Identification of physical interactions between genomic regions by enChIP-Seq

Toshitsugu Fujita, Miyuki Yuno, Yutaka Suzuki, Sumio Sugano, and Hodaka Fujii*

Toshitsugu Fujita, Miyuki Yuno, Hodaka Fujii
Chromatin Biochemistry Research Group, Combined Program on Microbiology and Immunology,
Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan

Yutaka Suzuki, Sumio Sugano
Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University
of Tokyo, Kashiwa, Chiba, Japan

Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University
of Tokyo, Minato-ku, Tokyo, Japan

*Corresponding author
E-mail: hodaka@biken.osaka-u.ac.jp
ABSTRACT

Physical interactions between genomic regions play critical roles in the regulation of genome functions, including gene expression. However, the methods for confidently detecting physical interactions between genomic regions remain limited. Here, we demonstrate the feasibility of using engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) in combination with next-generation sequencing (NGS) (enChIP-Seq) to detect such interactions. In enChIP-Seq, the target genomic region is captured by an engineered DNA-binding complex, such as a CRISPR system consisting of a catalytically inactive form of Cas9 (dCas9) and a single guide RNA (sgRNA). Subsequently, the genomic regions that physically interact with the target genomic region in the captured complex are sequenced by NGS. Using enChIP-Seq, we found that the 5’HS5 locus, which regulates expression of the β-globin genes, interacts with multiple genomic regions upon erythroid differentiation in the human erythroleukemia cell line K562. Genes near the genomic regions inducibly associated with the 5’HS5 locus were transcriptionally up-regulated in the differentiated state, suggesting the existence of a coordinated transcription mechanism directly or indirectly mediated by physical interactions between these loci. Our data suggest that enChIP-Seq is a potentially useful tool for detecting physical interactions between genomic regions in a non-biased manner, which would facilitate elucidation of the molecular mechanisms underlying regulation of genome functions.
INTRODUCTION

Physical interactions between genomic regions play important roles in the regulation of genome functions, including transcription and epigenetic regulation [1]. Several techniques, such as fluorescence in situ hybridization (FISH) [2, 3], chromosome conformation capture (3C), and 3C-derived methods [4, 5], have been used to detect such interactions. Although these techniques are widely used, they have certain limitations. The resolution of FISH is low, i.e., apparent co-localization of FISH signals does not necessarily mean that the loci in question physically interact. In addition, FISH cannot be used in a non-biased search for interacting genomic regions.

In 3C and related methods, molecular interactions are maintained by crosslinking with formaldehyde prior to digestion with a restriction enzyme(s). The digested DNA is purified after ligation of the DNA ends within the same complex. Interaction between genomic loci is detected by PCR using locus-specific primers or next-generation sequencing (NGS). As with FISH, 3C-based approaches also have intrinsic drawbacks. For example, these methods require enzymatic reactions, including digestion with restriction enzyme(s) and ligation of crosslinked chromatin; the difficulty of achieving complete digestion of crosslinked chromatin can result in detection of artifactual interactions. In addition, in 3C and its derivatives, it is difficult to distinguish the products of intra-molecular and inter-molecular ligation reactions, making it difficult to ensure that the detected signals truly reflect physical interactions between different genomic regions.

An alternative approach to detecting physical interactions between genomic regions is to purify specific genomic regions engaged in molecular interactions and then analyze the genomic DNA in the purified complexes. To purify specific genomic regions, we recently developed two
locus-specific chromatin immunoprecipitation (locus-specific ChIP) technologies, insertional ChIP (iChIP) [6-10] (see review [11]) and engineered DNA-binding molecule-mediated ChIP (enChIP) [12-16] (see reviews [17, 18]). enChIP consists of the following steps (Figure 1): (i) A DNA-binding molecule or complex (DB) that recognizes a target DNA sequence in a genomic region of interest is engineered. Zinc-finger proteins [19], transcription activator-like (TAL) proteins [20], and a clustered regularly interspaced short palindromic repeats (CRISPR) system [21] consisting of a catalytically inactive Cas9 (dCas9) plus a single guide RNA (sgRNA) can be used as the DB. Tag(s) and a nuclear localization signal (NLS)(s) can be fused with the engineered DB, and the fusion protein(s) can be expressed in the cells of interest. (ii) If necessary, the DB-expressing cells are stimulated and crosslinked with formaldehyde or other crosslinkers. (iii) The cells are lysed, and chromatin is fragmented by sonication or digested with nucleases. (iv) Chromatin complexes containing the engineered DB are affinity-purified by immunoprecipitation or other methods. (v) After reverse crosslinking (if necessary), DNA, RNA, proteins, or other molecules are purified and identified by various methods including NGS and mass spectrometry.

As a model locus in this study, we focused on 5′HS, which plays critical roles in developmentally regulated expression of the β-globin genes and has been extensively analyzed (Figure 2A) [22, 23]. The 5′HS2-4 regions in the 5′HS locus behave as enhancers for β-globin expression [24-26]. By contrast, 5′HS5 functions as an insulator to prevent invasion of heterochromatin into the β-globin genes [27]. In addition, the 5′HS5 locus interacts with the 3′HS1 locus in the 3′ region of the globin locus [28, 29]. Moreover, CTCF, a major component of the insulator complex, plays a critical role in insulation and formation of a chromatin loop [27, 29]. However, the molecular mechanisms underlying the functions of 5′HS5 remain incompletely understood.
Here, we combined enChIP with NGS (enChIP-Seq) to detect genomic regions that physically interact with the 5’HS5 locus. Using enChIP-Seq, we showed that the 5’HS5 locus physically interacts with multiple genomic regions upon erythroid differentiation in the human erythroleukemia cell line K562. Thus, enChIP-Seq represents a potentially useful tool for analysis of genome functions.
RESULTS AND DISCUSSION

Isolation of the 5’HS5 locus by enChIP using the CRISPR system

To purify the 5’HS5 locus by enChIP using the CRISPR system (Figures 1 and 2A) [12, 14], we generated human erythroleukemia K562-derived cells expressing 3xFLAG-dCas9 [14] and sgRNA targeting the 5’HS5 locus (Figure 2A). We designed two sgRNAs (#6 and #17) to target different sites in the 5’HS5 locus separated by 52 bp (Figure 2B). Like the parental K562 cells, the derivative cells were white, suggesting that they did not spontaneously express globin genes as a result of introduction of the CRISPR complex or its binding to the target locus. Crosslinked chromatin was fragmented by sonication and subjected to affinity purification with anti-FLAG antibody (Ab) (Figure 1B). Subsequently, crosslinking was reversed and DNA was purified from the isolated chromatin. As shown in Figure 2C, 0.1-0.4% of the input 5’HS5 locus was isolated, whereas the irrelevant Sox2 locus was not enriched, suggesting that enChIP using either of the two sgRNAs (#6 and #17) could isolate the 5’HS5 locus.

Detection of genomic regions that physically interact with the 5’HS5 locus

To identify genomic regions associated with the 5’HS5 locus on a genome-wide scale in erythroid cells under undifferentiated or differentiated conditions, K562-derived cells were mock-treated or treated with sodium butyrate (NaB) for 4 days before being crosslinked with formaldehyde and subjected to enChIP. The K562-derived cells changed from white to pink upon NaB treatment, suggesting that they had begun to express the globin genes, as wild-type K562 cells do in response to NaB. After isolating the 5’HS5 locus by enChIP, we subjected the purified DNA to NGS analysis. As expected, reads corresponding to the 5’HS5 locus were clearly
detected in cells expressing either sgRNA #6 or #17, but not in cells expressing neither sgRNA
(Figure 2D and Table 1). By contrast, no peak was detected at the irrelevant Sox2 locus (Figure
2E).

To identify genomic regions physically interacting with 5’HS5 upon erythroid differentiation, we
analyzed the NGS peaks observed in the differentiated state. The CRISPR complex can interact
with multiple genomic sites containing sequences similar to the sgRNA sequence [30-33]. In
addition, in the absence of sgRNA, it is possible that dCas9 could bind non-specifically to some
genomic sites in vivo. Therefore, the peaks identified for sgRNA #6 and #17 may include such
off-target sites. To remove those off-target sites, we first eliminated peaks derived from
non-specific binding of dCas9 in the absence of sgRNA from those identified for each sgRNA
(Steps 1 and 2 in Figure 3). To identify peaks with confidence, we next established two criteria
for choosing peaks based on NGS information from the target 5’HS5 locus: (1) tag number >5%
of that of the target 5’HS5 locus, and (2) fold enrichment relative to input genomic DNA >10. As
shown in Figure 3 (Step 2), 19 and 228 peaks for sgRNA #6 and #17, respectively, fulfilled these
criteria. Next, to eliminate sgRNA-dependent off-target sites, we compared the peaks for sgRNA
#6 and #17 and selected peaks detected in common by both sgRNA #6 and #17. These peaks
were considered to represent regions engaged in bona fide physical interactions with the 5’HS5
locus (Step 3 in Figure 3). The six identified peaks could be classified into two categories: (1)
peaks that were larger in the differentiated state, and (2) peaks constitutively observed in both the
undifferentiated and differentiated states. The first category should contain genomic regions that
inducibly associate with the 5’HS5 locus upon erythroid differentiation, whereas the second
category should contain genomic regions constitutively associated with the 5’HS5 locus. To
extract the peaks that grew larger specifically in the differentiated state, we selected peaks
observed constitutively or in the undifferentiated state (Step 4 in Figure 3) and compared them
with the six peaks extracted in Step 3 (Step 5 in Figure 3). As shown in Figure 3 (Step 5) and
Table 2, the 5’HS5 site was the unique peak constitutively observed in the undifferentiated and
differentiated states, whereas the five other peaks were larger specifically in the differentiated
state. These included one intra-chromosomal interaction and four inter-chromosomal interactions
(Table 2); the two peaks on chromosome 1 corresponding to inter-chromosomal interactions were
adjacent to each other in the primary sequence. Next, we attempted to extract genomic regions
that interacted with the 5’HS5 locus specifically in the undifferentiated state. However,
bioinformatics analysis based on the aforementioned criteria identified no regions in this category
(Figure S1).

We visualized some of the identified peaks in the UCSC Genome Browser (Figure 4). The peaks
were clearly visible for both sgRNA #6 and #17. The adjacent NaB-specific peaks in
chromosome 1 were both located in the first intron of the ZFN670 and ZFN670-695 genes. The
other NaB-specific peaks were located in the vicinity of the MIR422A gene in chromosome 15
and between the TMEM151A and YIF1A genes in chromosome 11.

To confirm the interactions identified by enChIP-Seq, we used the ligation-mediated approach
used in 3C-based assays (Figure S2). In this approach, cells are subjected to crosslinking, and
then chromatin is randomly fragmented by sonication. After proximity ligation of genomic DNA,
the junction between the target locus and a potential interacting site is amplified by PCR.
Subsequently, a part of the amplified region in the potential interacting locus is detected by a
second PCR. Amplification of a region in the second PCR suggests that the potential interacting
locus is physically proximal to the target locus. When we used this assay to examine the
interaction between the 5’HS5 locus and the ZNF670/ZFN670-695 locus, the second PCR amplified the ZNF670/ZFN670-695 locus in a NaB-specific manner only when the proximal ligation step was performed (Figure 5). This observation is consistent with the enChIP-Seq result (Figure 4) showing that the 5’HS5 locus interacts with the ZNF670/ZFN670-695 locus in the differentiated state in K562 cells. Thus, we were able to confirm the chromosomal interaction identified by enChIP-Seq by another independent method, suggesting that it is feasible to use enChIP-Seq to perform non-biased identification of physical interactions between genomic regions.

Transcription of genes near the 5’HS5-interacting genomic regions identified by enChIP-Seq could be directly or indirectly regulated by the induced association with the 5’HS locus. Such regulation could involve the 5’HS2-4 regions, which function as enhancers [24-26]. Therefore, we investigated whether mRNA levels of the genes in the vicinity (±10 kbp) of the 5’HS5-interacting genomic regions changed after NaB treatment. As shown in Figure 6, mRNA levels of the ZNF670, MIR422A, and CNIH2 genes were clearly up-regulated in the NaB-induced differentiated state, suggesting that the identified chromosomal interactions are involved in transcriptional regulation of these genes. At this time, it is not clear whether these gene products play any roles in erythroid development. Future studies should attempt to elucidate how the 5’HS locus regulates transcription of these genes. It is possible that the enhancer function of the 5’HS locus directly activates transcription of these genes via interactions with their promoters under differentiated conditions. Alternatively, these loci may be incorporated into the “transcription factory” [34] upon erythroid differentiation, independent of the enhancer function of the 5’HS locus.
Lack of interaction between the 5’HS5 and β-globin loci

Studies using 3C and derived methods suggested that the 5’HS5 and β-globin loci interact [28, 29]. Therefore, we sought to detect a physical interaction between these loci by enChIP-Seq. As shown in Figure 3, bioinformatics analysis based on the criteria described above (fold enrichment: >10, tag number: >5% of that of the target positions) did not extract genomic regions around the β-globin locus. In fact, the peak images did not indicate any physical interactions between 5’HS5 and the β-globin locus or 3’HS1 (Figure S3). Importantly, in this regard, it is not likely that bound dCas9/sgRNA complexes abrogate the formation of chromosomal loops between 5’HS5 and the β-globin or 3’HS1 locus, because the target positions of the sgRNA do not overlap with the CTCF-binding site and the 5’HS5 core region (Figure 2), and the cells maintained their capability to differentiate in response to NaB.

Several phenomena might explain the discrepancy between the results of this study and those of 3C-derived methods regarding the interaction of 5’HS5 with the β-globin locus. First, 3C-derived methods might be much more sensitive than enChIP-Seq. Specifically, because 3C-derived methods use PCR amplification to detect ligated regions consisting of different genomic regions, they may be able to detect transient or weak interactions that did not pass the criteria we used in this study. Second, the fragmentation of chromatin by sonication in enChIP-Seq may be too harsh to retain weak chromosomal interactions. By contrast, 3C-derived methods employ restriction digestion, which is much milder than sonication, to fragment chromatin. Third, physical interactions between genomic regions are likely to be regulated in a cell cycle-dependent manner, and it is possible that interactions between these regions may occur only in a certain phase of a cell cycle, making it difficult to detect by enChIP-Seq. Alternatively, our results raise the
possibility that the ‘interactions’ detected by 3C and its derivatives reflect accessibility of the loci
to the nucleases and ligases employed in these techniques, but do not necessarily reflect physical
interactions between genomic regions. In fact, discrepancies between results of 3C or its
derivatives and those of FISH have been suggested [35]. This possibility highlights the
importance of confirming chromosomal interactions by independent methods.

Managing potential contamination of off-target sites

In this study, we identified physical interactions between genomic regions based on signals
detected by enChIP-Seq. dCas9 can bind to multiple sites containing sequences similar to the
sgRNA sequence [30, 31]. To eliminate potential contamination of our findings by off-target sites,
we propose several strategies:

1) Carefully examine the sequences of the detected peaks and remove those containing
sequences similar to the target sequence.

2) Use different conditions or cell types. Signals specifically detected in one condition or cell
type should reflect true physical interactions between genomic regions.

3) Use multiple different sgRNAs. Because different sgRNA are unlikely to engage in off-target
binding at the same genomic regions, signals observed in common using different sgRNAs
should reflect true physical interactions between genomic regions. In addition, cells expressing
dCas9 without sgRNA should be used as a negative control to eliminate off-target sites associated
with dCas9 in the absence of sgRNA.

4) Use a sequential purification scheme. Cas9 orthologs derived from different bacterial species
recognize distinct proto-spacer adjacent motif (PAM) sequences and can be used for genome
Tagging of a given locus with dCas9s derived from different lineages and bearing distinct tags would make it feasible to sequentially purify the locus, minimizing contamination with off-target sites.

Using these techniques, we believe that we can effectively manage potential contamination of dCas9 off-target sites in enChIP analyses.
CONCLUSIONS

In this study, we used enChIP-Seq analysis to detect physical interactions between genomic regions. In K562-derived cells, the 5’HS5 locus physically interacted with multiple regions in the genome (Figure 4, Table 2). These interactions were induced by erythroid differentiation in response to NaB treatment (Figure 4, Table 2). Transcription of genes around the interacting genomic region was up-regulated in the differentiated state (Figure 6), suggesting a direct or indirect involvement of 5’HS enhancer activity in transcription of genes proximal to the interacting sites. Our results suggest that enChIP-Seq represents a potentially useful tool for performing non-biased searches for physical interactions between genomic regions, which would facilitate elucidation of the molecular mechanisms underlying regulation of genome functions.
MATERIALS AND METHODS

Plasmids
3xFLAG-dCas9/pMXs-puro (Addgene #51240) was described previously [14]. To construct vectors for expression of sgRNAs, two oligos for each sgRNA were annealed and extended using Phusion polymerase (New England Biolabs) to make 100 bp double-stranded DNA fragments, as described previously [12]. The nucleotide sequences were as follows: hHS5 #6,

5'-TTTCTTGCTTTATATCTTGTGGAAAGGACGAAACACCAGATTCATAGCAGACA
GCTA-3'

5'-GACTAGCCTTATTTTAACTTGCTATTTCAGCTCTAAAACCTAGCTGTCTGCTATGAA
TCC-3'; hHS5 #17,

5'-TTTCTTGCTTTATATCTTGTGGAAAGGACGAAACACCAGAAGATAGGGTAA
GAGAC-3'

5'-GACTAGCCTTATTTTAACTTGCTATTTCAGCTCTAAAACCTAGCTGTCTGCTATGAA
CCC-3'. Fragments were purified following agarose gel electrophoresis and subjected to Gibson assembly (New England Biolabs) with the linearized sgRNA cloning vector (Addgene #41824), a gift from George Church [37], to yield sgRNA-hHS5 #6 and sgRNA-hHS5 #17. The gBlocks were excised with XhoI and HindIII and cloned into XhoI/HindIII-cleaved pSIR vector to generate self-inactivating retroviral vectors for sgRNAs, as described previously [14].

Cell culture
K562-derived cells [14] were maintained in RPMI (Wako) supplemented with 10% fetal calf serum (FCS).
Establishment of cells stably expressing 3xFLAG-dCas9 and sgRNA

Establishment of K562-derived cells expressing 3xFLAG-dCas9 was described previously [14]. To establish cells expressing both 3xFLAG-dCas9 and sgRNAs targeting the 5’HS5 locus, 2 µg of sgRNA-hHS5 #6/pSIR or sgRNA-hHS5 #17/pSIR was transfected along with 2 µg of pPAM3 into 1 × 10^6 293T cells [38]. Two days after transfection, K562-derived cells expressing 3xFLAG-dCas9 were infected with the supernatant (5 ml) of 293T cells containing the virus particles. K562-derived cells expressing both 3xFLAG-dCas9 and sgRNA-hHS5 #6 or sgRNA-hHS5 #17 were selected in RPMI medium containing 10% FCS, puromycin (0.5 µg/ml), and G418 (0.8 mg/ml).

Induction of differentiation of K562-derived cells

To induce erythroid differentiation of the K562-derived cells, cells were incubated in the presence of 1 mM NaB for 4 days.

enChIP-real-time PCR

enChIP-real-time PCR was performed as previously described [12], except that ChIP DNA Clean & Concentrator (Zymo Research) was used for purification of DNA. Primers used in the analysis are shown in Table S3.

enChIP-Seq and bioinformatics analysis

Undifferentiated or differentiated K562-derived cells (2 × 10^7 each) expressing 3xFLAG-dCas9 and sgRNAs were subjected to the enChIP procedure as described previously [12], except that ChIP DNA Clean & Concentrator was used for purification of DNA. NGS and data analysis were
performed at the University of Tokyo as described previously [39, 40]. Additional data analysis for Step 2 in Figure 3 was performed at Hokkaido System Science Co., Ltd. Images of NGS peaks were generated using the UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgGateway).

**Proximity ligation assay to confirm interactions between genomic regions**

K562 cells (1 × 10^7) were fixed with 1% formaldehyde at 37°C for 5 min. The chromatin fraction was extracted and fragmented by sonication (fragment length, 2 kb on average) as described previously [41], except for the use of 800 µl of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and a UD-201 ultrasonic disruptor (TOMY SEIKO). Sonicated chromatin (34 µl) was treated with the End-It DNA End-Repair kit (Epicentre) in a 50 µl reaction mixture at room temperature for 45 min. After heating at 70°C for 10 min, reaction mixture (23.5 µl) was incubated in the presence or absence of T4 DNA ligase (Roche) at room temperature for 2 h. After reverse crosslinking at 65°C followed by RNase A and Proteinase K treatment, DNA was purified using ChIP DNA Clean & Concentrator. The purified DNA was used as a template for the first PCR with KOD FX (Toyobo) and a primer set including one primer containing an *I-SceI* site that was biotinylated at the 5’ end (Table S3). PCR conditions were as follows: denaturing at 94°C for 2 min; 30 cycles of 98°C for 10 sec, 60°C for 30 sec, and 68°C for 6 min. The reaction mixture (15 µl) was mixed with 15 µl of Dynabeads M-280 Streptavidin (Thermo Fisher Scientific) and 500 µl of RIPA buffer (50 mM Tris [pH 7.5], 150 mM NaCl, 1 mM EDTA, 0.5% sodium deoxycholate, 0.1% SDS, 1% IGEPAL-CA630) at 4°C for 1 h. After three washes with RIPA buffer and one wash with 1× NEBuffer 2 (New England Biolabs), the Dynabeads were treated with *I-Sce I* at 37°C for 2 h. The supernatant was collected, incubated at 65°C for 20 min, and used for the second PCR with AmpliTaq Gold 360 Master Mix (Applied Biosystems). PCR conditions were as follows:
denaturing at 95°C for 10 min; 27 cycles of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 1 min.

Primers used in the analysis are shown in Table S3.

**RNA extraction and quantitative RT-PCR**

Total RNA was extracted from mock- or NaB-treated K562 cells and used for quantitative RT-PCR as previously described [42]. Primers used in the analysis are shown in Table S3.
SUPPORTING INFORMATION

Table S1. List of genomic regions detected in the differentiated state (sgRNA #6) by enChIP-Seq. (.xlsx)

Table S2. List of genomic regions detected in the differentiated state (sgRNA #17) by enChIP-Seq. (.xlsx)

Table S3. Primers used in this study. (.xlsx)

Figure S1. Filtering of NGS peaks to identify genomic regions that interact with the 5’HS5 locus in the undifferentiated state. (Step 1) Extraction of genomic regions that interact with the 5’HS5 locus specifically in the undifferentiated state. After the extraction step, regions fulfilling the defined criteria (one for sgRNA #6 and seven for sgRNA #17) were analyzed in step 2. (Step 2) Identification of peaks detected in common using sgRNA #6 and #17.

Figure S2. Scheme for proximity ligation assay to confirm interactions between genomic regions. Cells are crosslinked and lysed, and genomic DNA is randomly fragmented by sonication or other methods. After repair of DNA ends, proximal ligation, and reversal of crosslinking, the junction between the target locus and its potential interacting region is amplified by PCR. If a biotinylated primer is used, the amplicon can be purified using streptavidin. Subsequently, a part of the amplified region in the potential interacting locus is detected by the second PCR. Amplification of the region in the second PCR suggests that the potential interacting locus is physically proximal to the target locus.
Figure S3. NGS peak images in the \( \beta \)-globin locus. Although a peak at 5’HS5 was clearly observed, no peak indicated a physical interaction between 5’HS5 and other regions in the \( \beta \)-globin locus. Raw ChIP-Seq read data were displayed as density plots in the UCSC Genome Browser. The vertical viewing range (y-axis shown as Scale) was set at 1-300, based on the magnitude of the noise peaks. Black vertical bars show locus positions in the human genome (hg19 assembly). Positions of genes, including the \( \beta \)-globin genes, are shown under the plots.
ACKNOWLEDGEMENTS

We thank G.M. Church for providing a plasmid (Addgene plasmid #41824), F. Kitaura for technical assistance, and T. Kikuchi, H. Horiuchi, and M. Tosaka for NGS analysis.
REFERENCES


FIGURE LEGENDS

Figure 1. Schematic of the use of enChIP-Seq analysis to identify physical interactions between genomic regions. (A) The system consists of 3xFLAG-dCas9 (a fusion protein of the 3xFLAG-tag, dCas9, and a nuclear localization signal [NLS]) and a single guide RNA (sgRNA). (B) 3xFLAG-dCas9 and the sgRNA are expressed in the cells to be analyzed. The cells are crosslinked, if necessary, and lysed. Chromatin is purified and fragmented by sonication or other methods. Complexes containing the CRISPR complex are immunoprecipitated with anti-FLAG Ab. After reversal of crosslinking, if necessary, DNA is purified and subjected to the NGS analysis.

Figure 2. Isolation of the 5’HS5 locus by enChIP. (A) Schematic depiction of the β-globin locus. The positions of the 5’HS and 3’HS1 loci and the β-globin genes are indicated. (B) Positions of sgRNA target sites. Purple: CTCF-binding site [29]; red: 5’HS5 core region (NCBI Reference Sequence: NG_000007.3); blue: sgRNA target site (5’HS5 #6); green: sgRNA target site (5’HS5 #17); italic and underline: primer positions used in enChIP-real-time PCR in (C). (C) Yields of enChIP for the 5’HS5 locus. (D and E) NGS peaks at the target 5’HS5 locus (D) and the irrelevant Sox2 locus (E). Raw ChIP-Seq read data were displayed as density plots in the UCSC Genome Browser. The vertical viewing range (y-axis shown as Scale) was set at 1-300 based on the magnitude of the noise peaks. Black vertical bars show locus positions in the human genome (hg19 assembly). The position of the Sox2 gene is shown under the plot (E).

Figure 3. Filtering of the NGS peaks to identify genomic regions that interact with the 5’HS5 locus in the differentiated state. (Step 1) Extraction of enChIP-specific NGS peaks by
comparison with enChIP and input peaks detected in the differentiated state, using Model-based Analysis of ChIP-Seq (MACS) (http://liulab.dfci.harvard.edu/MACS/). The extracted enChIP-specific peaks include NaB-specific and constitutively detected peaks. (Step 2) Elimination of peaks derived from non-specific binding of dCas9. After the elimination step, peaks fulfilling the defined criteria (19 for sgRNA #6 and 228 for sgRNA #17) were analyzed in step 3. (Step 3) Identification of peaks detected in common by both sgRNA #6 and #17. Genomic regions corresponding to the identified peaks were considered to physically interact with the 5’HS5 locus. (Step 4) Extraction of enChIP-specific NGS peaks by comparison with enChIP and input peaks detected in the undifferentiated state, using MACS. The extracted enChIP-specific peaks include peaks specifically detected in undifferentiated cells and constitutively detected peaks. (Step 5) Identification of genomic regions that interact with the 5’HS5 locus in a NaB-specific manner.

Figure 4. Identification of genomic regions that physically interact with the 5’HS5 locus by enChIP-Seq. Raw ChIP-Seq read data were displayed as density plots in the UCSC Genome Browser. The vertical viewing range (y-axis shown as Scale) was set at 1-160 based on the noise peaks. Black vertical bars reveal locus positions in the human genome (hg19 assembly). Positions of genes are shown under the plots.

Figure 5. Confirmation of enChIP-Seq results by proximity ligation assay. (Upper) Results of PCR. K562 cells cultured in the presence or absence of NaB were crosslinked, and the chromatin was extracted, fragmented, and subjected to the proximity ligation assay (Figure S2) to confirm interaction between the 5’HS5 and ZNF670/ZFN670-695 loci. A region of the ZNF670/ZFN670-695 locus was detected by the second PCR. (Lower) The result of PCR with
DNA used for the first PCR and the primer set used for the second PCR. This result shows that the reactions were run with comparable amounts of input DNA. Two representatives of each sample (±NaB treatment) are shown.

Figure 6. NaB-induced expression of genes in the vicinity of genomic regions that interact with the 5’HS5 locus. Total RNA was used for quantitative RT-PCR analysis. Expression levels of the indicated genes were normalized to those of GAPDH, and the levels of mRNA in the absence of NaB were defined as 1 (mean ± SD, n = 3). N.D.: not detected, *: t-test p value <0.05, **: t-test p value <0.01.
### Table 1. NGS information regarding the 5’HS5 target locus.

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>Condition</th>
<th>Fold enrichment (compared with input gDNA)</th>
<th>Number of tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>#6</td>
<td>minus (mock)</td>
<td>120</td>
<td>655</td>
</tr>
<tr>
<td></td>
<td>NaB</td>
<td>168</td>
<td>783</td>
</tr>
<tr>
<td>#17</td>
<td>minus (mock)</td>
<td>65</td>
<td>273</td>
</tr>
<tr>
<td></td>
<td>NaB</td>
<td>122</td>
<td>217</td>
</tr>
</tbody>
</table>
### Table 2. List of peak positions detected in common by both sgRNA #6 and #17.

<table>
<thead>
<tr>
<th>chr</th>
<th>start</th>
<th>end</th>
<th>length</th>
<th>tags</th>
<th>fold enrichment</th>
<th>chr</th>
<th>start</th>
<th>end</th>
<th>length</th>
<th>tags</th>
<th>fold enrichment</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>247240307</td>
<td>247240894</td>
<td>588</td>
<td>68</td>
<td>17.62</td>
<td>chr1</td>
<td>247240610</td>
<td>247240804</td>
<td>195</td>
<td>18</td>
<td>12.84</td>
<td>NaB-specific</td>
</tr>
<tr>
<td>chr1</td>
<td>247240925</td>
<td>247241339</td>
<td>415</td>
<td>49</td>
<td>12.95</td>
<td>chr1</td>
<td>247240971</td>
<td>247241324</td>
<td>354</td>
<td>28</td>
<td>10.33</td>
<td>NaB-specific</td>
</tr>
<tr>
<td>chr7</td>
<td>132001863</td>
<td>132002364</td>
<td>502</td>
<td>71</td>
<td>27.24</td>
<td>chr7</td>
<td>132001929</td>
<td>132002385</td>
<td>457</td>
<td>50</td>
<td>22.29</td>
<td>NaB-specific</td>
</tr>
<tr>
<td>chr11</td>
<td>66057056</td>
<td>66057342</td>
<td>287</td>
<td>45</td>
<td>30.65</td>
<td>chr11</td>
<td>66057120</td>
<td>66057297</td>
<td>178</td>
<td>13</td>
<td>14.48</td>
<td>NaB-specific</td>
</tr>
<tr>
<td>chr15</td>
<td>64162769</td>
<td>64163183</td>
<td>415</td>
<td>51</td>
<td>29.06</td>
<td>chr15</td>
<td>64162762</td>
<td>64163061</td>
<td>300</td>
<td>31</td>
<td>27.51</td>
<td>NaB-specific</td>
</tr>
<tr>
<td>chr11</td>
<td>5311586</td>
<td>5312647</td>
<td>1062</td>
<td>783</td>
<td>168.11</td>
<td>chr11</td>
<td>5311878</td>
<td>5312599</td>
<td>722</td>
<td>217</td>
<td>122.27</td>
<td>constitutive</td>
</tr>
</tbody>
</table>
Figure 1

A

3xFLAG-dCas9

gRNA

3xFLAG tag
dCas9
NLS

B

Interacting genomic regions

Crosslink (if necessary)
Fragmentation

IP with anti-FLAG Ab

Reverse crosslink

DNA analysis by next-generation sequencing (NGS)
Figure 2

The figure shows the results of CRISPR-Cas9 targeting of the HBE1 and HBG2 genes in human erythroid progenitors. The top panel (A) illustrates the genomic locations of the genes and sgRNAs used. The middle panel (B) displays the sequences of the target sites for the sgRNAs. The bottom panels (C and D) depict the enrichment of dCas9 and NaB at the target sites in the presence and absence of sgRNAs.

The y-axis in panel C represents the percentage of input, while the x-axis shows the genomic coordinates. The bars indicate the enrichment of dCas9 and NaB for sgRNA #6 and #17. The genomic regions are annotated with the gene names and the presence or absence of sgRNAs.

The figure also includes a scale bar indicating the genomic coordinates (chr11: 5,310,000-5,320,000) and the genomic context (20 kb). The figure is licensed under a CC-BY 4.0 International license.
Figure 3

**Step 1**
- dCas9 + sgRNA #6
  - Input (NaB treatment) vs enChIP (NaB treatment)
  - enChIP #6 peaks (NaB-specific or constitutive) (2297 peaks)

- dCas9
  - Input (NaB treatment) vs enChIP (NaB treatment)
  - Off-target sites (NaB-specific or constitutive) (607 peaks)

- dCas9 + sgRNA #17
  - Input (NaB treatment) vs enChIP (NaB treatment)
  - enChIP #17 peaks (NaB-specific or constitutive) (3059 peaks)

**Step 2**
- enChIP #6 peaks (NaB-specific or constitutive)
  - vs
  - Off-target sites (NaB-specific or constitutive)

- enChIP #17 peaks (NaB-specific or constitutive)
  - vs
  - Off-target sites (NaB-specific or constitutive)

- enChIP #6-specific sites (NaB-specific or constitutive) (2161 peaks)
  - Fold enrichment (vs Input) : > 10, tag : > 39
  - (19 peaks, Supplemental Table S1)

- enChIP #17-specific sites (NaB-specific or constitutive) (2933 peaks)
  - Fold enrichment (vs Input) : > 10, tag : > 11
  - (228 peaks, Supplemental Table S2)

**Step 3**
- enChIP #6-specific sites (NaB-specific or constitutive) (19 peaks)

- enChIP #17-specific sites (NaB-specific or constitutive) (228 peaks)

- enChIP #6 #17-common sites (NaB-specific or constitutive) (6 peaks)

**Step 4**
- dCas9 + sgRNA #6
  - Input (mock treatment [minus]) vs enChIP (mock treatment [minus])
  - enChIP #6 peaks (minus-specific or constitutive) (118 peaks)

- dCas9 + sgRNA #17
  - Input (mock treatment [minus]) vs enChIP (mock treatment [minus])
  - enChIP #17 peaks (minus-specific or constitutive) (595 peaks)

**Step 5**
- enChIP #6 #17-common sites (NaB-specific or constitutive) vs
  - enChIP #6 #6 peaks (minus-specific or constitutive)
  - enChIP #17 peaks (minus-specific or constitutive)

- enChIP #6 #17-common site (constitutive)
  - (1 peak = 5'HS5 site, Table 2)

- enChIP #6 #17-common sites (NaB-specific)
  - (5 peaks, Table 2)
Figure 4
Figure 5

Proximity ligation assay

Input

NaB ligation

M #1 #2 #1 #2 H₂O
Figure 6

![Bar chart showing relative mRNA expression levels for different genes under different conditions.](image_url)
<table>
<thead>
<tr>
<th>chr</th>
<th>start</th>
<th>end</th>
<th>length</th>
<th>summit</th>
<th>tags</th>
<th>10*LOG10(pvalue)</th>
<th>fold_enrichment</th>
<th>FDR(%)</th>
<th>gene1</th>
<th>gene2</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>33526002</td>
<td>33526540</td>
<td>539</td>
<td>188</td>
<td>95</td>
<td>889.16</td>
<td>51.67</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr1</td>
<td>15358701</td>
<td>153587167</td>
<td>467</td>
<td>81</td>
<td>49</td>
<td>259.09</td>
<td>15.83</td>
<td>0</td>
<td>1654:NM_080388:S100A16-1640:NM_020672:S100A14</td>
<td></td>
</tr>
<tr>
<td>chr1</td>
<td>247240925</td>
<td>247241339</td>
<td>415</td>
<td>243</td>
<td>49</td>
<td>221.87</td>
<td>12.95</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr10</td>
<td>15387617</td>
<td>15388319</td>
<td>703</td>
<td>369</td>
<td>382</td>
<td>3100</td>
<td>101.06</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr11</td>
<td>5011586</td>
<td>5012847</td>
<td>1062</td>
<td>595</td>
<td>763</td>
<td>3100</td>
<td>168.11</td>
<td>0</td>
<td>-1220:NM_001204220:NM_033213:ZNF670</td>
<td></td>
</tr>
<tr>
<td>chr11</td>
<td>3612944</td>
<td>3613372</td>
<td>425</td>
<td>281</td>
<td>63</td>
<td>441.98</td>
<td>27.79</td>
<td>0</td>
<td>-1220:NM_001204220:NM_033213:ZNF670</td>
<td></td>
</tr>
<tr>
<td>chr11</td>
<td>57659714</td>
<td>57670066</td>
<td>353</td>
<td>129</td>
<td>42</td>
<td>199.96</td>
<td>11.97</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr11</td>
<td>66057056</td>
<td>66057342</td>
<td>287</td>
<td>127</td>
<td>43</td>
<td>329.48</td>
<td>20.65</td>
<td>0</td>
<td>0.0317:NM_153266:TMEM151A</td>
<td>705:NM_0020470:YIF1A</td>
</tr>
<tr>
<td>chr11</td>
<td>66131388</td>
<td>66131805</td>
<td>418</td>
<td>189</td>
<td>56</td>
<td>388.52</td>
<td>33.37</td>
<td>0</td>
<td>-7485:NM_001532:SLC29A2</td>
<td></td>
</tr>
<tr>
<td>chr12</td>
<td>76347754</td>
<td>76348240</td>
<td>487</td>
<td>266</td>
<td>66</td>
<td>513.86</td>
<td>22.08</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr15</td>
<td>64182789</td>
<td>64183183</td>
<td>415</td>
<td>80</td>
<td>51</td>
<td>400.95</td>
<td>29.06</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr16</td>
<td>1359061</td>
<td>1359401</td>
<td>341</td>
<td>284</td>
<td>41</td>
<td>299.03</td>
<td>25.24</td>
<td>0</td>
<td>0.0332345:NM_194259:NM_194260:NM_194261:UBE2</td>
<td></td>
</tr>
<tr>
<td>chr16</td>
<td>67522988</td>
<td>67523488</td>
<td>500</td>
<td>265</td>
<td>100</td>
<td>883.92</td>
<td>38.37</td>
<td>0</td>
<td>8400:NM_004691:ATP6V0D1:5773:NM_001133:AGRP</td>
<td></td>
</tr>
<tr>
<td>chr5</td>
<td>18103377</td>
<td>18103798</td>
<td>414</td>
<td>190</td>
<td>44</td>
<td>286.82</td>
<td>19.07</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr5</td>
<td>18118465</td>
<td>18178515</td>
<td>599</td>
<td>363</td>
<td>44</td>
<td>318.58</td>
<td>31.93</td>
<td>0</td>
<td>3285:NM_000414:NM_00119291:NM_00119292:HSD17B4</td>
<td></td>
</tr>
<tr>
<td>chr6</td>
<td>116781574</td>
<td>116781978</td>
<td>405</td>
<td>138</td>
<td>41</td>
<td>208.29</td>
<td>13.35</td>
<td>0</td>
<td>982:NM_0010109:VAMP2</td>
<td></td>
</tr>
<tr>
<td>chr7</td>
<td>132001863</td>
<td>132002364</td>
<td>502</td>
<td>272</td>
<td>71</td>
<td>488.49</td>
<td>27.24</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>108639633</td>
<td>108640362</td>
<td>430</td>
<td>283</td>
<td>41</td>
<td>171.13</td>
<td>14.3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Common** (sgRNA #6 and #17)

**Target site**
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Symbol</th>
<th>Fold Enrichment</th>
<th>Gene ID</th>
<th>Gene Symbol</th>
<th>Fold Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>51183996</td>
<td>NM_014965</td>
<td>4.39</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>51183950</td>
<td>NM_014965</td>
<td>6.28</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>51183996</td>
<td>NM_014965</td>
<td>4.39</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>51183950</td>
<td>NM_014965</td>
<td>6.28</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>119195471</td>
<td>NM_004965</td>
<td>6.28</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>119195471</td>
<td>NM_004965</td>
<td>6.28</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>119195471</td>
<td>NM_004965</td>
<td>6.28</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
<tr>
<td>119195471</td>
<td>NM_004965</td>
<td>6.28</td>
<td>119098812</td>
<td>NM_004965</td>
<td>11.43</td>
</tr>
</tbody>
</table>

**Table S2.**

The table above shows the fold enrichment for different gene IDs and symbols, with a focus on those with a fold enrichment of at least 4.39. The table includes multiple entries for some gene IDs, indicating the presence of multiple symbols or variations. The data is presented in a tabular format with columns for Gene ID, Gene Symbol, and Fold Enrichment. The fold enrichment values range from 4.39 to 11.43, with the highest values indicating a significant enrichment for those gene symbols.
<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Sequence (5’→3’)</th>
<th>Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>27420</td>
<td>hHS5-TAL-Target-F</td>
<td>ccagtttctccagtttcccttt</td>
<td>ChIP assays in Figure 2C (5’HS5)</td>
</tr>
<tr>
<td>27421</td>
<td>hHS5-TAL-Target-R</td>
<td>tttcaaaatgcaaggtgatgtc</td>
<td>ChIP assays in Figure 2C (5’HS5)</td>
</tr>
<tr>
<td>27222</td>
<td>hSox2-prom-F</td>
<td>atgggccctgaanaaccattatt</td>
<td>ChIP assays in Figure 2C (Sox2)</td>
</tr>
<tr>
<td>27223</td>
<td>hSox2-prom-R</td>
<td>ctgccttgacaactctgtatcct</td>
<td>ChIP assays in Figure 2C (Sox2)</td>
</tr>
<tr>
<td>27915</td>
<td>h5’HS5-3D-3’-I-SceI-F</td>
<td>(Biotin)cggtgTGGGATTACGGTAAACAGGTTTCCGTTTTGTTTCTAA</td>
<td>1st PCR in 3C-based assay in Figure 5 (Upper cases: I-Sce I site)</td>
</tr>
<tr>
<td>27880</td>
<td>hZNF670-3D-F2</td>
<td>gatgcttgcccgtgggctc</td>
<td>1st PCR in 3C-based assay in Figure 5</td>
</tr>
<tr>
<td>27863</td>
<td>hZNF670-3D-F</td>
<td>acgtaggggggaaggtcttt</td>
<td>Nested PCR in 3C-based assay in Figure 5</td>
</tr>
<tr>
<td>27864</td>
<td>hZNF670-3D-R</td>
<td>ccagagaacagactgctgaacac</td>
<td>Nested PCR in 3C-based assay in Figure 5</td>
</tr>
<tr>
<td>27771</td>
<td>hZNF670-ZNF695-Ex1_2-F</td>
<td>aagttgctttgctgggacgtc</td>
<td>RT-PCR in Figure 6 (ZNF670-ZNF695)</td>
</tr>
<tr>
<td>27772</td>
<td>hZNF670-ZNF695-Ex1_2-R</td>
<td>cacatctctgtgaccaactagtgc</td>
<td>RT-PCR in Figure 6 (ZNF670-ZNF695)</td>
</tr>
<tr>
<td>27779</td>
<td>hZNF670-Ex4-F2</td>
<td>gctcccagagcagtgtctgtgc</td>
<td>RT-PCR in Figure 6 (ZNF670)</td>
</tr>
<tr>
<td>27780</td>
<td>hZNF670-Ex4-R2</td>
<td>acagagagctggagccccacagag</td>
<td>RT-PCR in Figure 6 (ZNF670)</td>
</tr>
<tr>
<td>27773</td>
<td>hMIR422A-F</td>
<td>gcctccctcatcttctttgatc</td>
<td>RT-PCR in Figure 6 (MIR422A)</td>
</tr>
<tr>
<td>27774</td>
<td>hMIR422A-R</td>
<td>aagcttcgcagcagcagacag</td>
<td>RT-PCR in Figure 6 (MIR422A)</td>
</tr>
<tr>
<td>27782</td>
<td>hTMEM151A-Ex2-F</td>
<td>ctagagctccgagctgtctgtg</td>
<td>RT-PCR in Figure 6 (TMEM151A)</td>
</tr>
<tr>
<td>27783</td>
<td>hTMEM151A-Ex2-R</td>
<td>ctagagctccgagctgtctgtg</td>
<td>RT-PCR in Figure 6 (TMEM151A)</td>
</tr>
<tr>
<td>27777</td>
<td>hYF1A-Ex1_2-F</td>
<td>agacagctggcctcggttctgtg</td>
<td>RT-PCR in Figure 6 (YF1A)</td>
</tr>
<tr>
<td>27778</td>
<td>hYF1A-Ex1_2-R</td>
<td>gaatcgtacccgctgtctgtg</td>
<td>RT-PCR in Figure 6 (YF1A)</td>
</tr>
<tr>
<td>27784</td>
<td>hCNNH2-Ex1_4-F</td>
<td>gctctccctctctctctgtg</td>
<td>RT-PCR in Figure 6 (CNNH2)</td>
</tr>
<tr>
<td>27785</td>
<td>hCNNH2-Ex1_4-R</td>
<td>agttttcgcggcaccacagag</td>
<td>RT-PCR in Figure 6 (CNNH2)</td>
</tr>
</tbody>
</table>
Step 1

enChIP #6 peaks
(NaB-specific or constitutive)
(2297 peaks, Step 1 in Figure 3)
vs
enChIP #17 peaks
(NaB-specific or constitutive)
(3059 peaks, Step 1 in Figure 3)

enChIP #6 peaks
(minus-specific or constitutive)
(118 peaks, Step 4 in Figure 3)
vs
enChIP #17 peaks
(minus-specific or constitutive)
(595 peaks, Step 4 in Figure 3)

enChIP #6-specific peaks
(minus-specific)
(59 peaks)
Fold enrichment (vs Input) : > 10, tag : > 33
(1 peak)

enChIP #17-specific peaks
(minus-specific)
(507 peaks)
Fold enrichment (vs Input) : > 10, tag : > 13
(7 peaks)

Step 2

enChIP #6-specific peaks
(minus-specific)
(1 peak)
vs
enChIP #17-specific peaks
(minus-specific)
(7 peaks)
0 peak
Crosslinking DNA fragmentation (e.g. sonication)

DNA-end repair ligation

Reverse crosslinking purification of DNA

1st PCR

2nd PCR

Purification of amplified DNA (if necessary)