chromatin interactions from RNA Polymerase II ChIA-PET data (Li et al., Cell, 2012)

Identify enhancers/H3k4me3 domains associated with chromatin interactions (Fig. 3a-b, Fig. 4d, Table S1)

Gene list (Gencode)

List of tumor suppressors and oncogenes (Davoli et al., Cell, 2013)

Role of super-enhancer/broad H3k4me3 loops in gene expression and gene cell specificity levels (Fig. 3d-e, Fig. 4e-f, Fig. S4d-e, Fig S5e-f, Fig. S6)

Investigation of chromatin interactions in clinical samples by Episwitch, enhancer levels by H3K27ac ChIP-qPCR, gene expression levels by RT-qPCR and DNA mutation status by Sanger DNA sequencing (Fig. S9)

Roles of regions with different super-enhancer and broad H3k4me3 contents in gene expression and gene cell specificity by looping or by proximity. (Fig S10-13)

Validation of super-enhancers that loop to oncogenes and tumor suppressors in K562 cells by Oxford Biodynamics Episwitch™ (Fig. 6b-c, Fig. S8)

Identification of enhancers/H3k4me3 domains associated with chromatin interactions. (Fig. 3a-b, Fig. 4d, Table S1)

Overlap between super-enhancers and broad H3K4me3 peaks (Fig 5, Fig S7, Fig. S10a-b, Fig S12a-b)

CAGE data: Gene expression and cell specificity levels and enhancer RNA levels (Forrest et al., Nature, 2014; Andersson et al., Nature, 2015))

Analysis of regions with different super-enhancer and broad H3k4me3 contents on their enrichment at oncogenes, tumor suppressors. (Fig 5c, Fig S7c)

Examine properties of proximal and distal super-enhancers (Fig. 1e-k)

Identify genes, tumor suppressors and oncogenes regulated by super-enhancer/broad H3k4me3-associated chromatin interactions (Fig. 3f, Fig. 4f, Table S1-2)

Investigation of chromatin interactions in clinical samples by Episwitch, enhancer levels by H3K27ac ChIP-qPCR, gene expression levels by RT-qPCR and DNA mutation status by Sanger DNA sequencing (Fig. S9)

Roles of regions with different super-enhancer and broad H3k4me3 contents in gene expression and gene cell specificity by looping or by proximity. (Fig S10-13)
Figure S2

By Proximity

By Looping

Enhancers ranked by H3K27ac signal

C

H3k27ac

H3k4me1

H3k4me3

Average

Median

Average

Median

Average

Median

PSE (155)

PTE (8010)

DSE (142)

DTE (6685)

Expression

(155)

(8010)

(142)

(6685)

Expression (TPM)

PSE (155)

PTE (8010)

DSE (142)

DTE (6685)

Fraction transcribed

DSE (142)

DTE (6685)

Element type
Figure S3
Figure S4
Figure S5
Figure S6

(a) Box plots showing expression (TPM) for different conditions.

(b) Box plots showing specificity scores for different conditions.
Figure S7
### Table: CML BM (%)

| Sample detection (%) | CML BM (%) | Control BM (%) | CML PB (%) | Control PB (%) |
|----------------------|------------|----------------|------------|----------------|
| Total CML (%)        |            |                |            |                |
| Total Control (%)    |            |                |            |                |

### Figure S8a-b

**a**

Proximal super-enhancer, proximal broad domain: near the KDM6B gene promoter; spanning over CD68

**b**

Proximal super-enhancer, proximal broad domain: near the MPP1 gene promoter; spanning over CD68

**Table**

| C-MYC Hit 1 | C-MYC Hit 2 | C-MYC Bait (1) | C-MYC Bait (2) |
|-------------|-------------|----------------|----------------|
| Distal super-enhancer, near the MR1205 intron | Distal typical-enhancer, near the P117 intron | AD294 | AD316 |
| Proximal super-enhancer, near the MPP1 gene promoter | Proximal super-enhancer, near the KDM6B gene promoter | AD346 | AD315 |

**Figures**

- **Figure S8a:** Proximal super-enhancer, near the KDM6B gene promoter.
- **Figure S8b:** Proximal super-enhancer, near the MPP1 gene promoter.

**Notes**

- Sample detection (%): CML BM, control BM, CML PB, control PB, total CML, total control.
- Genomic characteristics include: c-MYC Negative, Hit 1: broad domain; near the intron; quiescent region.
| Genomic Loci | Interaction (F/R) | Primer Set and Dilution | DNA Input (ng) | CML (Bone Marrow) | Control (Bone Marrow) | CML (Peripheral Blood) | Control (Peripheral Blood) | Sample detection (%) | Sample BM (%) | BM Control (%) | PB Control (%) | Total CML (%) | Total BM (%) | Total PB (%) |
|-------------|------------------|-------------------------|---------------|------------------|----------------------|----------------------|----------------------|----------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| **TPS3**    | **Bait - Hi**    | FULL2 3/1_1X            | 1.35          | 5                 | 5                    | 0                    | 0                    | 0                    | 97              | 100            | 100            | 100            | 94             | 100            |
|             |                  | FULL2 3/1_1/2X          | 0.65          | 0                 | 0                    | 0                    | 0                    | 0                    | 81              | 89             | 78             | 81             | 83             | 79             |
|             |                  | FULL2 3/1_1/4X          | 3.25          | 0                 | 0                    | 0                    | 0                    | 0                    | 75              | 75             | 70             | 75             | 70             | 70             |
|             |                  | FULL2 3/1_1/6X          | 1.625         | 0                 | 0                    | 0                    | 0                    | 0                    | 10              | 11             | 4              | 11             | 4              | 11             |
|             |                  | FULL2 3/1_1/16X         | 0.8125        | 0                 | 0                    | 0                    | 0                    | 0                    | 6               | 6              | 2             | 6              | 2              | 6              |
|             |                  | FULL2 3/1_1/32X         | 0.40625       | 0                 | 0                    | 0                    | 0                    | 0                    | 3               | 3              | 0             | 3              | 0              | 3              |
|             | **Bait - Lo**    | FULL2 3/1_1X            | 1.35          | 5                 | 5                    | 0                    | 0                    | 0                    | 97              | 100            | 100            | 100            | 94             | 100            |
|             |                  | FULL2 3/1_1/2X          | 0.65          | 0                 | 0                    | 0                    | 0                    | 0                    | 81              | 89             | 78             | 81             | 83             | 79             |
|             |                  | FULL2 3/1_1/4X          | 3.25          | 0                 | 0                    | 0                    | 0                    | 0                    | 75              | 75             | 70             | 75             | 70             | 70             |
|             |                  | FULL2 3/1_1/6X          | 1.625         | 0                 | 0                    | 0                    | 0                    | 0                    | 10              | 11             | 4              | 11             | 4              | 11             |
|             |                  | FULL2 3/1_1/16X         | 0.8125        | 0                 | 0                    | 0                    | 0                    | 0                    | 6               | 6              | 2             | 6              | 2              | 6              |
|             |                  | FULL2 3/1_1/32X         | 0.40625       | 0                 | 0                    | 0                    | 0                    | 0                    | 3               | 3              | 0             | 3              | 0              | 3              |
|             | **Bait - Control** | FULL2 3/1_1X           | 1.35          | 5                 | 5                    | 0                    | 0                    | 0                    | 97              | 100            | 100            | 100            | 94             | 100            |
|             |                  | FULL2 3/1_1/2X          | 0.65          | 0                 | 0                    | 0                    | 0                    | 0                    | 81              | 89             | 78             | 81             | 83             | 79             |
|             |                  | FULL2 3/1_1/4X          | 3.25          | 0                 | 0                    | 0                    | 0                    | 0                    | 75              | 75             | 70             | 75             | 70             | 70             |
|             |                  | FULL2 3/1_1/6X          | 1.625         | 0                 | 0                    | 0                    | 0                    | 0                    | 10              | 11             | 4              | 11             | 4              | 11             |
|             |                  | FULL2 3/1_1/16X         | 0.8125        | 0                 | 0                    | 0                    | 0                    | 0                    | 6               | 6              | 2             | 6              | 2              | 6              |
|             |                  | FULL2 3/1_1/32X         | 0.40625       | 0                 | 0                    | 0                    | 0                    | 0                    | 3               | 3              | 0             | 3              | 0              | 3              |

Figure S8c
### Table

| Primer set | AD35 | AD35a | AD96 | AD256 | AD294 | AD301a | AD454 | AD548 | AD454 | AD548 | AD278 | AD283 | AD419 | AD446 | AD394 | AD412 | AD289 | AD339 |
|------------|------|-------|------|--------|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| c-MYC Bait-Hit 1 | 0/3 3/3 0/3 2/3 | 0/3 2/3 2/3 3/3 | 0/3 0/3 3/3 3/3 | 3/3 3/3 2/3 2/3 | 0/3 2/3 | 3/3 3/3 |
| c-MYC Bait-Hit 2-1 | 1/3 3/3 3/3 3/3 | 2/3 3/3 3/3 2/3 | 3/3 3/3 3/3 2/3 | 3/3 3/3 3/3 2/3 | 1/3 3/3 |
| c-MYC Bait-Hit 2-2 | 3/3 3/3 3/3 3/3 | 2/3 3/3 3/3 3/3 | 2/3 1/3 3/3 3/3 | 3/3 3/3 3/3 3/3 | 3/3 3/3 |
| Negative control 1 | 0/3 2/3 1/3 0/3 | 0/3 1/3 0/3 1/3 | 0/3 0/3 1/3 0/3 | 0/3 1/3 1/3 0/3 | 1/3 2/3 |

### Figures

#### Figure S9

- **a**: Table showing expression levels for different primer sets across CML and Other patients.
- **b**: Bar graph showing expression levels for different samples.
- **c**: Bar graph showing % Input for different samples.

---

**Table Legend**
- MYC-335
- MIR1205 Superenhancer
- MYC Promoter
- Negative Control
Figure S10
Figure S11
### Count of broad domains

|       | DSE | PSE | DTE | PTE | Non-overlapping | Total |
|-------|-----|-----|-----|-----|-----------------|-------|
| DBD   | 4   | 1   | 11  | 1   | 9               | 26    |
| PBD   | 1   | 62  | 6   | 1049| 148             | 1264  |
| DTD   | 115 | 85  | 1783| 96  | 7339            | 9418  |
| PTD   | 2   | 108 | 39  | 5945| 9007            | 15101 |

### Count of super-enhancers

|       | DBD | PBD | DTD | PTD | Non-overlapping | Total |
|-------|-----|-----|-----|-----|-----------------|-------|
| DSE   | 4   | 1   | 81  | 2   | 59              | 147   |
| PSE   | 1   | 61  | 57  | 78  | 22              | 219   |
| DTE   | 11  | 6   | 1725| 39  | 4917            | 6698  |
| PTE   | 1   | 1047| 92  | 5746| 1216            | 8102  |

---

**Figure S12**

- **Figure S12a**: Table showing the count of broad domains.
- **Figure S12b**: Table showing the count of super-enhancers.
- **Figure S12c** to **Figure S12h**: Box plots illustrating various types of data distributions.
Figure S13
Figure S14