Three-dimensional Genome Organization Maps in Normal Haematopoietic Stem Cells and Acute Myeloid Leukemia

Benny Wang[1,2*], Lingshi Kong[1,2*], Deepak Babu[2*], Ruchi Choudhary[1,2], Winnie Fam[2], Jia Qi Tng[2], Yufen Goh[2], Xin Liu[3], Fang Fang Song[3], Priscella Chia[3], Ming Chun Chan[4], Omer An[2], Cheng Yong Tham[2], Touati Benoukraf,[2,5], Henry Yang[2], Wilson Wang[4], Wee Joo Chng[2,3,6], Daniel Tenen[2], Melissa Jane Fullwood[1,2,7^]

[1]School of Biological Sciences, Nanyang Technological University, [2]Cancer Science Institute of Singapore, National University of Singapore, [3]National University Cancer Institute, National University Health System, [4]Department of Orthopaedic Surgery, Yong Loo Lin School of Medicine, National University of Singapore, [5]Memorial University of Newfoundland, [6]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, [7]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore.
*Co-first authors
^Corresponding author

Corresponding author contact: Melissa Jane Fullwood, mfullwood@ntu.edu.sg

Abstract
Acute Myeloid Leukemia (AML) is a highly lethal blood cancer arising due to aberrant differentiation of haematopoietic stem cells. *MEIS1* and *HOXA9* regulate stemness-related transcriptional programs in normal haematopoietic stem cells and AML. Here we obtained 3D genome organization maps in the CD34+ haematopoietic stem cells from 3 healthy individuals and 3 individuals with AML. The *MEIS1* oncogenic transcription factor is regulated by a Frequently Interacting Region (FIRE). This FIRE is present in normal bone marrow samples, and an AML sample with high *MEIS1* levels. However, it is absent in two AML samples that show low *MEIS1* levels. CRISPR excision of the FIRE led to loss of *MEIS1* and reduced cell growth. Moreover, MEIS1 can bind to the promoter of *HOXA9*, and HOXA9 can also auto-regulate by binding to its own promoter as well as an Acute Myeloid Leukemia-specific super-enhancer that interacts with the *HOXA9* promoter via chromatin interactions.

Significance
Many oncogenes, such as *MEIS1* and *HOXA9*, are overexpressed in some but not all cancers. We identified two key epigenetic mechanisms underlying this heterogeneity in oncogene expression in Acute Myeloid Leukemia. This mechanism could be potentially exploited to utilize epigenetic inhibitors to specifically target oncogene expression in cancer.

Introduction

Acute Myeloid Leukemia (AML) is a highly lethal cancer which is characterized by a block in the differentiation and abnormal proliferation of hematopoietic progenitor cells (1). While complete remissions in patients are generally achieved, relapses are commonly observed and eventually fatal (2, 3). The DNA mutation load of AML is lower than that of solid cancers, and analyses of the mutated genes have revealed that epigenetic factors and transcription factors are frequently mutated in AML (4, 5), pointing to epigenetic dysregulation as being important to AML pathogenesis. HOXA9 and MEIS1 oncogenic transcription factors are overexpressed in more than half of AML cases (6-9), and overexpression of *HOXA9* and *MEIS1* oncogenes is associated with poor prognosis (9-11). Here we asked whether there are epigenetic mechanisms behind the heterogeneity of oncogene expression in AML.

One potential source of heterogeneity is that of 3D genome organization. The genome is organized into chromatin interactions, which refer to two or more separate genome loci that come together in close spatial proximity (12), such as enhancer-promoter loops and also large domains of interacting regions, called Topologically-Associated Domains (TADs). Clusters of enhancers marked by high H3K27ac signals called super-enhancers (SEs) can be acquired by cancer cells, and have been identified to be associated with oncogene activation in cancer (13). We and others have found that SEs can loop via chromatin interactions to distant oncogenes (14). Moreover, cancer cells can show altered TADs (15) and altered chromatin loops at key oncogenes such as *TERT* (16, 17). Thus, we reasoned that the interplay between SEs and chromatin interactions might underlie the heterogeneity of *HOXA9* and *MEIS1* expression in AML.

Here we performed Hi-C to obtain 3-dimensional genome organization maps in 3 samples of CD34+-enriched AML bone marrow clinical samples and 3 CD34+ enriched bone marrow samples from healthy individuals. Interestingly, the *MEIS1* locus shows a "Frequently Interacting Region" (FIRE) that can be found in all healthy bone marrow samples, but only one of three AML patient samples. FIREs are regions of unusually high local contact frequency that tend to be depleted near Topologically-Associated Domain (TAD) boundaries and enriched towards the centre of TADs (18). The AML patient with intact *MEIS1* FIRE has high *MEIS1* gene expression, but this was absent in the other two samples with absent *MEIS1* FIRE. Removal of a FIRE boundary at the *MEIS1* gene locus leads to loss of *MEIS1* gene expression, slower leukemic cell growth and reorganisation of existing chromatin loops. *HOXA9*, in contrast to *MEIS1*, shows chromatin interactions that are present in all AML patient samples and all healthy samples, however, analysis of a large dataset of H3K27ac ChIP-Seq in AML patient samples (19) indicates the loci that interact with the *HOXA9* show differences in occupancy by SEs. Patient samples that contain SEs at the interacting loci show higher *HOXA9* expression.

Taken together, in this paper, we identified two key epigenetic mechanisms underlying the heterogeneity of *MEIS1* and *HOXA9* oncogene expression in AML. First, heterogeneity of chromatin interaction structures in AML cells can explain the heterogeneous expression of *MEIS1* in AML. Second, heterogeneity of AML-acquired SEs at genomic loci that show chromatin interactions to *HOXA9* can explain heterogeneous expression of *HOXA9* in AML. This information could be useful in the development of targeted epigenetic therapies against *HOXA9* and *MEIS1* in AML while sparing healthy cell types such as normal haematopoietic stem cells.

Results

A compendium of Hi-C maps of Topologically Associated Domains and Chromatin Loops in AML and normal Haematopoietic Stem Cells

To date there is no detailed study of topologically associated domains (TADs) and chromatin loops in AML as compared with normal haematopoietic stem cells, which are the precursor cells for AML development. Here we performed deep sequencing Hi-C in 3 samples of CD34+-enriched AML bone marrow clinical samples ("AML28", "AML29", "AML30") and 3 CD34+-enriched bone marrow samples from healthy individuals obtained from knee replacement surgery ("Knee47", "Knee49", "Knee50") (**Table S1**). All samples were obtained with patient consent.

We obtained more than a billion sequencing reads for each Hi-C library, and analysed the data by Juicer (20) (**Table S2**). TADs were called using Arrowhead (20) (**Table S3**), while loops were called using HiCCUPS (20) (**Table S4**). Heatmaps were visualized with Juicebox (21). The statistics of the

Hi-C analyses are shown in **Figure 1A**. Several thousands of TADs and over ten thousand loops were called from our Hi-C data, reflecting that our libraries are of high resolution (AML28: 1,682 TADs and 24,394 loops; AML29: 1,108 TADs and 19,541 loops; AML30: 1,296 TADs and 10,733 loops; Knee47: 1,153 TADs and 4,733 loops; Knee49: 927 TADs and 13,107 loops; Knee50: 1,234 TADs and 13,795 loops). Examples of TADs at the *MEIS1* and *MYC* oncogenes are shown in **Figure 1B**.

Hi-C analyses are capable of revealing translocations (22), and therefore we analysed the Hi-C data for structural variants in these patient samples by Dovetail Genomics' "Selva" approach for analysing structural variants (**Supplementary Methods**). AML28 and AML29 showed fewer than 5 translocations and no known oncofusions from Hi-C analyses (**Figure 1C**, **Figure S1**), while AML30 showed a common t(8; 21) translocation which fuses *AML1* (also called *RUNX1*) with the *ETO* gene (also called *RUNX1T1*) (**Table S5**, **Figure S1**).

Identification of a Frequently Interacting Region (FIRE) at MEIS1 in haematopoietic stem cells and specific AML samples

Next, we examined the chromatin interactions that are different between the different clinical samples. At the *MEIS1* locus, analysis of the Hi-C heatmaps indicated that all the knees and AML29 have an unusual region consisting of many interactions in a local space (small, tight, dense square shown in the Hi-C heatmap in **Figure 1D**), and this is absent in both AML28 and AML30 (**Figure 1E**). The tight loop structure is shown in more detail through close-ups in **Figure 1F**. Gene expression analysis by digital droplet polymerase chain reaction (ddPCR) of *MEIS1* shows that AML28 and AML30, which have both lost the tight loop structure region, have lower levels of *MEIS1* as compared with AML29 (**Figure 1G**). Furthermore, the gene expression levels of *MEIS1* in AML28 and AML30 as examined by RNA-Seq (23) (24) are also observed to be lower than those of normal CD34+ haematopoietic stem cells, while the gene expression level of *MEIS1* in AML29 is much higher than those of normal CD34+ haematopoietic stem cells (**Figure 1G**). Examination of Juicer-called TADs and loops of this interacting structure indicates that the interacting structure encapsulates the *MEIS1* gene from promoter to terminator.

Juicebox (21), (25) images of the region in GM12878 human lymphoblastoid cell line show that this structure does not exist in GM12878 lymphoblastoid cells. However, this structure exists in the K562 (Chronic Myelogenous Leukemia cell line) and other cell lines of different tissue types such as HUVEC (human umbilical vein endothelial cells), HMEC (human mammary epithelial cells), IMR90 (human fetal lung fibroblast cell line), HeLa (human cervical carcinoma cell line), HAP1 (human near-haploid Chronic Myelogenous Leukemia cell line) and THP-1 (AML cell line) (**Figure 1H & Figure S2).**

We characterized this structural region in detail using Circular Chromosome Conformation Capture (4C) (26, 27) in the THP-1 AML cell line using the *MEIS1* promoter as a bait and found that many interactions to *MEIS1* promoter occur within the small, dense square region, confirming that the region indeed consists of many interactions. While most interactions occur within the structural region, there are a few chromatin interactions that extend outwards to additional regions within the larger TAD structure (**Figure 1H**). Additionally, 4C analysis of the *MEIS1* promoter in HL-60, another AML cell line, also shows similar chromatin interactions as those found in the THP-1 AML cell line (**Figure 1H**).

Given that the small TAD-like structure appears in the context of a larger TAD structure but possess chromatin interactions that are unconfined within the small TAD-like structure, we reason that this pattern is characteristic of a recently-reported novel class of chromatin interactions called "Frequently Interacting Regions" ("FIRES"), which are regions of unusually high local contact frequency that tend to be depleted near TAD boundaries and enriched towards the center of TADs (18). Examination of the heatmap in **Figure 1H** reveals that the FIRE is located away from the boundary and towards the center of a large TAD. Moreover, FIRES are often reported to be located near cell identity genes and since *MEIS1* is important in maintaining stem cell-like identities (28), supporting the notion that the region of interest is likely to be a FIRE.

One phenomenon of FIREs is their tendency to be enriched for SEs. We hypothesized that AML cells will contain SEs that interact with the FIRE that contains the *MEIS1* gene. We investigated this idea by using a dataset of H3K27ac ChIP-Seq libraries that had been performed on bone marrow from 66 AML patients and 2 healthy individuals (19) to identify AML-acquired SEs. We found that approximately 87% (51 out of 58) of AML patient show the presence of strong H3K27ac ChIP-Seq

signals at the *MEIS1* FIRE region **(Figure S3A)**, suggesting the presence of active transcriptional machinery within the FIRE. Interestingly, PBX3, which cooperates with MEIS1 in AML (29) (30), also shows a FIRE in AML29 (**Figure S4**) with about 71% (40 out of 56) of AML patients showing high H3K27ac ChIP-Seq levels and SEs within the *PBX3* FIRE region (**Figure S6**). Taken together, our results indicate that FIREs at *MEIS1* and *PBX3* are associated with high H3K27ac ChIP-Seq levels and SE presence, indicating that these FIREs are associated with active chromatin in a large fraction of AML patients.

Next, we examined whether there are any SEs in AML samples (19) that occupy the genomic regions that we identified as having chromatin interactions with *MEIS1* promoter. We found three SEs (SE 1, 2 & 3) that are located at the *MEIS1* FIRE region. SE1 and SE3 are upstream and downstream of the *MEIS1* gene respectively and can be found in about 79% (46 out of 58) of AML patient samples. SE2, which is located within the *MEIS1* intron, can be found in approximately 58% (34 out of 58) of the AML patients. SE1 was found in two different clinical samples of healthy CD34+ haematopoietic stem cells, and SE2 and SE3 were both present in one of two clinical samples (**Figure 1H**), suggesting that several SEs that drive *MEIS1* at high levels are present in haematopoietic stem cells and retained in certain AML patients, while additional SEs may be acquired in AML pathogenesis.

By contrast, other cell types that possess the FIRE but do not show high *MEIS1* expression such as Human Umbilical Vein Endothelial Cells (HUVEC) show strong H3K27me3 ChIP-Seq signals in the FIRE region, indicating that the FIRE is associated with repressive chromatin marks in these cells (**Figure 2B & S3B**). Other cell types such as Human Mammary Epithelial Cells (HMEC) do not show strong H3K27me3 ChIP-Seq signals, but do not show strong H3K27ac ChIP-Seq signals, indicating the FIRE may be present but not associated with active transcription (**Figure S3B**).

Maintenance of a Frequently Interacting Region (FIRE) chromatin loop boundary near MEIS1 is important for high MEIS1 gene expression in the K562 cell line

Next, we investigated the consequences of perturbation of the *MEIS1* FIRE through CRISPR deletion of a FIRE border. FIRE formation is partially dependent on CTCF (18), and we observed that the 3' end of *MEIS1* which is near a border of the FIRE is marked by a strong CTCF binding site located at a CpG island that can be found in multiple cell lines (**Figure 2A-B**). We asked whether maintenance of the CTCF binding site was important in maintaining high levels of *MEIS1* gene expression and proceeded to excise the CTCF binding site in K562 cells via CRISPR. We confirmed that the FIRE is also present in K562 cells, thus making it a suitable model for the study of this particular FIRE (**Figure 2B**). Our 4C results of the CTCF excised K562 cells as compared with empty vector control cells show a clear loss of *MEIS1*-promoter linked chromatin interactions within the *MEIS1* locus. We observed losses of chromatin interactions between the *MEIS1* promoter and the SEs (SE1, 2 & 3) **(Figure 2C).** Furthermore, we see an apparent loss of *MEIS1* associated distal chromatin loops that were beyond the immediate CTCF boundaries of *MEIS1* (**Figure 2C**). Taken together, our results indicate that excision of the CTCF binding site at the border of the *MEIS1* FIRE can lead to severe losses of chromatin interactions.

Examination of *MEIS1* by reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) in cells treated with empty vector showed high levels of *MEIS1*, but a marked loss or downregulation of *MEIS1* was observed in the CTCF CRISPR knockout cells (**Figure 2D**). *MYC*, which was previously shown to be a downstream target of *MEIS1* in zebrafish (31), also showed significantly lower gene expression in CTCF CRISPR knockout cells (**Figure 4D**). The CTCF CRISPR knockout cells also grew more slowly in culture as compared to the control empty vector K562 cells (**Figure 2E**). Overall, we found that maintenance of the FIRE CTCF sites is important in maintaining *MEIS1* gene expression levels and cell growth in myeloid leukemia cells.

Taken together, our results suggest a model of *MEIS1* gene regulation in which blood stem cells have a FIRE at *MEIS1* that shows many chromatin interactions to other genomic regions, which can be occupied by SEs (**Figure 2F-G**). The interplay between SEs and the FIRE may drive *MEIS1* at high levels in blood stem cells and AML samples with intact FIREs. In our AML28 and AML30 samples, we observed altered loops, resulting in the *MEIS1* promoter losing contact with other regulatory elements, including the three genomic loci that are frequently occupied by SEs in AML patient samples. These samples had lower MEIS1 expression, which we speculate may be due to the loss of SE-promoter

contacts in these samples. Therefore, the presence of different chromatin conformations in various AMLs at *MEIS1* and their varying interactions with SEs may explain the heterogeneity of *MEIS1* expression.

Genomic loci that can loop to the *HOXA9* oncogene display heterogeneity in SE occupancy in AMLs

Next, we investigated *HOXA9*, an oncogenic transcription factor that tends to be co-expressed with *MEIS1*. *HOXA9* is expressed at a moderately high level in CD34+ normal haematopoietic stem cells and overexpressed in half of AMLs (**Figure 3A)**. To characterize the heterogeneity of *HOXA9* levels in AML, we performed digital droplet PCR to compare *HOXA9* levels with progenitor (CD34+ haematopoietic stem cells) and more differentiated myeloid cells (CD33+ myeloid cells) derived from healthy bone marrow samples. We found *HOXA9* expression levels to be higher in CD34+ cells, while lower *HOXA9* expression levels were observed in myeloid cells. By contrast, AML samples exhibited a range of *HOXA9* expression levels; from very low levels to very high levels but all of these expression levels are higher than those of the CD34+ cells from knees (**Figure 3A**). Next, we performed ddPCR of *HOXA9* on the three AML samples for which we have Hi-C data as well as three other CD34+ enriched cell populations from healthy patients ("Knee25, 26, 27"). We found that all had varying levels of *HOXA9* (**Figure 3B**).

We found strong chromatin interactions and/or TADs around the *HOXA9* gene, which also show interactions with the *HOXA9* gene promoter (**Figure 3C-D**). However, the *HOXA9*-associated chromatin interactions and TADs were found in all knee samples as well as AML samples (**Figure 3C-D**), thus indicating that different chromatin interaction usage was not the reason behind the different levels of *HOXA9* oncogene expression in these cells. We then investigated AML cell lines that have high or low levels of *HOXA9* oncogene expression. ddPCR analysis of *HOXA9* expression levels in THP-1 and HL-60 AML cell lines indicate that *HOXA9* is specifically upregulated in THP-1 but not HL-60 (**Figure S6**). Next, we investigated the chromatin interactions using Circular Chromosome Conformation Capture (4C), which was performed with the *HOXA9* gene promoter as the bait region in three different myeloid leukemic cell lines (**Table S8**). We found that the chromatin interactions were similar between HL-60 and THP-1 cell lines (**Figure 3E**), confirming that different chromatin interaction usage was not the reason behind the different levels of *HOXA9* oncogene expression in these cells.

Since the chromatin interactions at *HOXA9* are similar between healthy CD34+ haematopoietic stem cells as well as AML patients, we hypothesized that SE acquisition in AML might be heterogenous, with some patients acquiring certain SEs and not others, possibly leading to the heterogeneous overexpression of *HOXA9* seen in some AML patients. Upon examining the patient SEs identified earlier from McKeown et al (19), we found that 63% of SEs could be found in fewer than 5 patients, and fewer than 2% of them could be found in more than 55 of 66 patients, indicating that SEs in AML are heterogeneous (**Figure 4A**).

To characterize the interplay between SEs and chromatin interactions, we then intersected the SE information with the chromatin loops detected in the normal and AML Hi-C data. In AML28, we observed that 74% of SEs were associated with chromatin interactions. In AML29, 70% SEs were associated with chromatin loops. In AML30, 54% super enhancers were associated with chromatin loops (**Figure 4A**). Some of the observed SEs connect to genes with potential roles in cancer via chromatin interactions. For example, an AML SE can connect with *PKP4* gene, known to be differentially expressed in human lung squamous cell carcinoma (32) through two chromatin interactions (**Figure S8B**).

We further observed that approximately half of the chromatin interactions associated with AML-specific SEs could be also found in normal CD34+ haematopoietic stem cells (4907 pre-existing interactions out of 10548 = 46.5% in AML28, 4235 pre-existing interactions out of 8132 = 52.07 % in AML29 and 2613 pre-existing interactions out of 4575 = 57.11% in AML30) (**Table S7)**. This indicates that while some SEs that are newly acquired in certain AML patients might be associated with newly formed chromatin interactions, other SEs that are newly acquired in AML might occupy pre-existing chromatin interactions that are also found in the precursor cells (haematopoietic stem cells) to regulate target genes.

We then investigated SE occupancy of chromatin interactions at *HOXA9*. We detected the presence of multiple SEs (indicated by blue regions in **Figure 4B**) that show interactions with the

HOXA9 promoter through 4C and/or Hi-C. In particular, we observed a SE near the *SNX10* gene that loops over to *HOXA9* (**Figure 4B**), which we analysed further. This SE is specifically present in 18 of 58 of the AML patients and is not found in the normal CD34+ haematopoietic stem cell data from McKeown et al. (19). Our analysis revealed that about 31% of AML patients carry this *SNX10* SE (**Figure 4C**). The presence of the *SNX10* SE was associated with upregulated high *HOXA9* levels in the AML patient samples (**Figure 4D**).

Next, because HOXA9 and MEIS1 are transcription factors, we investigated how HOXA9 and MEIS1 bind to SE and promoters in the *HOXA9* and *MEIS1* gene regions. ChIP-qPCR of THP-1 at regions of the *SNX10* SE and the *HOXA9* promoter showed the binding of HOXA9 and MEIS1 proteins at both locations (**Figure 4E-F)**. Interestingly, we also observed that the HOXA9 and MEIS1 proteins bind to the *MEIS1* promoter regions, suggesting a tight circuit between HOXA9 and MEIS1. Altogether, our results suggest a model whereby the *MEIS1* FIRE can lead to high *MEIS1* levels. The MEIS1 transcription factor can auto-regulate the *MEIS1* promoter, and also bind to the *HOXA9* promoter. HOXA9 can bind to the *HOXA9* and *MEIS1* promoters, as well as the *SNX10* SE that loops over to the *HOXA9* promoter. This can explain the observation that *MEIS1* and *HOXA9* tend to be upregulated together in AML. Additionally, the heterogeneity of super-enhancer acquisition at chromatin interactions that loop to *HOXA9* in AML patients may explain the heterogeneity of *HOXA9* oncogene expression observed in AML patients (**Figure 4G**).

Discussion

A major focus in understanding cancer is identifying the molecular underpinnings of the heterogeneity of oncogene expression in different cancer clinical samples. We identified two mechanisms involving chromatin interactions that may give rise to heterogeneity of oncogene expression. The first mechanism is that different AMLs may show different patterns of chromatin interactions around oncogenes, thus leading to different oncogene expression levels, such as *MEIS1*, which is regulated by a FIRE. A second mechanism is that SEs may be heterogeneously acquired in AML. The AMLs which have acquired such SEs may show high levels of oncogene expression because the SEs may occupy, or "hijack", pre-existing enhancer-promoter chromatin interaction circuits present in precursor cells such as haematopoietic stem cells, thus driving oncogenes strongly in certain AMLs but not others.

HOXA9 and MEIS1 have been widely reported to be important contributors to the progression of AML (9, 10). The synergistic effect of HOXA9 and MEIS1 also leads to the development of aggressive AML with poor prognosis. Despite the progress in AML research, the understanding of *HOXA9* and *MEIS1* regulation at the epigenomic level remains limited. Here, we performed an in-depth analysis of 3D genome organization by Hi-C in AML and its precursor cells, haematopoietic stem cells (marked by CD34+), and discovered auto-regulatory and cross-regulatory chromatin interaction circuits necessary for maintaining high levels of *MEIS1* and *HOXA9* which are seen in some subtypes of leukemia.

Our results demonstrate for the first time that a FIRE at the *MEIS1* region is critical in maintaining high *MEIS1* levels. Interestingly, the FIRE at *MEIS1* is not observed in AML28 and AML30, but seen in AML29, suggesting that certain AMLs can develop or be maintained without having FIREs around oncogenes such as *MEIS1*. These AMLs are likely to be driven by different oncogenes or other genetic and epigenetic alterations. For example, AML30 contains a translocation to the *RUNX1* oncogene, suggesting that AML30 does not require high *MEIS1* and *HOXA9* levels to function but might be addicted to *RUNX1* instead.

Our data showed that perturbing a CTCF region at a *MEIS1* FIRE boundary could lead to altered chromatin interactions and gene expression of *MEIS1*. Aza-cytidine, a drug commonly used in AML treatment, has recently been shown to alter CTCF levels (33). We speculate that these drugs may work in part by leading to alterations in chromatin interactions, such as the FIRE at *MEIS1*, leading to downregulation of gene expression. Additionally, SE analyses have pinpointed specific oncogenes to which certain AMLs may be addicted, such as retinoic acid receptor alpha (*RARA*), allowing for the development of RARA inhibitors for this category of AMLs (19). Also, a study characterizing chromatin interactions in Acute Lymphocytic Leukemia patient samples by Hi-C indicates that inhibition of super-enhancers by drugs such as THZ1 can lead to reductions in chromatin interactions at target oncogenes and reduced gene expression (34). Our Hi-C maps in this manuscript provide a

framework for us to investigate chromatin interactions that are affected by various AML drugs and identify distal oncogenes that are the targets of distal SE that loop over to the gene promoters via chromatin interactions.

In conclusion, we demonstrated that the interplay between SEs and FIREs is important in upregulating *MEIS1* and *HOXA9* in AML. Different chromatin conformations and/or different occupancy by SEs are associated with different levels of oncogene expression. Thus, our results present a molecular basis for the development of epigenetic inhibitors to target SEs that loop over to *MEIS1* and *HOXA9* in AML.

Methods

Detailed methods are given in the **Supplementary Materials**. Briefly, Bone marrow samples from AML patients were taken from the back of the pelvic (hip) bone while bone marrow from healthy counterparts was withdrawn during Total Knee Arthroplasty as part of a standard operative procedure. All clinical samples were obtained from the National University Hospital Singapore and collected according to the requirements of the Human Biomedical Research Act. Informed consent was obtained for all clinical samples used in the study. Isolation of CD34+ cells from normal sample MNCs was performed according to the manufacturer's instructions using CD34 MicroBead Kit UltraPure. Acute Myeloid Leukemia cells THP-1 and HL-60 and Chronic Myelogenous Leukemia cells K562 were cultured at 5% CO_2 at 37^0C. THP-1 and K562 were cultured with Roswell Park Memorial Institute (RPMI) 1640 media (Hyclone) supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS; Hyclone) and 1% penicillin/streptomycin (Hyclone). HL-60 cells were cultured using Iscove's Modified Dulbecco's Medium (IMDM; Gibco), supplemented with 20% heat inactivated FBS (Hyclone) and 1% penicillin/streptomycin (Hyclone). Hi-C was performed through Dovetail Genomics and sequenced by Illumina sequencing. Hi-C data were aligned and processed by Juicer (version 1.5). The reference genome was hg38. RNA was extracted using the DNA/RNA Allprep Kit (Qiagen). Reverse transcription of RNA into cDNA was performed using the qScript cDNA Supermix (Quantabio). Quantitative polymerase chain reaction (qPCR) was performed with the GoTaq qPCR Mastermix (Promega) and QuantStudio 5 Real Time PCR (Applied Biosystems). ddPCR experiments on cDNA were performed with the EvaGreen Mastermix (Biorad) and the QX200 Droplet Digital PCR system. 4C-seq was performed as previously described with some modifications as described in the Supplementary Materials (35). CRISPR-Cas9 excision was performed with the All-in-One vector system as described previously (36). Transfection of K562 cells was performed with the Neon Transfection System (ThermoFisher). All primer sequences are reported in **Table S9**.

Acknowledgements

This research is supported by the National Research Foundation (NRF) Singapore through an NRF Fellowship awarded to M.J.F (NRF-NRFF2012-054) and NTU start-up funds awarded to M.J.F. This research is supported by the RNA Biology Center at the Cancer Science Institute of Singapore, NUS, as part of funding under the Singapore Ministry of Education Academic Research Fund Tier 3 awarded to Daniel Tenen (MOE2014-T3-1-006). This research is supported by the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiative.

Author contributions

B.W, L.K, D.B and M.J.F conceived of the research. M.J.F, B.W, L.K, D.B, R.C and D.G.T. contributed to the study design. B.W generated the CRISPR clones and performed 4C, ddPCR, qPCR, ChIP-qPCR and growth assays. L.K performed bioinformatics analysis of the clinical Hi-C and RNA-Sequencing data. D.B prepared the AML and Knee samples for Hi-C by Dovetail genomics. D.B and W.F prepared the clinical samples for RNA-Sequencing. D.B performed manual curation of the Hi-C data. R.C performed bioinformatics analysis on published AML cases to identify SEs and their looping patterns in AML. W.F, J.Q.T and Y.G collected knee and bone marrow samples and isolated CD34+ and CD33+ cells. MCC and WW provided normal bone marrow clinical samples. O.A. and H.Y provided the bioinformatics pipeline and CSI web portal for the 4C analyses. C.Y.T and T.B provided the access for the online storage of clinical data. F.F.S, P.C and C.W.J provided AML clinical samples

and annotation. B.W, L.K, D.B, R.C, D.G.T. and M.J.F reviewed the data. B.W, L.K, D.B and M.J.F wrote the manuscript. All authors reviewed and approved of the manuscript.

Competing interests

M.J.F declares two patents on methodologies related to ChIA-PET. No other conflicts of interest are declared.

A

Sample	Reads	#TADs	#loops
AML28	1,290,109,443	1,682	24,394
AML29	1,245,251,169	1,108	19,541
AML30	1,387,562,168	1,296	10,733
Knee 47	1,337,973,222	1,153	4,733
Knee49	1,156,332,234	927	13,107
Knee50	1,165,296,399	1,234	13,795

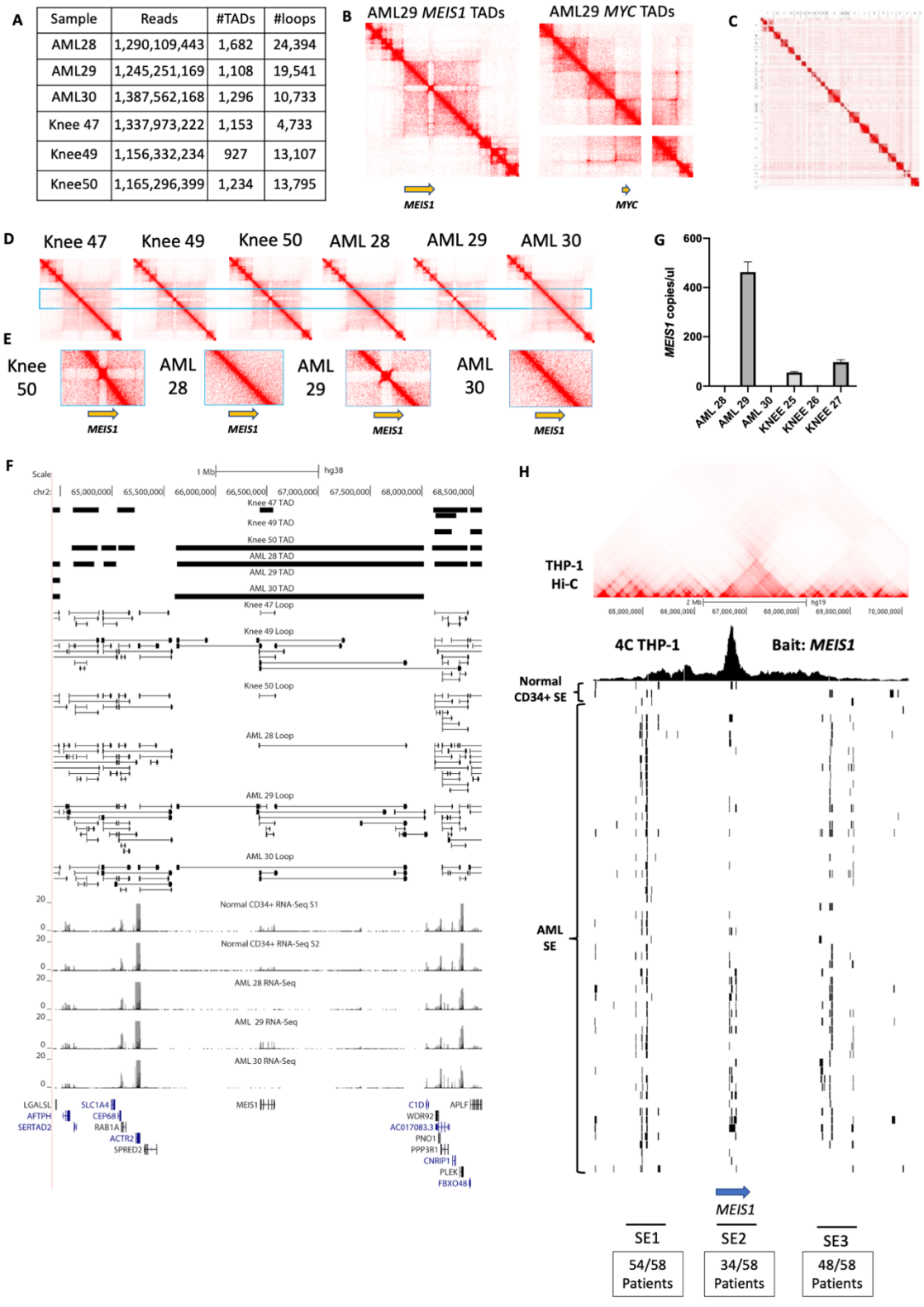


Figure 1. Analyses of chromatin interactions at the *MEIS1* region. A. Table of statistics of Hi-C data. **B.** Coverage normalized heatmaps of AML29 at the oncogenes *MEIS1* (genomic coordinates - chr2:64,377,585-69,377,584) and *MYC* region (genomic coordinates - chr8:125,253,102-130,253,101). **C.** Heatmap showing all chromosomes for AML29. **D.** Hi-C heatmaps of knee 47, 49 and 50, AML28, 29 and 30 indicate that AML 28 & 30 has an unusual loss of a sub-TAD (genomic coordinates - chr2:64,427,585-68,581,684). **E.** Zoomed in view of FIRE region of the clinical samples. **F.** Screenshot on the UCSC genome browser of the loops corresponding to the sub-TAD at the *MEIS1* gene (genomic coordinates - chr2:64,427,585-68,581,684). Coverage normalization was used to visualize the Hi-C heatmaps. **G.** ddPCR of *MEIS1* in CD34+ cells of clinical AML and normal bone marrow samples. **H.** Heatmap of the region in THP-1 cells from published data (37) is shown followed by 4C of THP-1 indicating that *MEIS1* shows chromatin interactions to three genomic regions that are marked by SEs in AML patients (SE1, SE2 & SE3). Genomic coordinates for view: chr2:63,160,137-71,161,280.

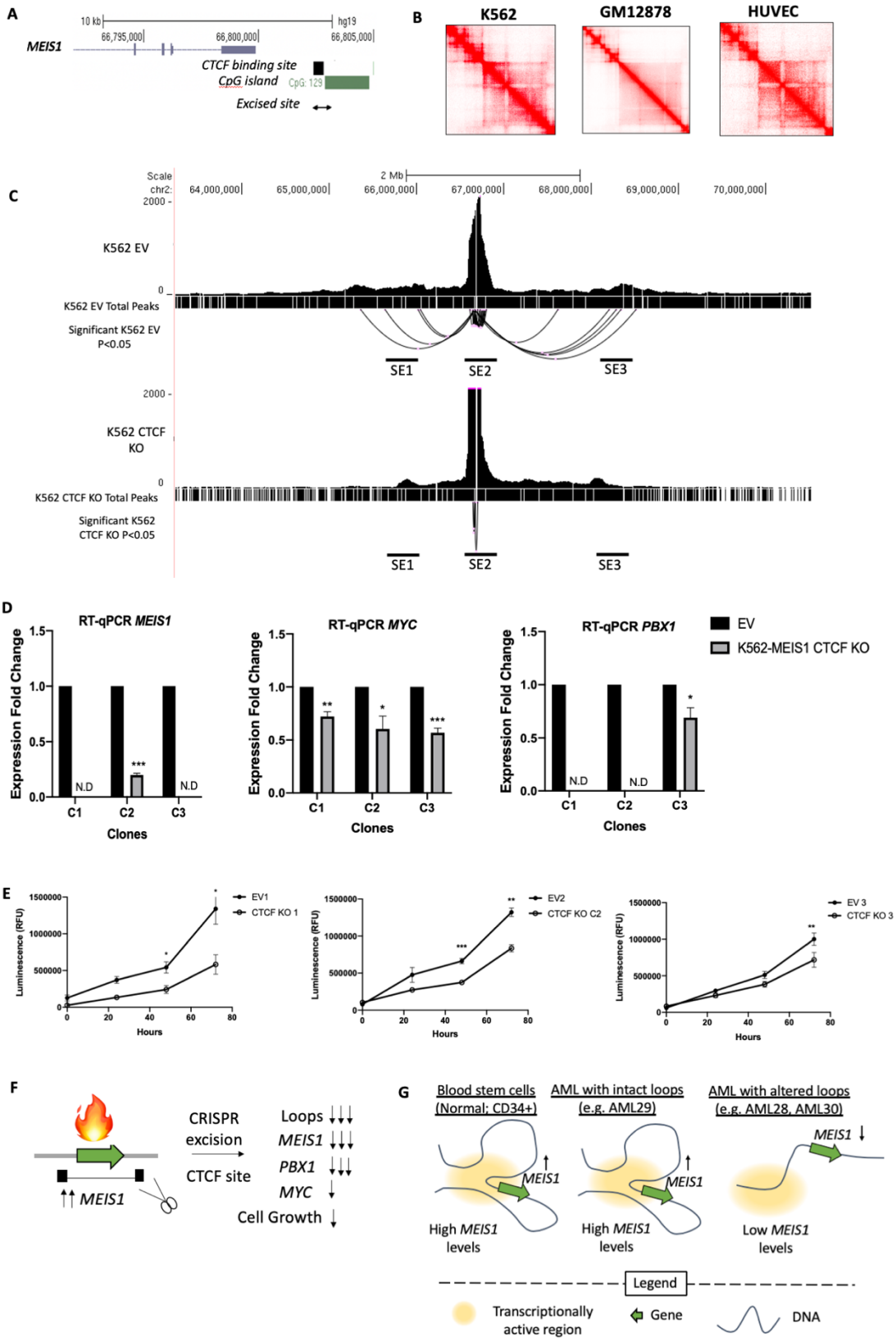


Figure 2. CRISPR excision of a CTCF binding site near the *MEIS1* FIRE leads to loss of *MEIS1*. **A.** A close-up view of the *MEIS1* 3' end which is the right side of the *MEIS1* FIRE, together with the CTCF binding site indicated from analysis of ENCODE CTCF ChIP-Seq data (38, 39) and a nearby CpG island. The excised site is indicated by a double-headed arrow, is several hundred nucleotides long, and covers the CTCF binding site. **B.** The CRISPR excision was performed in K562, and the heatmap of the *MEIS1* FIRE region is shown to confirm that K562 cells also have this FIRE. Hi-C heatmap of the same *MEIS1* FIRE in GM12878 and HUVEC cell lines show the FIRE is only present in selected cell types. Previously published Hi-C data was obtained to generate the heatmaps seen in panel B (25). Coverage normalization was used to visualize the Hi-C heatmaps. The genomic coordinates visualized in the Hi-C contact map is chr2:64,427,585-68,581,684. **C.** 4C of K562 empty vector and CTCF KO are shown in comparison with their total peaks and significant loops (P<0.05). The locations of the SEs are marked as SE1-3. **D.** *MEIS1* RT-qPCR (left), *MYC* RT-qPCR and *PBX1* RT-qPCR (far right) in negative control cells which went through the CRISPR process with empty vector ("empty vector") and CRISPR knockout ("CRISPR KO") cells. Data shown are average +/- standard error from three biological replicates. "N.D." indicates no gene expression was detected. Two-tailed t-test was performed to evaluate significance, and the asterisks indicate the data is significant as per *** - P< 0.001 ** - P< 0.01, *- P < 0.05 level. **E.** Cell viability assay was performed, and viable cell levels are shown on the y-axis. Data shown are average +/- standard error from three biological replicates. One-tailed t-test was performed to evaluate significance, and the asterisks indicate the data is significant as per ** - P< 0.01, *- P < 0.05 level. **F.** Schematic summary of the data presented. CRISPR excision of the FIRE (indicated by the "fire" icon) at *MEIS1* leads to *MEIS1* gene expression and other cellular changes in myeloid leukemia. **G.** A proposed schematic of how the loss of FIRE boundary at the *MEIS1* region results in the dysregulation of chromatin loops and downregulated *MEIS1* expression in myeloid leukemia.

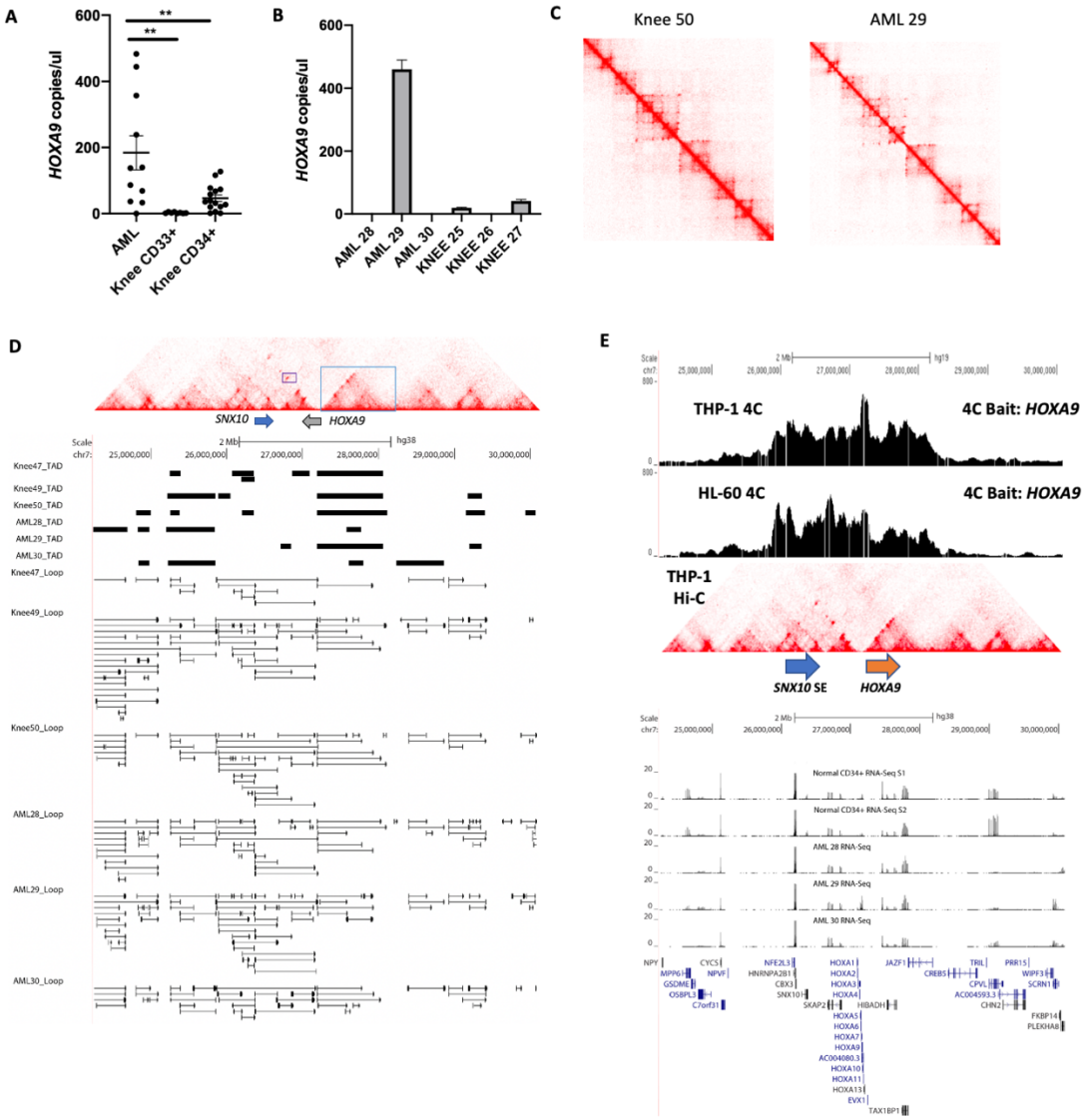


Figure 3. Hi-C analyses of CD34+ blood stem cells from Acute Myeloid Leukemia patient samples and normal knee bone marrow shows abundant loops and Topologically Associated Domains (TADs) around the *HOXA9* gene. **A.** Digital droplet (ddPCR) investigations of *HOXA9* gene expression level in AML, CD33+ myeloid cells from healthy donors, and CD34+ cells from healthy donors. Significance testing was performed by Student's two-tailed t-test. Two asterisks ** indicate $p<0.01$. **B.** ddPCR of *HOXA9* in CD34+ AML28, 29, 30 and Knee 25, 26 and 27 samples. Data shown indicates the average value of technical replicates performed on the same clinical sample and error bars indicate standard error. **C.** Heatmaps (coverage normalized) of the *HOXA9* genomic region in Knee50 (left) and AML29 (right) depicted have the X axis that list a set of genomic coordinates (in the case of the *HOXA9* region it is chr7:24,243,983-30,083,982). The same coordinates are then shown on the Y axis, and regions with strong interactions are shown in dark red. Loops are indicated as dots on the map. Distinct squares indicating regions with strong interactions can be seen in the heatmap, indicating TADs which are regions of the genome with strong interactions within the TAD boundaries. **D.** The heatmap shows the TAD of AML29 indicated in the heatmap in C, with the called TAD indicated below. The loop indicated in the heatmap in D is depicted (as a dot), and the loop depicted is marked by a purple box in the loop screenshot. The purple dotted lines indicate the genomic coordinates of the loop. The tracks corresponding to loops and TADs are shown via the UCSC genome browser. The genomic locations of the *SNX10* and *HOXA9* genes in this area are indicated. A light blue box surrounds an example of a TAD on the heatmap which looks like a triangle, and a dark blue box surrounds an example of a loop on the heatmap. The loops are depicted as anchor – loop – anchor, where the thick black box indicates an anchor, meaning the base of a loop which physically interacts with the other anchor in the loop. The region that is looped out and away is indicated by a thin black connecting line. **E.** 4C at the *HOXA9* viewpoint in HL-60 and THP-1 cells. Hi-C heatmap of TADs shown in THP-1. RNA-seq of two normal CD34+ samples and three CD34+ AML samples were aligned with 4C region of *HOXA9* and the *SNX10* SE (genomic coordinates - chr7:24,243,983-30,083,982)

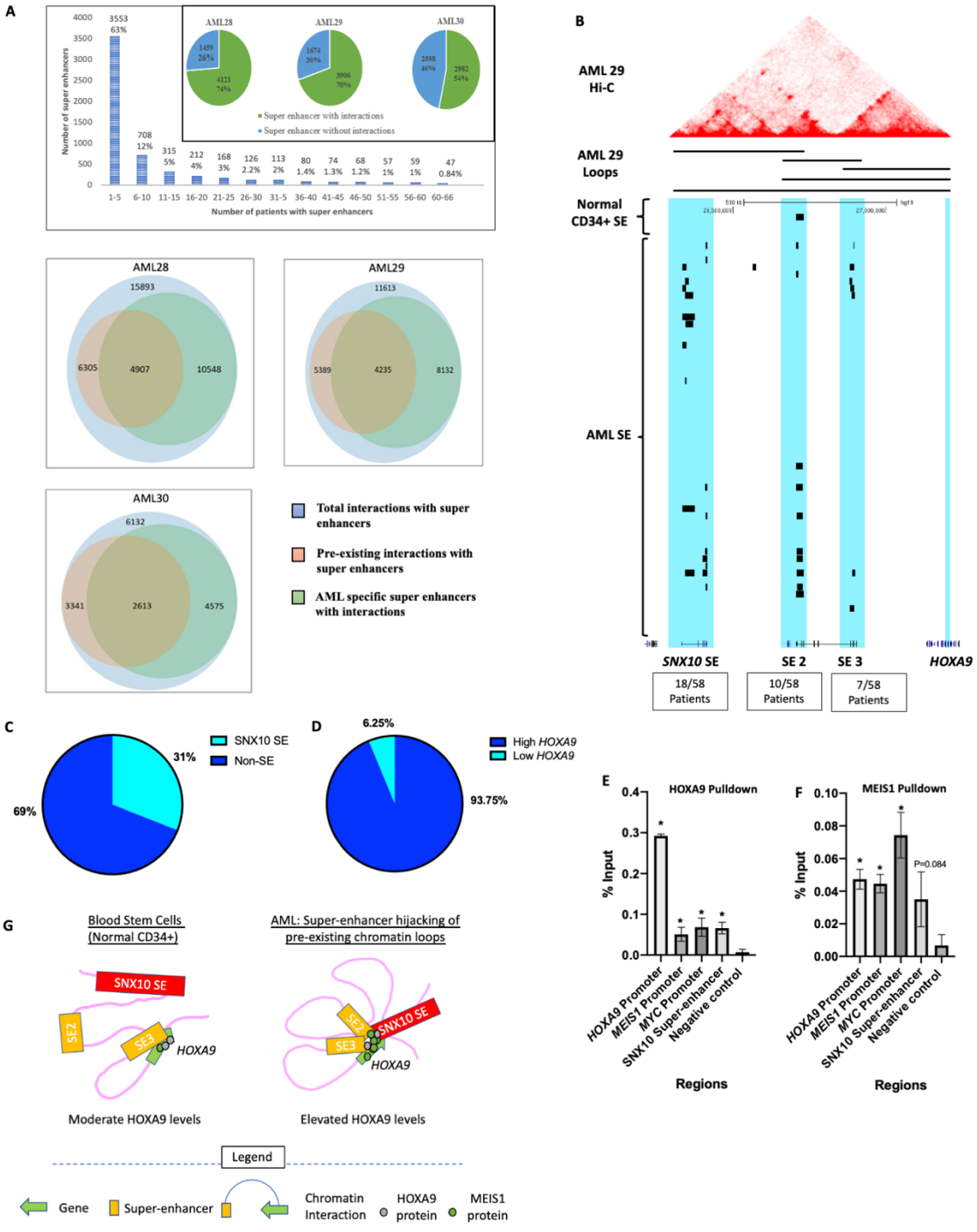


Figure 4. An AML-specific SE hijacks pre-existing chromatin interactions between *SNX10* gene and *HOXA9* and is associated with high *HOXA9* levels. A. Chart depicting the number of SEs found in 66 AML patients. Individual boxes indicate the SE profiles of three AML samples. Pie chart shows the number of pre-existing SE associated chromatin loops in AML. **B.** Loops between *SNX10* and *HOXA9* are shown together with SEs from two normal CD34+ donors and 66 AML patients (each row represents one patient), and the black bars indicate SEs. Regions with many SEs found in many patients are indicated by blue stripes, as well as the *HOXA9* promoter. Genomic coordinates for region Chr7:25,906,537-27,652,334. **C.** Pie chart showing the proportion of AML patients from the 66 AML patients who have the *SNX10* SE ("SNX10 SE") or not ("no SE"). **D.** Pie chart showing the proportion of AML patients with the *SNX10* SE with high *HOXA9* as compared with low *HOXA9*. **E.** Chromatin-immunoprecipitation (ChIP)-qPCR from HOXA9 antibody pulldown at selected oncogenes promoter regions. **F.** Chromatin-immunoprecipitation (ChIP)-qPCR from MEIS1 antibody pulldown at selected oncogenes promoter regions. **G.** Schematic of the hypothesized mechanism of SEs hijacking pre-existing chromatin loops at the *HOXA9* locus.

References

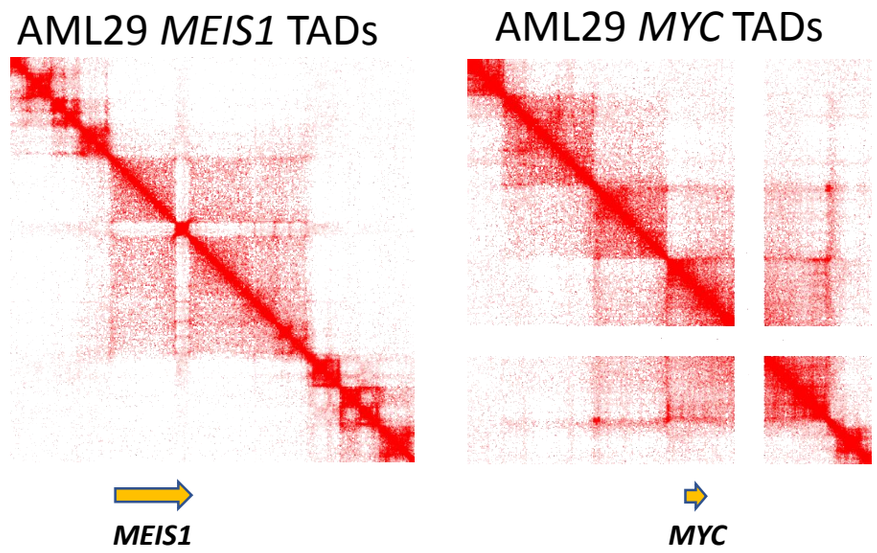
1. Estey E, Döhner H. Seminar Acute myeloid leukaemia. Lancet. 2006;368:1894-907.
2. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. New England Journal of Medicine. 2015;373:1136-52.
3. Ritchie EK, Feldman EJ, Christos PJ, Rohan SD, Lagassa CB, Ippoliti C, et al. Decitabine in patients with newly diagnosed and relapsed acute myeloid leukemia. Leukemia & lymphoma. 2013;54:2003-7.
4. Network CGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl j Med. 2013;2013:2059-74.
5. Wouters BJ, Delwel R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. Blood. 2016;127:42-52.
6. Gao L, Sun J, Liu F, Zhang H, Ma Y. Higher expression levels of the HOXA9 gene, closely associated with MLL-PTD and EZH2 mutations, predict inferior outcome in acute myeloid leukemia. OncoTargets and therapy. 2016;9:711.
7. Andreeff M, Ruvolo V, Gadgil S, Zeng C, Coombes K, Chen W, et al. HOX expression patterns identify a common signature for favorable AML. Leukemia. 2008;22:2041-7.
8. Collins C, Wang J, Miao H, Bronstein J, Nawer H, Xu T, et al. C/EBPα is an essential collaborator in Hoxa9/Meis1-mediated leukemogenesis. Proceedings of the National Academy of Sciences. 2014;111:9899-904.
9. Collins CT, Hess JL. Deregulation of the HOXA9/MEIS1 axis in acute leukemia. Current opinion in hematology. 2016;23:354.
10. Mohr S, Doebele C, Comoglio F, Berg T, Beck J, Bohnenberger H, et al. Hoxa9 and Meis1 cooperatively induce addiction to Syk signaling by suppressing miR-146a in acute myeloid leukemia. Cancer cell. 2017;31:549-62. e11.
11. Sun Y, Zhou B, Mao F, Xu J, Miao H, Zou Z, et al. HOXA9 reprograms the enhancer landscape to promote leukemogenesis. Cancer cell. 2018;34:643-58. e5.
12. Babu D, Fullwood MJ. 3D genome organization in health and disease: emerging opportunities in cancer translational medicine. Nucleus (Austin, Tex). 2015;6:382-93.
13. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155:934-47.
14. Cao F, Fang Y, Tan HK, Goh Y, Choy JYH, Koh BTH, et al. Super-Enhancers and Broad H3K4me3 Domains Form Complex Gene Regulatory Circuits Involving Chromatin Interactions. Sci Rep. 2017;7:2186.
15. Taberlay PC, Achinger-Kawecka J, Lun AT, Buske FA, Sabir K, Gould CM, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome research. 2016;26:719-31.
16. Zhang Y, Yang L, Kucherlapati M, Chen F, Hadjipanayis A, Pantazi A, et al. A pan-cancer compendium of genes deregulated by somatic genomic rearrangement across more than 1,400 cases. Cell reports. 2018;24:515-27.
17. Akıncılar SC, Khattar E, Boon PLS, Unal B, Fullwood MJ, Tergaonkar V. Long-range chromatin interactions drive mutant TERT promoter activation. Cancer discovery. 2016;6:1276-91.
18. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 2016;17:2042-59.
19. McKeown MR, Corces MR, Eaton ML, Fiore C, Lee E, Lopez JT, et al. Superenhancer Analysis Defines Novel Epigenomic Subtypes of Non-APL AML, Including an RARalpha Dependency Targetable by SY-1425, a Potent and Selective RARalpha Agonist. Cancer Discov. 2017;7:1136-53.
20. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016;3:95-8.
21. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 2016;3:99-101.
22. Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biol. 2017;18:125.
23. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57-63.
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.
25. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665-80.
26. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic acids research. 2013;41:e132-e.

27. An Ö, Tan K-T, Li Y, Li J, Wu C-S, Zhang B, et al. CSI NGS Portal: An online platform for automated NGS data analysis and sharing. 2019.
28. Kumar AR, Sarver AL, Wu B, Kersey JH. Meis1 maintains stemness signature in MLL-AF9 leukemia. Blood. 2010;115:3642-3.
29. Garcia-Cuellar MP, Steger J, Fuller E, Hetzner K, Slany RK. Pbx3 and Meis1 cooperate through multiple mechanisms to support Hox-induced murine leukemia. Haematologica. 2015;100:905-13.
30. Li Z, Chen P, Su R, Hu C, Li Y, Elkahloun AG, et al. PBX3 and MEIS1 Cooperate in Hematopoietic Cells to Drive Acute Myeloid Leukemias Characterized by a Core Transcriptome of the MLL-Rearranged Disease. Cancer Res. 2016;76:619-29.
31. Bessa J, Tavares MJ, Santos J, Kikuta H, Laplante M, Becker TS, et al. meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. Development. 2008;135:799-803.
32. Sun W, Zhang K, Zhang X, Lei W, Xiao T, Ma J, et al. Identification of differentially expressed genes in human lung squamous cell carcinoma using suppression subtractive hybridization. Cancer Letters. 2004;212:83-93.
33. Mujahed H, Miliara S, Neddermeyer AH, Bengtzen S, Nilsson C, Deneberg S, et al. AML Displays Increased CTCF Occupancy Associated to Aberrant Gene Expression and Transcription Factor Binding. Blood. 2020.
34. Kloetgen A, Thandapani P, Ntziachristos P, Ghebrechristos Y, Nomikou S, Lazaris C, et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. Nat Genet. 2020;52:388-400.
35. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods. 2012;58:221-30.
36. Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Scientific reports. 2014;4:1-6.
37. Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, et al. Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. Mol Cell. 2017;67:1037-48 e6.
38. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57-74.
39. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018;46:D794-d801.

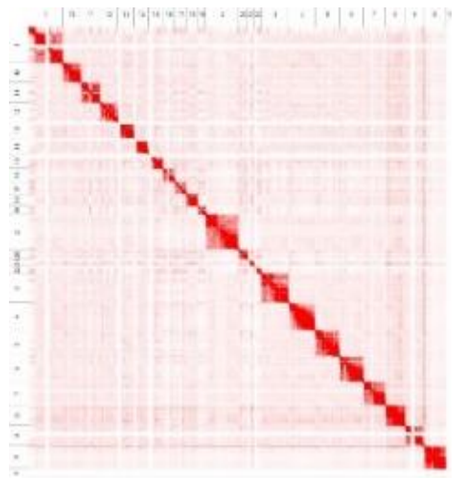
A

Sample	Reads	#TADs	#loops
AML28	1,290,109,443	1,682	24,394
AML29	1,245,251,169	1,108	19,541
AML30	1,387,562,168	1,296	10,733
Knee 47	1,337,973,222	1,153	4,733
Knee49	1,156,332,234	927	13,107
Knee50	1,165,296,399	1,234	13,795

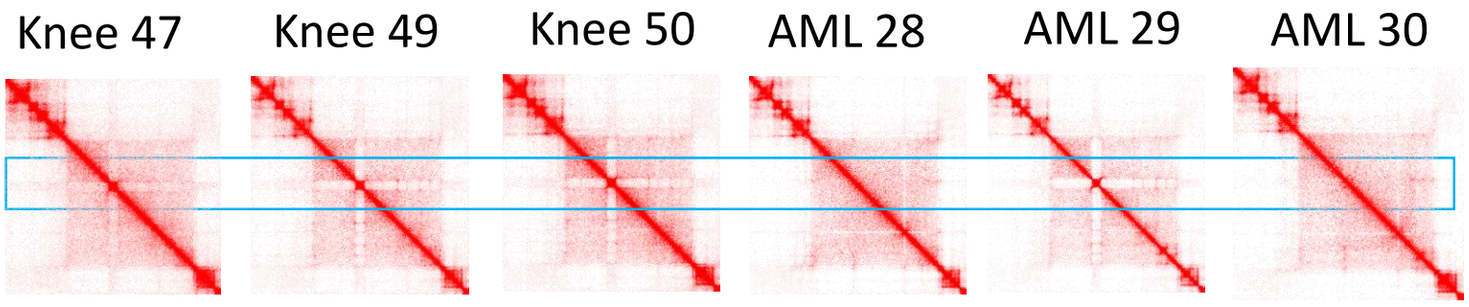
B



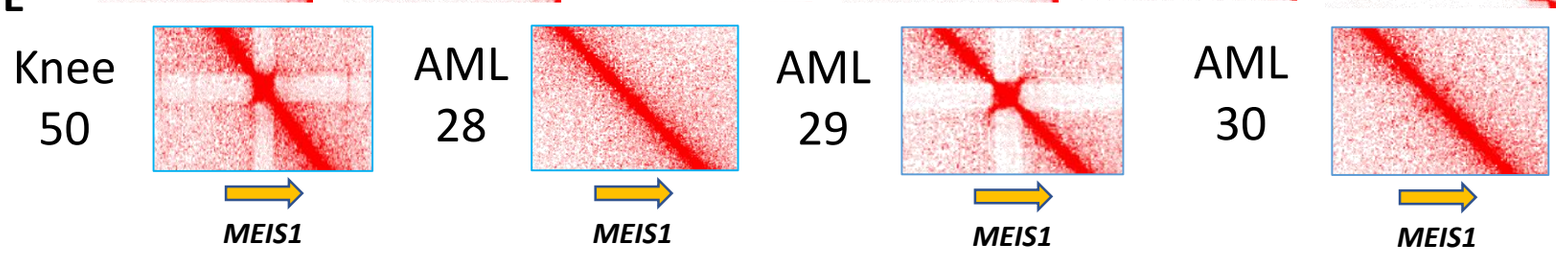
C



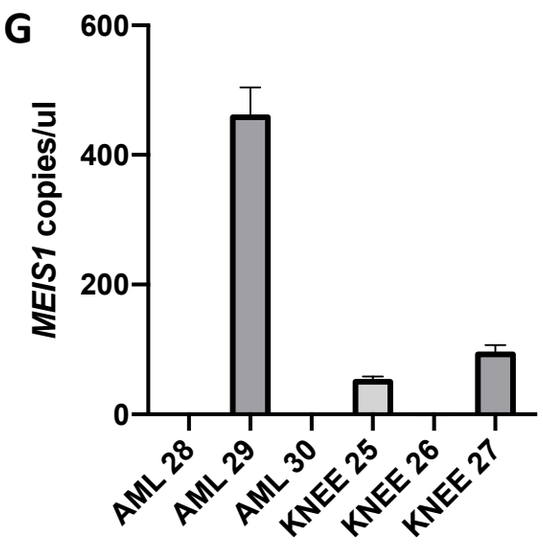
D



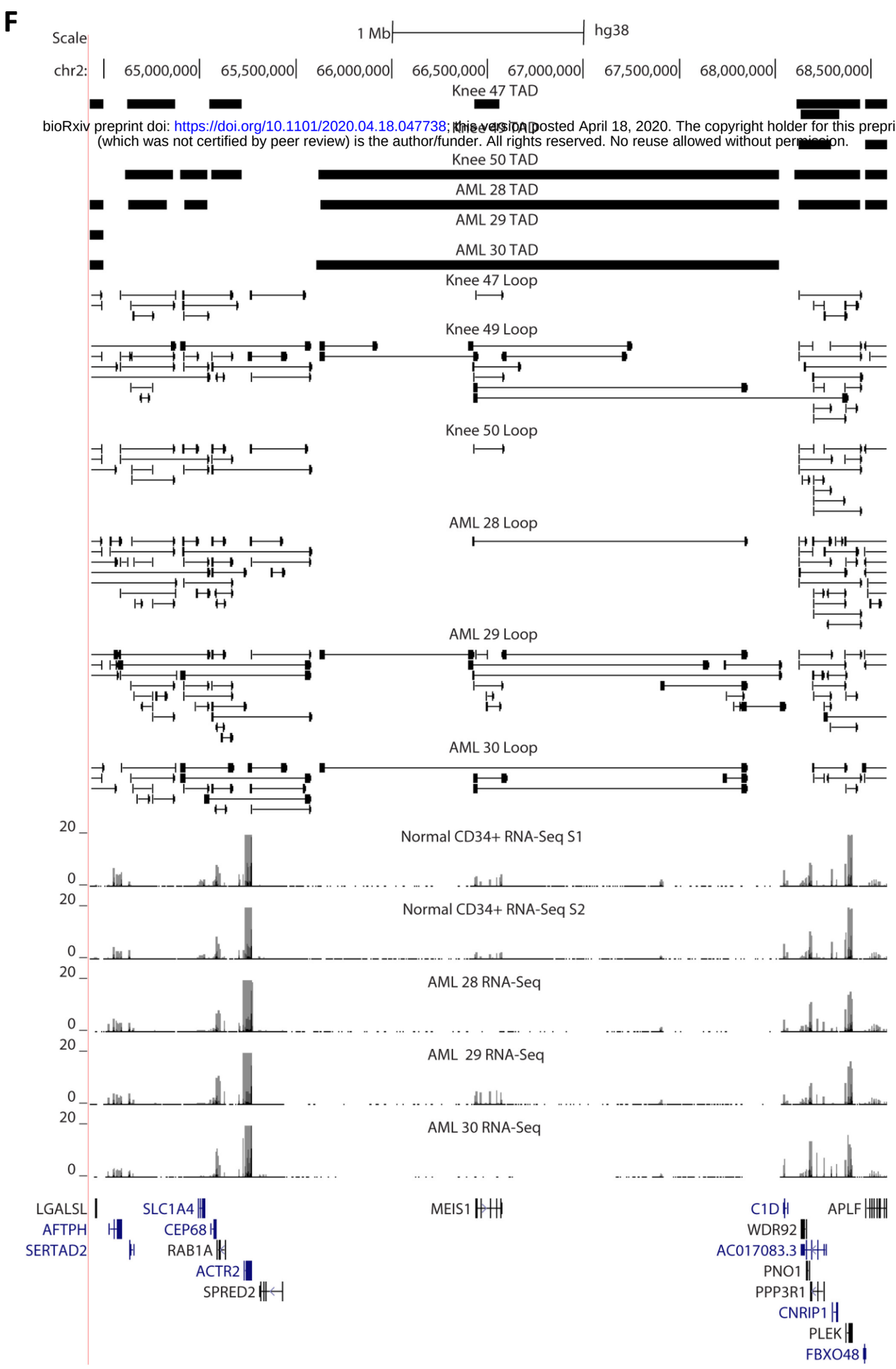
E



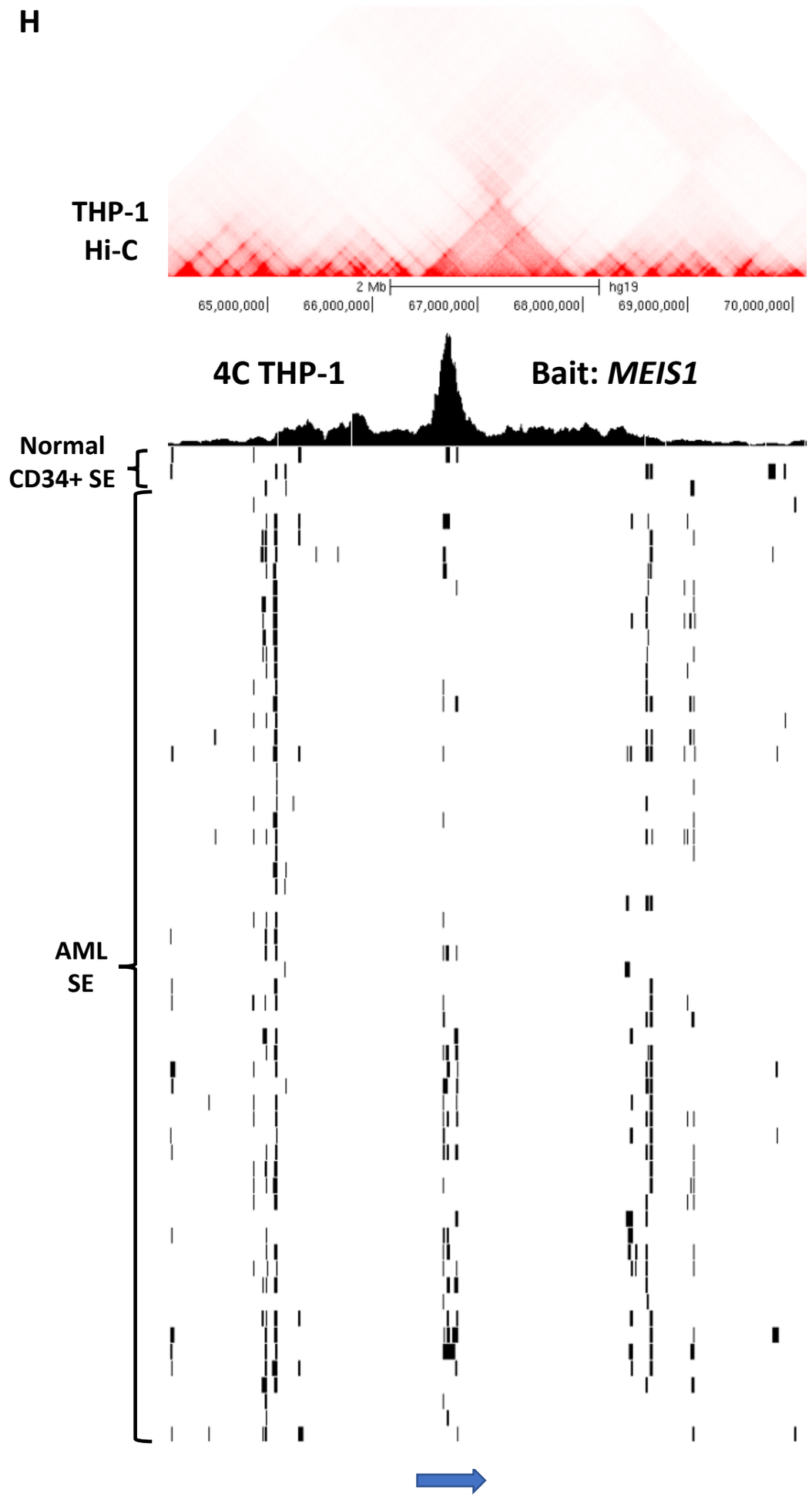
G

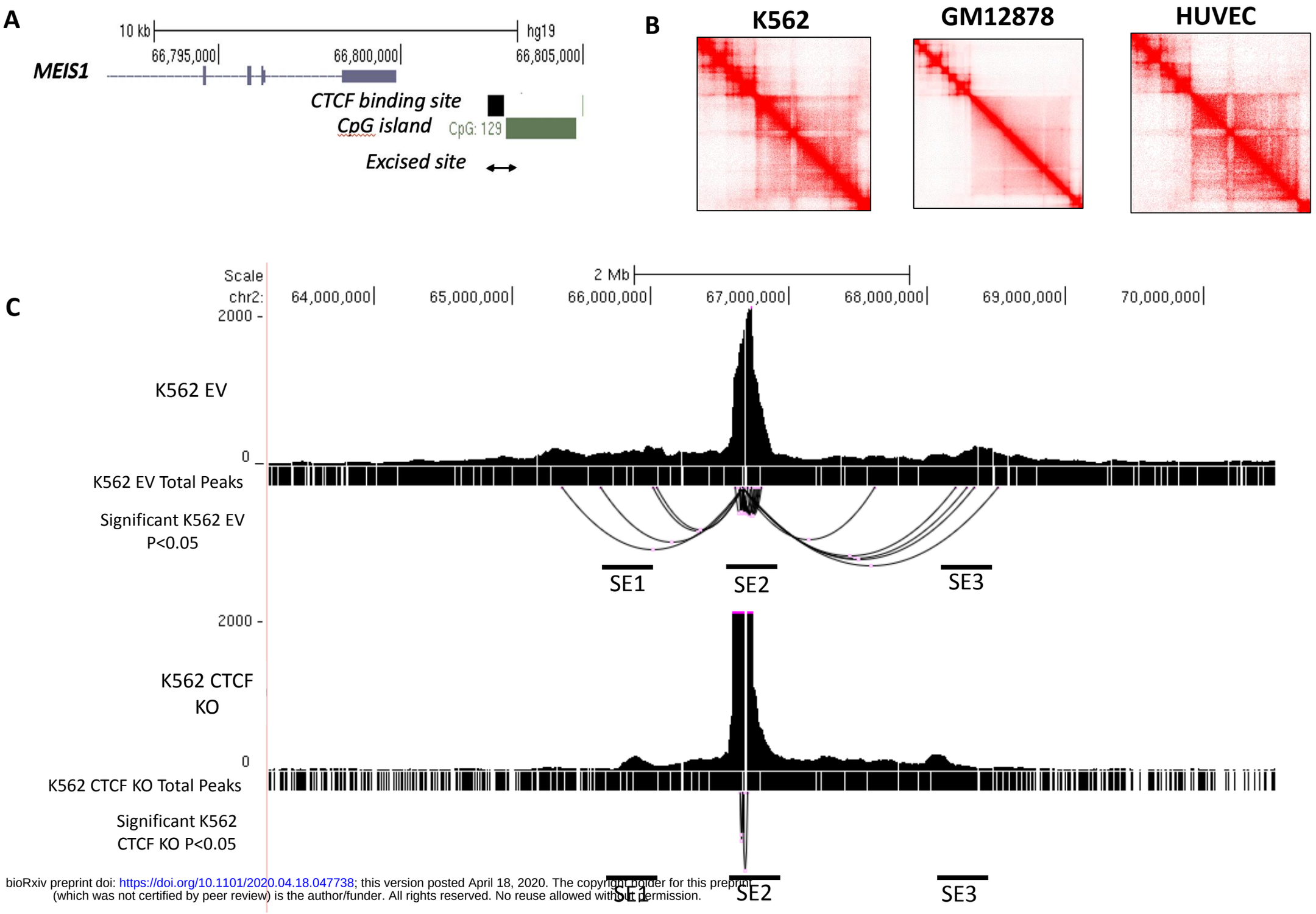


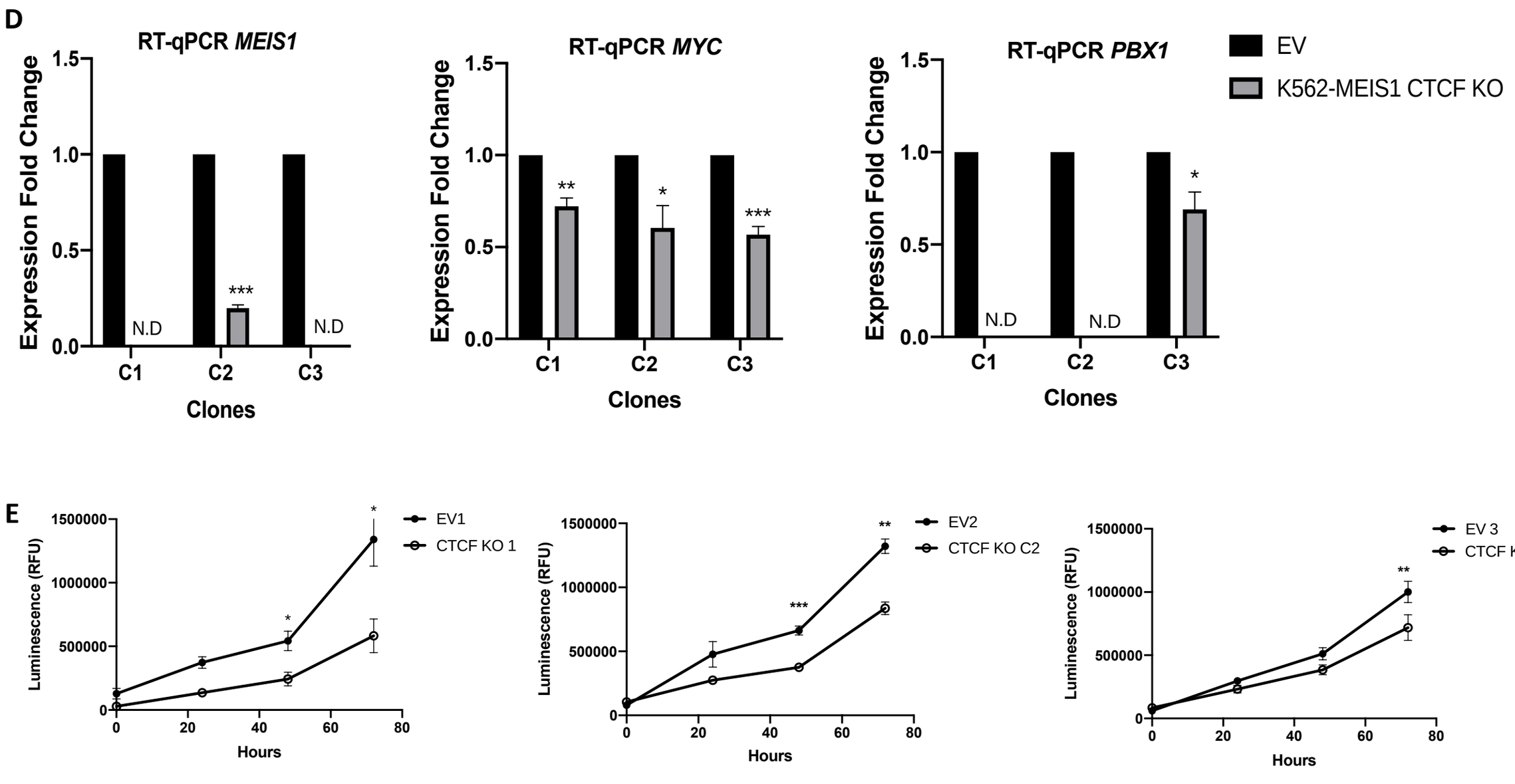
F

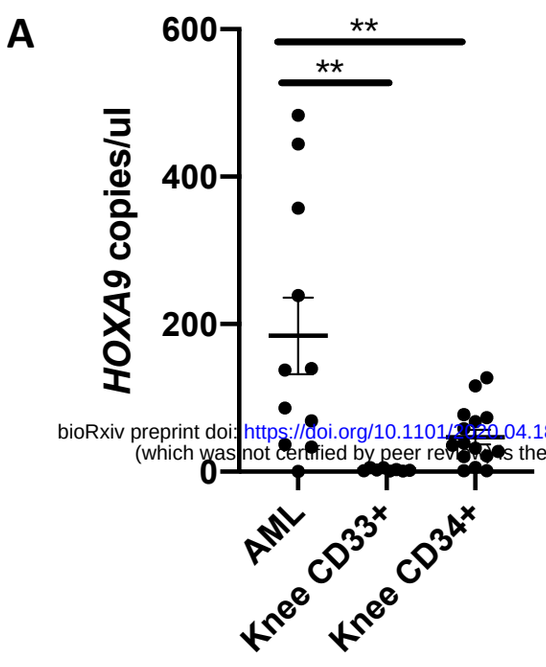


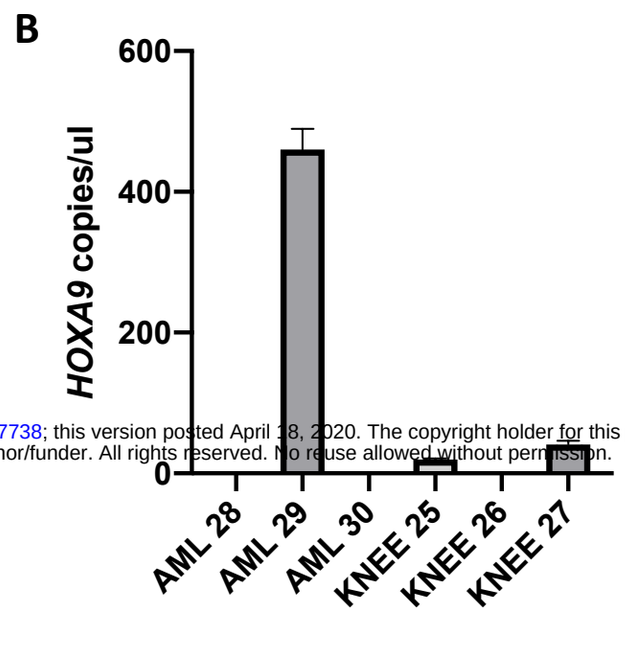
H

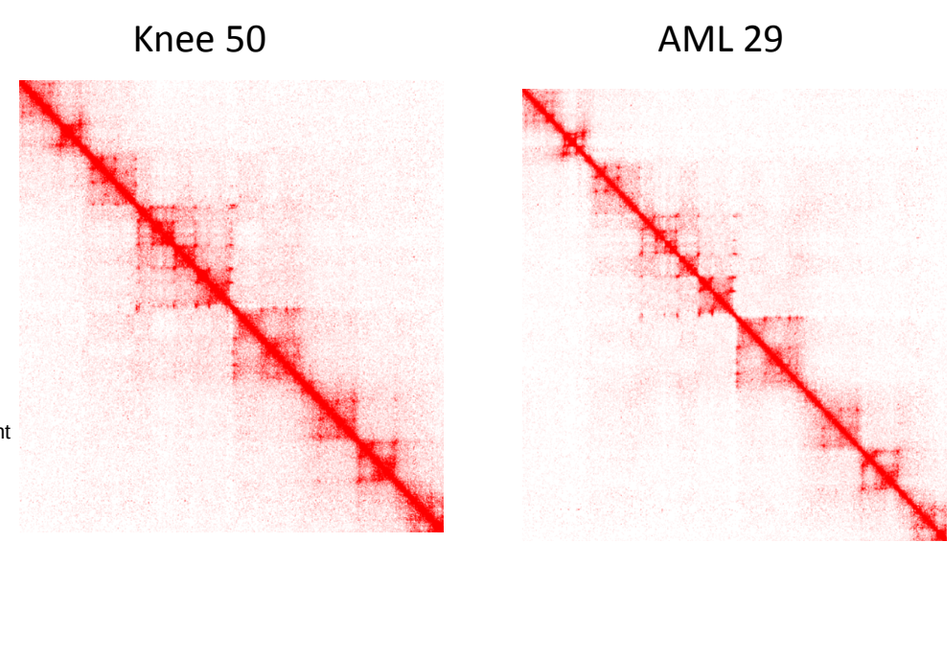


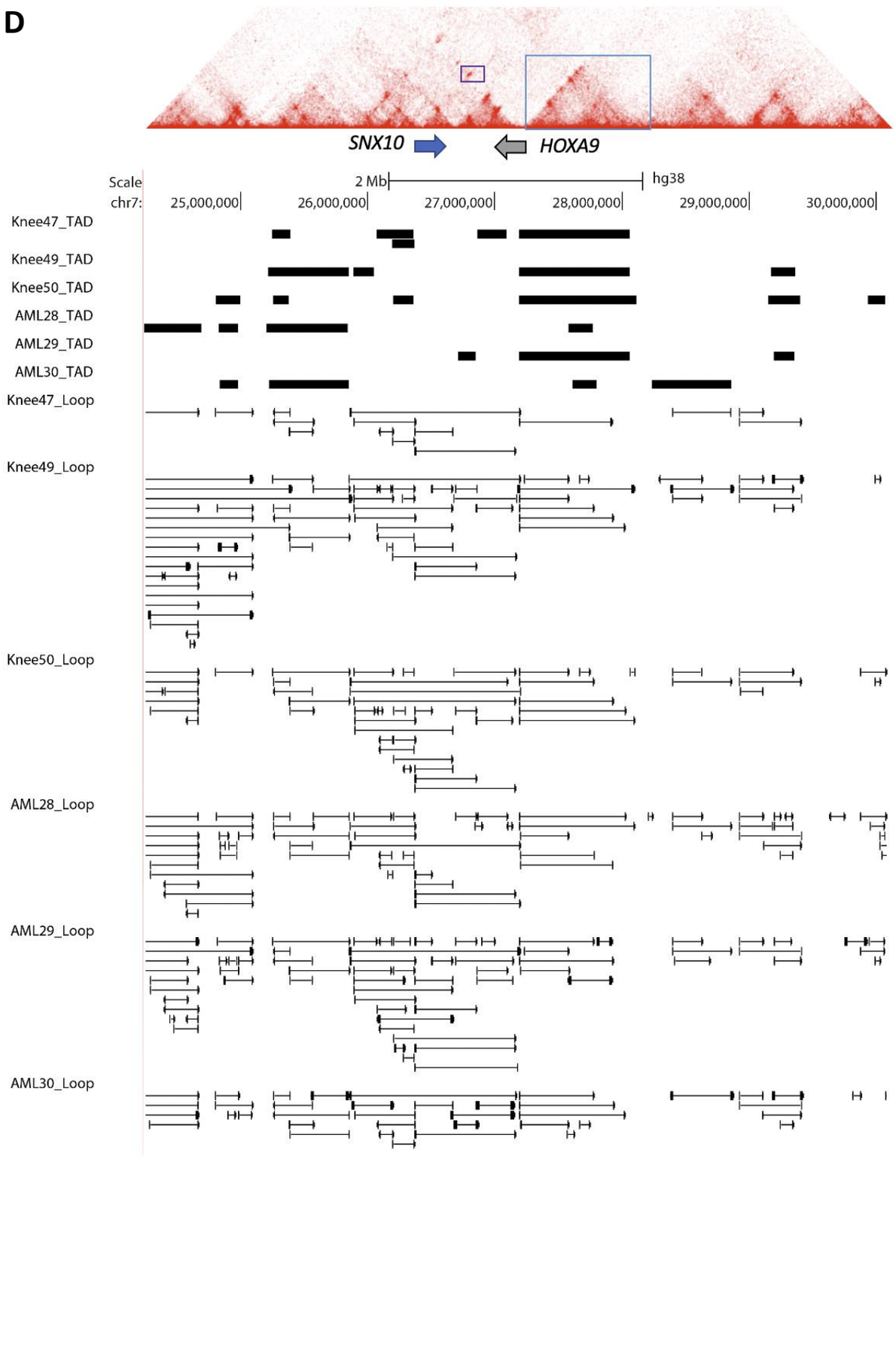


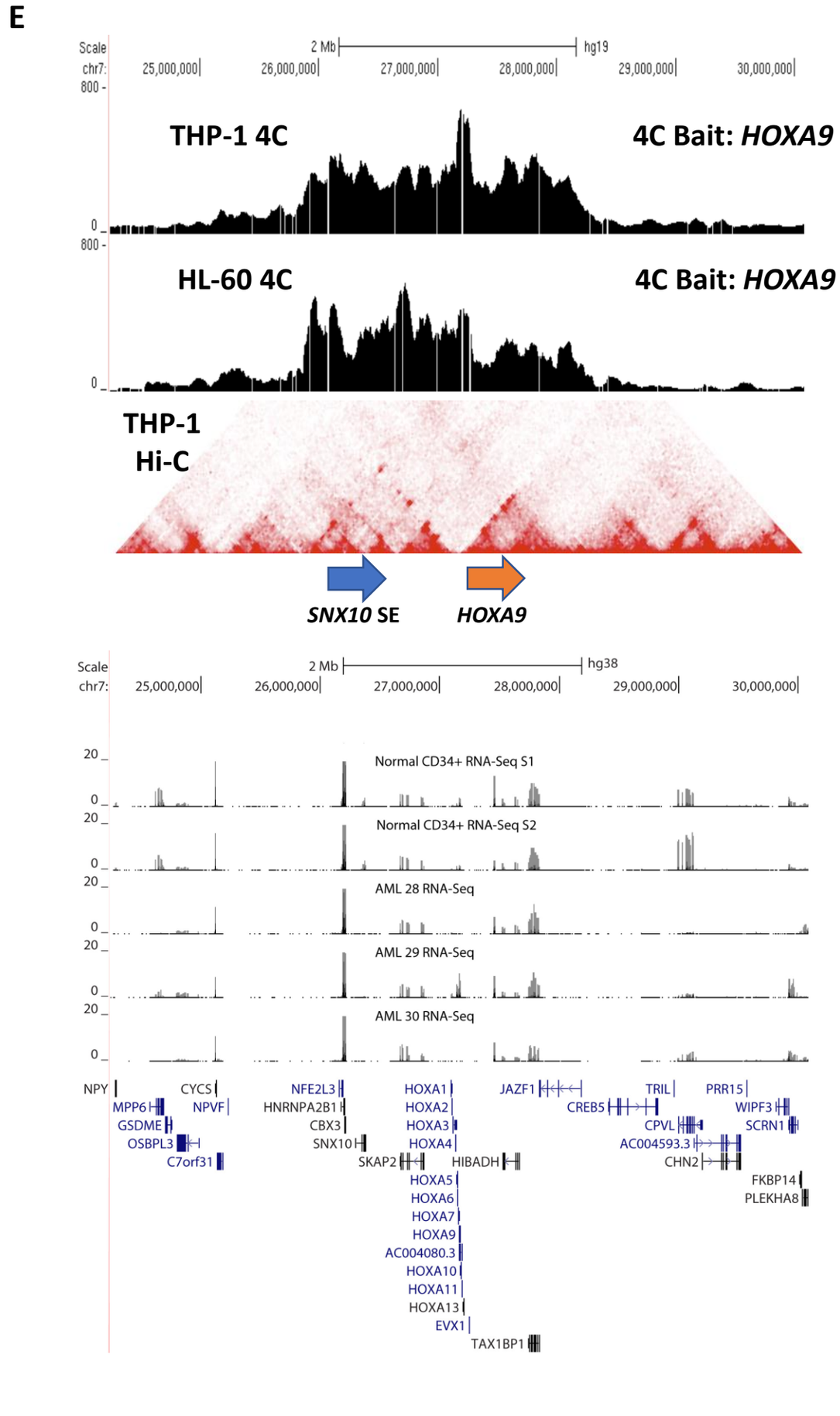


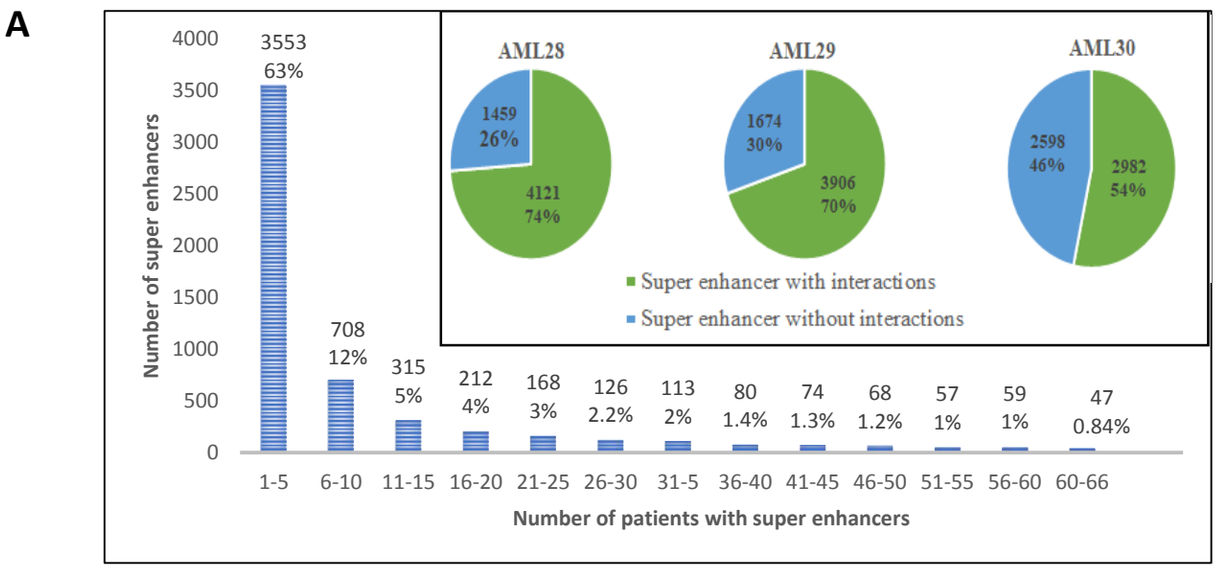


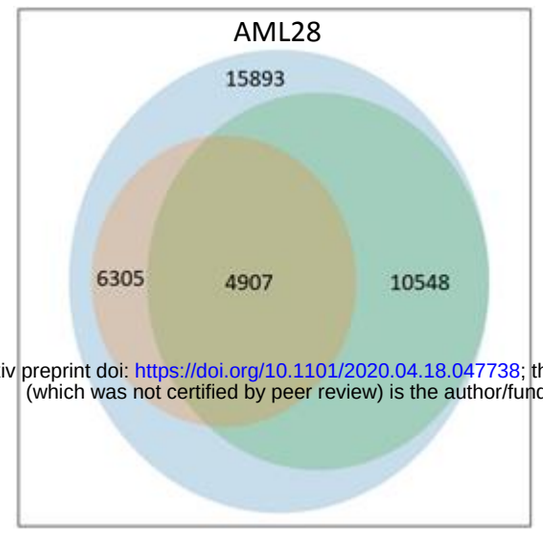


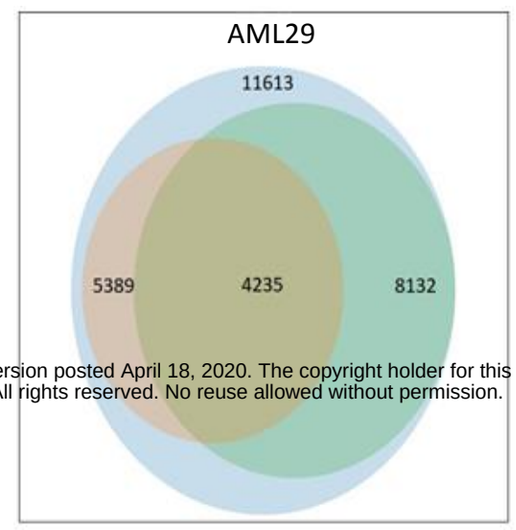


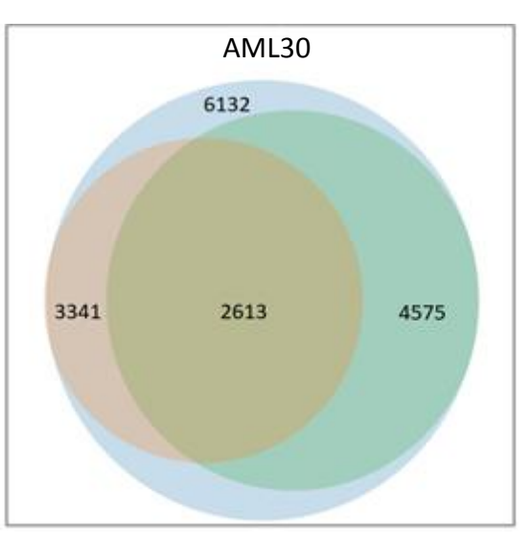












Total interactions with super enhancers

Pre-existing interactions with super enhancers

AML specific super enhancers with interactions