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

112 detected in cells expressing either sgRNA #6 or #17, but not in cells expressing neither sgRNA
113 (Figure 2D and Table 1). By contrast, no peak was detected at the irrelevant *Sox2* locus (Figure
114 2E).
115
116 To identify genomic regions physically interacting with 5'HS5 upon erythroid differentiation, we
117 analyzed the NGS peaks observed in the differentiated state. The CRISPR complex can interact
118 with multiple genomic sites containing sequences similar to the sgRNA sequence [30-33]. In
119 addition, in the absence of sgRNA, it is possible that dCas9 could bind non-specifically to some
120 genomic sites *in vivo*. Therefore, the peaks identified for sgRNA #6 and #17 may include such
121 off-target sites. To remove those off-target sites, we first eliminated peaks derived from
122 non-specific binding of dCas9 in the absence of sgRNA from those identified for each sgRNA
123 (Steps 1 and 2 in Figure 3). To identify peaks with confidence, we next established two criteria
124 for choosing peaks based on NGS information from the target 5'HS5 locus: (1) tag number >5%
125 of that of the target 5'HS5 locus, and (2) fold enrichment relative to input genomic DNA >10. As
126 shown in Figure 3 (Step 2), 19 and 228 peaks for sgRNA #6 and #17, respectively, fulfilled these
127 criteria. Next, to eliminate sgRNA-dependent off-target sites, we compared the peaks for sgRNA
128 #6 and #17 and selected peaks detected in common by both sgRNA #6 and #17. These peaks
129 were considered to represent regions engaged in bona fide physical interactions with the 5'HS5
130 locus (Step 3 in Figure 3). The six identified peaks could be classified into two categories: (1)
131 peaks that were larger in the differentiated state, and (2) peaks constitutively observed in both the
132 undifferentiated and differentiated states. The first category should contain genomic regions that
133 inducibly associate with the 5'HS5 locus upon erythroid differentiation, whereas the second
134 category should contain genomic regions constitutively associated with the 5'HS5 locus. To
135 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

229 editing and gene regulation [36]. Tagging of a given locus with dCas9s derived from different
230 lineages and bearing distinct tags would make it feasible to sequentially purify the locus,
231 minimizing contamination with off-target sites.

232 Using these techniques, we believe that we can effectively manage potential contamination of
233 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'-TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGATTCATAGCAGACA GCTA-3' and 5'-GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTAGCTGTCTGCTATGAA TCC-3'; hHS5 #17, 5'-TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGAAGATAGGGTAA GAGAC-3' and 5'-GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTCTCTTACCCTATCTT 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 *Xho*I and *Hind*III and cloned into *Xho*I/*Hind*III-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

291 performed at the University of Tokyo as described previously [39, 40]. Additional data analysis
292 for Step 2 in Figure 3 was performed at Hokkaido System Science Co., Ltd. Images of NGS
293 peaks were generated using the UCSC Genome Browser
294 (https://genome.ucsc.edu/cgi-bin/hgGateway).
295
296 **Proximity ligation assay to confirm interactions between genomic regions**
297 K562 cells (1×10^7) were fixed with 1% formaldehyde at 37°C for 5 min. The chromatin fraction
298 was extracted and fragmented by sonication (fragment length, 2 kb on average) as described
299 previously [41], except for the use of 800 µl of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and
300 a UD-201 ultrasonic disruptor (TOMY SEIKO). Sonicated chromatin (34 µl) was treated with the
301 End-It DNA End-Repair kit (Epicentre) in a 50 µl reaction mixture at room temperature for 45
302 min. After heating at 70°C for 10 min, reaction mixture (23.5 µl) was incubated in the presence
303 or absence of T4 DNA ligase (Roche) at room temperature for 2 h. After reverse crosslinking at
304 65°C followed by RNase A and Proteinase K treatment, DNA was purified using ChIP DNA
305 Clean & Concentrator. The purified DNA was used as a template for the first PCR with KOD FX
306 (Toyobo) and a primer set including one primer containing an *I-SceI* site that was biotinylated at
307 the 5' end (Table S3). PCR conditions were as follows: denaturing at 94°C for 2 min; 30 cycles
308 of 98°C for 10 sec, 60°C for 30 sec, and 68°C for 6 min. The reaction mixture (15 µl) was mixed
309 with 15 µl of Dynabeads M-280 Streptavidin (Thermo Fisher Scientific) and 500 µl of RIPA
310 buffer (50 mM Tris [pH 7.5], 150 mM NaCl, 1 mM EDTA, 0.5% sodium deoxycholate, 0.1%
311 SDS, 1% IGEPAL-CA630) at 4°C for 1 h. After three washes with RIPA buffer and one wash
312 with 1× NEBuffer 2 (New England Biolabs), the Dynabeads were treated with *I-Sce I* at 37°C for
313 2 h. The supernatant was collected, incubated at 65°C for 20 min, and used for the second PCR
314 with AmpliTaq Gold 360 Master Mix (Applied Biosystems). PCR conditions were as follows:

315 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.

316 Primers used in the analysis are shown in Table S3.

317

318 **RNA extraction and quantitative RT-PCR**

319 Total RNA was extracted from mock- or NaB-treated K562 cells and used for quantitative

320 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.

345

346 **Figure S3.** NGS peak images in the *β-globin* locus. Although a peak at 5'HS5 was clearly
347 observed, no peak indicated a physical interaction between 5'HS5 and other regions in the
348 *β-globin* locus. Raw ChIP-Seq read data were displayed as density plots in the UCSC Genome
349 Browser. The vertical viewing range (*y*-axis shown as Scale) was set at 1-300, based on the
350 magnitude of the noise peaks. Black vertical bars show locus positions in the human genome
351 (hg19 assembly). Positions of genes, including the *β-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.

355 **REFERENCES**

356

357 1. Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation
358 in trans. Trends Genet. 2010;26:188-97.
359 2. Fisher AG, Merkenshlager M. Gene silencing, cell fate and nuclear organization. Curr Opin
360 Genet Dev. 2002;12(2):193-7.
361 3. Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene
362 regulation. Nature. 2007;447(7143):413-7.
363 4. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science.
364 2002;295(5558):1306-11.
365 5. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes
366 Dev. 2012;26(1):11-24.
367 6. Hoshino A, Fujii H. Insertional chromatin immunoprecipitation: a method for isolating
368 specific genomic regions. J Biosci Bioeng. 2009;108(5):446-9. doi:
369 10.1016/j.jbiosc.2009.05.005.
370 7. Fujita T, Fujii H. Direct idenification of insulator components by insertional chromatin
371 immunoprecipitation. PLoS One. 2011;6(10):e26109. doi: 10.1371/journal.pone.0026109.
372 8. Fujita T, Fujii H. Efficient isolation of specific genomic regions by insertional chromatin
373 immunoprecipitation (iChIP) with a second-generation tagged LexA DNA-binding domain.
374 Adv Biosci Biotechnol. 2012;3(5):626-9. doi: 10.4236/abb.2012.35081.
375 9. Fujita T, Fujii H. Efficient isolation of specific genomic regions retaining molecular
376 interactions by the iChIP system using recombinant exogenous DNA-binding proteins. BMC
377 Mol Biol. 2014;15:26. doi: 10.1186/s12867-014-0026-0.

10. Fujita T, Kitaura F, Fujii H. A critical role of the Thy28-MYH9 axis in B cell-specific expression of the *Pax5* gene in chicken B cells. PLoS One. 2015;10:e0116579. doi: 10.1371/journal.pone.0116579.

11. Fujita T, Fujii H. Biochemical analysis of genome functions using locus-specific chromatin immunoprecipitation technologies. Gene Regul Syst Bio. 2015;in press.

12. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochem Biophys Res Commun. 2013;439:132-6. doi: 10.1016/j.bbrc.2013.08.013.

13. Fujita T, Asano Y, Ohtsuka J, Takada Y, Saito K, Ohki R, et al. Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). Sci Rep. 2013;3:3171. doi: 10.1038/srep03171.

14. Fujita T, Fujii H. Identification of proteins associated with an IFNγ-responsive promoter by a retroviral expression system for enChIP using CRISPR. PLoS One. 2014;9(7):e103084. doi: 10.1371/journal.pone.0103084.

15. Fujita T, Yuno M, Okuzaki D, Ohki R, Fujii H. Identification of non-coding RNAs associated with telomeres using a combination of enChIP and RNA sequencing. PLoS One. 2015;10:e0123387. doi: 10.1371/journal.pone.0123387.

16. Fujita T, Yuno M, Fujii H. Efficient sequence-specific isolation of DNA fragments and chromatin by in vitro enChIP technology using recombinant CRISPR ribonucleoproteins. Genes Cells. 2016;in press.

17. Fujii H, Fujita T. Isolation of specific genomic regions and identification of their associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using the CRISPR system and TAL proteins. Int J Mol Sci. 2015;16:21802-12.

18. Fujita T, Fujii H. Applications of engineered DNA-binding molecules such as TAL proteins and the CRISPR/Cas system in biology research. Int J Mol Sci. 2015;16:23143-64.

19. Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys2his2 zinc finger poroteins. Annu Rev Biochem. 2001;70:313-40.

20. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. Science. 2011;333(6051):1843-6.

21. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152:1173-83. doi: 10.1016/j.cell.2013.02.022.

22. Ghirlando R, Giles K, Gowher H, Xiao T, Xu Z, Yao H, et al. Chromatin domains, insulators, and the regulation of gene expression. Biochim Biophys Acta. 2012;1819(644-651).

23. Holwerda SJ, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. Philos Trans R Soc Lond B Biol Sci. 2013;368:20120369.

24. Caterina JJ, Ryan TM, Pawlik KM, Palmiter RD, Brinster RL, Behringer RR, et al. Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. Proc Natl Acad Sci U S A. 1991;88:1626-30.

25. Peterson KR, Clegg CH, Navas PA, Norton EJ, Kimbrough TG, Stamatoyannopoulos G. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. Proc Natl Acad Sci U S A. 1996;93:6605-9.

26. Navas PA, Peterson KR, Li Q, McArthur M, Stamatoyannopoulos G. The 5'HS4 core element of the human beta-globin locus control region is required for high-level globin gene expression in definitive but not in primitive erythropoiesis. J Mol Biol. 2001;312:17-26.

27. Farrell CM, West AG, Felsenfeld G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol. 2002;22:3820-31.

28. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006;16:1299-309.

29. Chien R, Zeng W, Kawauchi S, Bender MA, Santos R, Gregson HC, et al. Cohesin mediates chromatin interactions that regulate mammalian β-globin expression. J Biol Chem. 2011;286:17870-8.

30. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32:670-6. doi: 10.1038/nbt.2889.

31. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol. 2014;32:677-83. doi: 10.1038/nbt.2916.

32. Cencic R, Miura H, Malina A, Robert F, Ethier S, Schmeing TM, et al. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. PLoS One. 2014;9:e109213. doi: 10.1371/journal.pone.0109213.

33. O'Green H, Henry IM, Bhakta MS, Meckler JF, Segal DJ. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 2015;43:3389-404. doi: 10.1093/nar/gkv137.

34. Deng B, Melnik S, Cook PR. Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. Semin Cancer Biol. 2013;23:65-71.

447 35. Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, et al. Spatial
448 genome organization: contrasting views from chromosome conformation capture and
449 fluorescence in situ hybridization. Genes Dev. 2014;28:2778-91.
450 36. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9
451 proteins for RNA-guided gene regulation and editing. Nat Methods. 2013;10:1116-21. doi:
452 10.1038/nmeth.2681.
453 37. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human
454 genome engineering via Cas9. Science. 2013;339:823-6. doi: 10.1126/science.1232033.
455 38. Miller AD, Buttimore C. Redesign of retrovirus packaging cell lines to avoid recombination
456 leading to helper virus production. Mol Cell Biol 1986;6:2895-902.
457 39. Yamashita R, Sathira NP, Kanai A, Tanimoto K, Arauchi T, Tanaka Y, et al. Genome-wide
458 characterization of transcriptional start sites in humans by integrative transcriptome analysis.
459 Genome Res. 2011;21:775-89.
460 40. Seki M, Masaki H, Arauchi T, Makauchi H, Sugano S, Suzuki Y. A comparison of the Rest
461 complex binding patterns in embryonic stem cells and epiblast stem cells. PLoS One.
462 2014;9(4):e95374.
463 41. Fujita T, Ryser S, Tortola S, Piuz I, Schlegel W. Gene-specific recruitment of positive and
464 negative elongation factors during stimulated transcription of the MKP-1 gene in
465 neuroendocrine cells. Nucleic Acids Res. 2007;35:1007-17.
466 42. Fujita T, Fujii H. Transcription start sites and usage of the first exon of mouse Foxp3 gene.
467 Mol Biol Rep. 2012;39:9613-9.
468
469

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

494 comparison with enChIP and input peaks detected in the differentiated state, using Model-based
495 Analysis of ChIP-Seq (MACS) (http://liulab.dfci.harvard.edu/MACS/). The extracted
496 enChIP-specific peaks include NaB-specific and constitutively detected peaks. (**Step 2**)
497 Elimination of peaks derived from non-specific binding of dCas9. After the elimination step,
498 peaks fulfilling the defined criteria (19 for sgRNA #6 and 228 for sgRNA #17) were analyzed in
499 step 3. (**Step 3**) Identification of peaks detected in common by both sgRNA #6 and #17. Genomic
500 regions corresponding to the identified peaks were considered to physically interact with the
501 5'HS5 locus. (**Step 4**) Extraction of enChIP-specific NGS peaks by comparison with enChIP and
502 input peaks detected in the undifferentiated state, using MACS. The extracted enChIP-specific
503 peaks include peaks specifically detected in undifferentiated cells and constitutively detected
504 peaks. (**Step 5**) Identification of genomic regions that interact with the 5'HS5 locus in a
505 NaB-specific manner.
506
507 **Figure 4. Identification of genomic regions that physically interact with the 5'HS5 locus by**
508 **enChIP-Seq.** Raw ChIP-Seq read data were displayed as density plots in the UCSC Genome
509 Browser. The vertical viewing range (*y*-axis shown as Scale) was set at 1-160 based on the noise
510 peaks. Black vertical bars reveal locus positions in the human genome (hg19 assembly). Positions
511 of genes are shown under the plots.
512
513 **Figure 5. Confirmation of enChIP-Seq results by proximity ligation assay.** (**Upper**) Results
514 of PCR. K562 cells cultured in the presence or absence of NaB were crosslinked, and the
515 chromatin was extracted, fragmented, and subjected to the proximity ligation assay (Figure S2) to
516 confirm interaction between the 5'HS5 and *ZNF670/ZFN670-695* loci. A region of the
517 *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.

527 **Table 1.** NGS information regarding the 5'HS5 target locus.

sgRNA	Condition	Fold enrichment (compared with input gDNA)	Number of tags
#6	minus (mock)	120	655
	NaB	168	783
#17	minus (mock)	65	273
	NaB	122	217

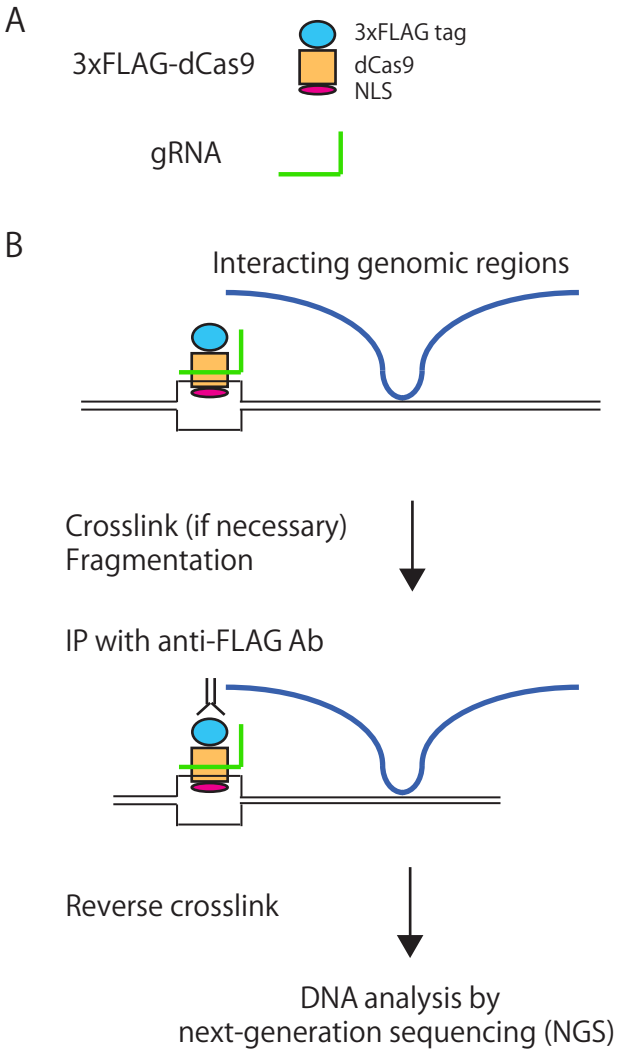
528

529 **Table 2.** List of peak positions detected in common by both sgRNA #6 and #17.

sgRNA #6 (NaB)						sgRNA #17 (NaB)						
chr	start	end	length	tags	fold enrichment	chr	start	end	length	tags	fold enrichment	status
chr1	247240307	247240894	588	68	17.62	chr1	247240610	247240804	195	18	12.84	NaB-specific
chr1	247240925	247241339	415	49	12.95	chr1	247240971	247241324	354	28	10.33	NaB-specific
chr7	132001863	132002364	502	71	27.24	chr7	132001929	132002385	457	50	22.29	NaB-specific
chr11	66057056	66057342	287	45	30.65	chr11	66057120	66057297	178	13	14.48	NaB-specific
chr15	64162769	64163183	415	51	29.06	chr15	64162762	64163061	300	31	27.51	NaB-specific
chr11	5311586	5312647	1062	783	168.11	chr11	5311878	5312599	722	217	122.27	constitutive

530

Figure 1



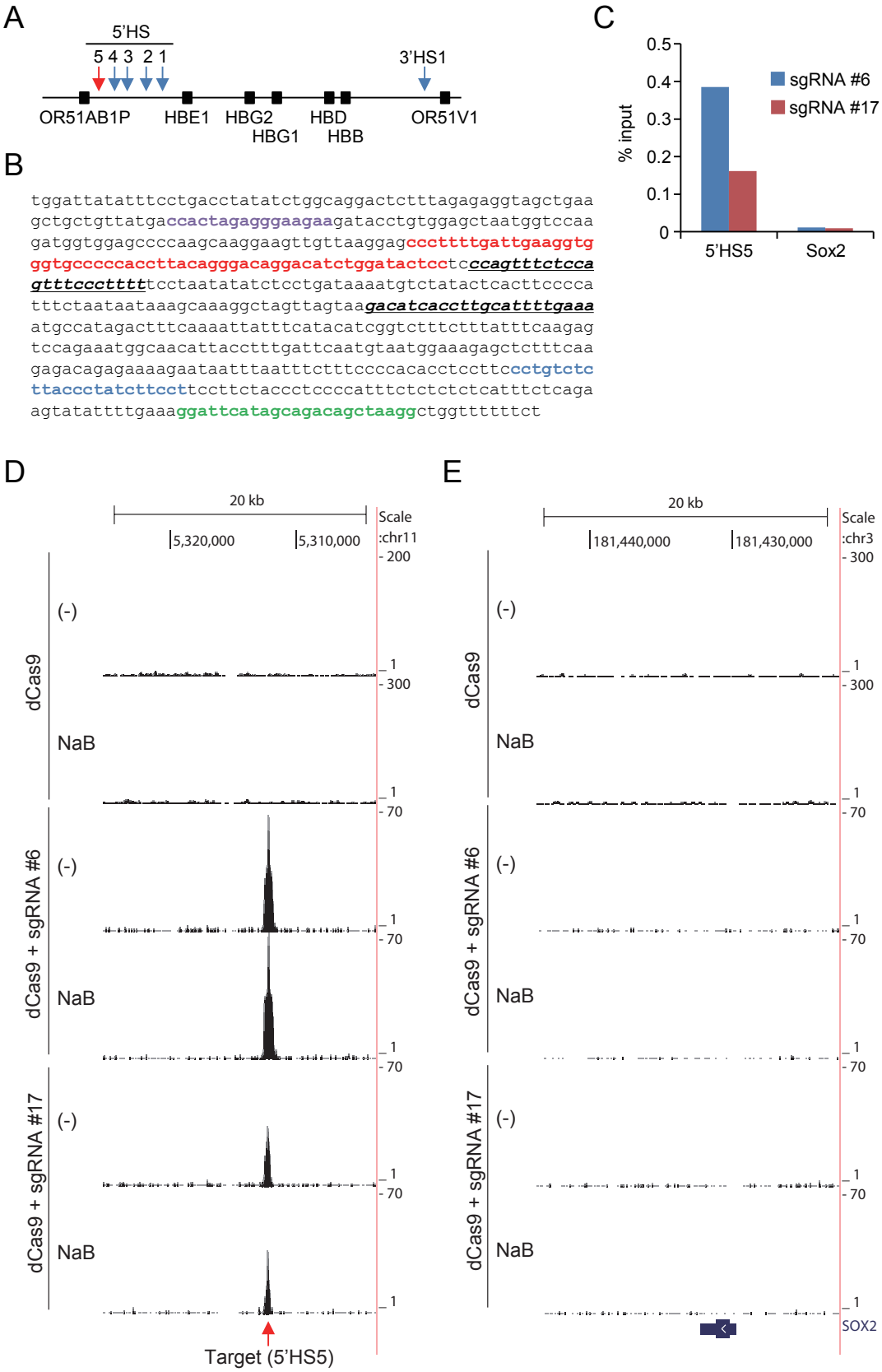


Figure 3

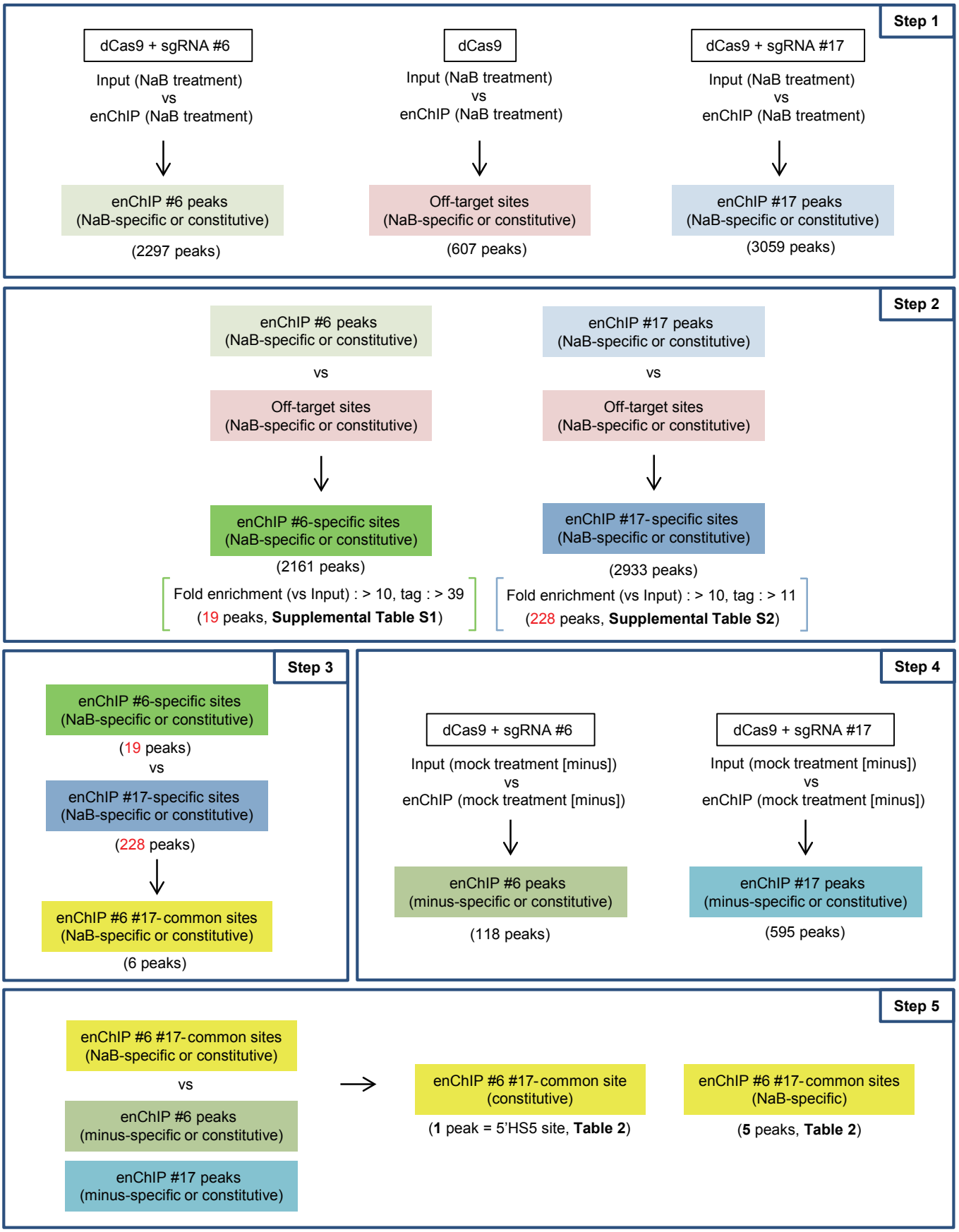


Figure 4

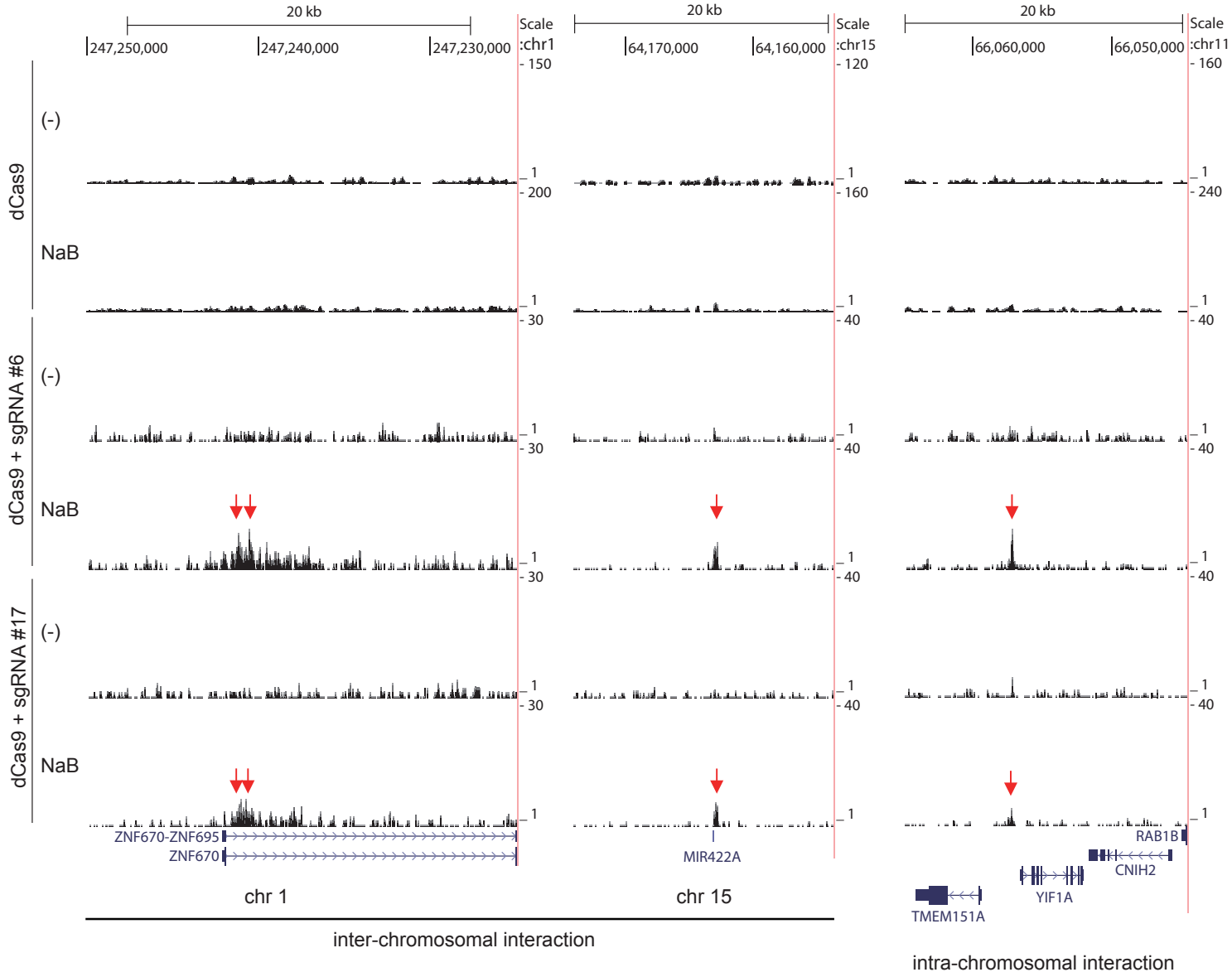


Figure 5

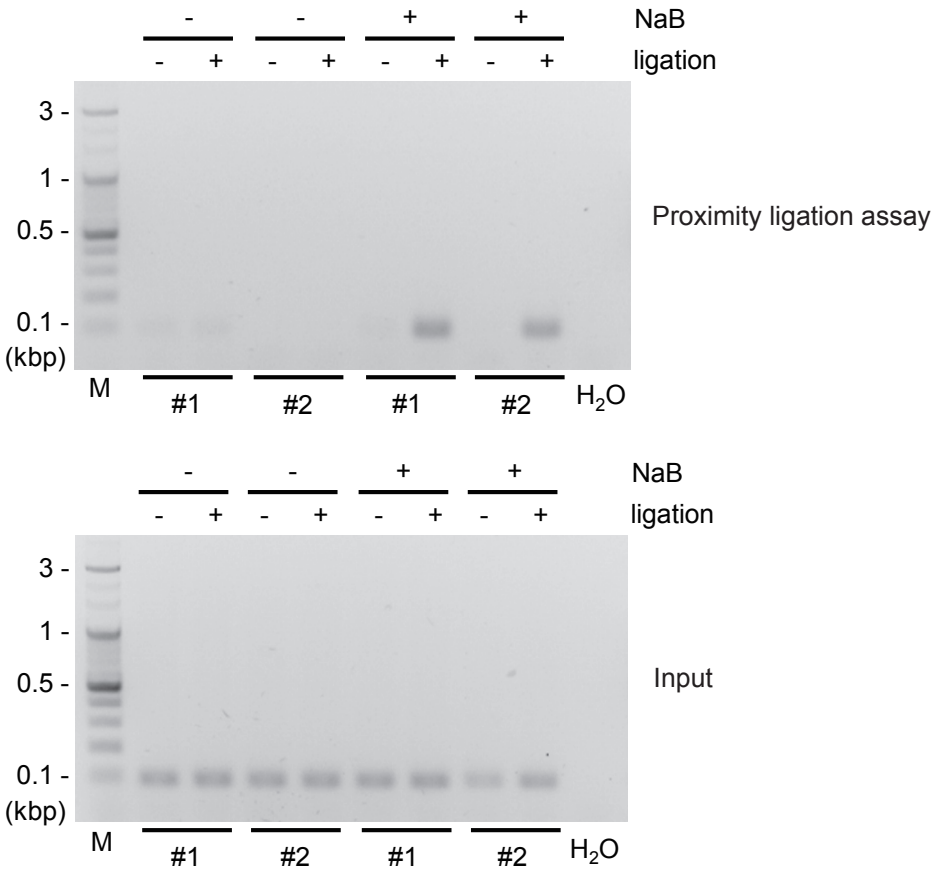


Figure 6

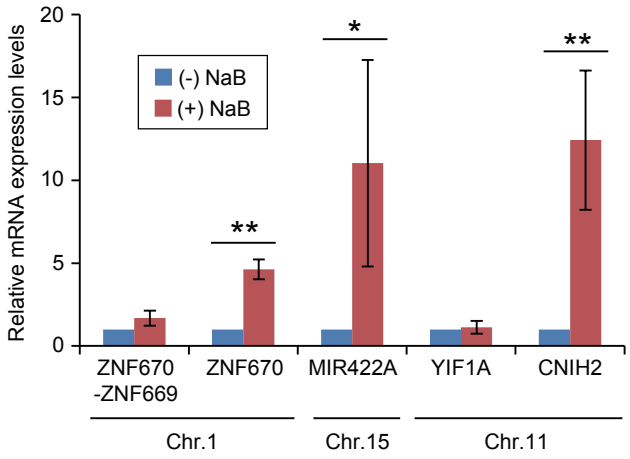


Table S1. List of genomic regions detected in the differentiated state (sgRNA #6) by enChIP-Seq.

sgRNA #6 (NaB-specific and constitutive peaks)										
chr	start	end	length	summit	tags	-10*LOG10(pvalue)	fold_enrichment	FDR(%)	gene1	gene2
chr1	33526002	33526540	539	188	95	869.16	51.67	0	.	.
chr1	153586701	153587167	467	81	49	259.09	15.83	0	.	1654:NM_080388:S100A16|-1640:NM_020672:S100A14|
chr1	247240307	247240894	588	246	68	322.89	17.62	0	.	-1220:NM_001204220,NM_033213:ZNF670|
chr1	247240925	247241339	415	243	49	221.87	12.95	0	.	-775:NM_001204220,NM_033213:ZNF670|
chr10	15387617	15388319	703	369	382	3100	101.06	0	.	.
chr11	5311586	5312647	1062	595	783	3100	168.11	0	.	.
chr11	36122948	36123372	425	281	63	441.98	27.79	0	.	.
chr11	57669714	57670066	353	129	42	199.98	11.97	0	.	.
chr11	66057056	66057342	287	127	45	329.48	30.65	0	2317:NM_153266:TMEM151A|	705:NM_020470:YIF1A|
chr11	66131388	66131805	418	189	56	388.52	33.37	0	.	-7485:NM_001532:SLC29A2|
chr12	76347754	76348240	487	266	76	513.86	22.08	0	.	.
chr15	64162769	64163183	415	80	51	400.95	29.06	0	.	.
chr16	1359061	1359401	341	284	41	299.03	25.24	0	93:NM_003345,NM_194259,NM_194260,NM_194261:UBE2I|	.
chr16	67522986	67523488	503	265	110	863.92	38.37	0	.	8400:NM_004691:ATP6V0D1|5773:NM_001138:AGRP|
chr4	81093373	81093786	414	190	46	266.82	19.07	0	.	.
chr5	118784853	118785151	299	163	44	318.58	31.93	0	3285:NM_000414,NM_001199291,NM_001199292:HSD17B4|	.
chr6	116781574	116781978	405	138	41	208.29	13.35	0	982:NM_001010919:FAM26F|	.
chr7	132001863	132002364	502	272	71	488.49	27.24	0	.	.
chr9	108639933	108640362	430	283	41	171.13	14.3	0	.	.

Common (sgRNA #6 and #17)

Target site

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

								sgRNA #17 (NaB-specific and constitutive peaks)		
chr	start	end	length	summit	tags	-10*LOG10(pvalue)	fold_enrichment	FDR(%)	gene1	gene2
chr1	11919466	11919666	201	155	16	106.97	21.5	0.62	.	675:NM_002521:NPPB|
chr1	12316588	12316738	151	51	11	61.36	11.46	1.52	.	.
chr1	12316760	12316874	115	56	11	61.65	14.06	1.46	.	.
chr1	16840452	16840625	174	129	28	105.1	12.77	0.61	.	.
chr1	25119803	25119904	102	65	11	87.09	18.88	1.02	.	.
chr1	26744588	26744690	103	56	11	90.74	26.45	0.81	−7319:NM_024674:LIN28A|	.
chr1	45187527	45187664	138	66	12	88.76	21.4	0.74	.	.
chr1	47201499	47201650	152	64	12	87.34	19.65	1.04	.	.
chr1	61511409	61511491	83	54	12	90.53	19.03	0.78	.	.
chr1	65583464	65583638	175	38	13	79.36	13.49	0.98	.	.
chr1	76238395	76238528	134	70	11	89.52	22.37	0.76	.	.
chr1	111289435	111289681	247	97	19	91.87	10.42	0.87	.	.
chr1	111492829	111493061	233	52	25	131.81	17.64	0	.	.
chr1	114755922	114756032	111	31	12	88.62	22.07	0.74	.	.
chr1	115300292	115300460	169	77	13	81.86	10.92	1.13	.	−210:NM_001007553,NM_001130523,NM_001242891,NM_001242892,NM_001242893,NM_007158:CSDE1|
chr1	118377797	118377896	100	33	16	154.8	30.77	0	.	.
chr1	118448119	118448224	106	32	11	83.6	18.24	1.21	.	.
chr1	118448271	118448398	128	33	11	75.27	15.81	0.83	.	.
chr1	119098643	119098812	170	114	16	88.27	17.54	0.72	.	.
chr1	119098833	119098996	164	92	12	76.75	17.19	0.89	.	.
chr1	142812735	142812886	152	55	15	56.75	11.62	1.62	.	.
chr1	154955985	154956103	119	42	12	68.62	14.11	1.15	−215:NM_001184891,NM_001184892,NM_025207,NM_201398:FLAD1|10	9145:NM_001130040,NM_001130041,NM_001202859,NM_003029,NM_183001:SHC1|
chr1	163151931	163152050	120	22	15	129.81	21.4	0	.	.
chr1	165366980	165367170	191	96	14	83.95	16.03	1.23	.	.
chr1	176760594	176760706	113	56	17	145.97	25.91	0	.	.
chr1	182156036	182156175	140	66	12	84.73	25.79	1.28	.	.
chr1	185033362	185033478	117	55	12	70.13	12.52	0.92	.	.
chr1	190281875	190281997	123	52	12	79.99	16.05	1.01	.	.
chr1	202662235	202662452	218	75	20	117.16	13.76	0	.	.
chr1	206600619	206600731	113	29	12	98.07	20.98	1.05	.	.
chr1	207682410	207682546	137	21	13	83.77	16.05	1.22	.	.
chr1	209779349	209779484	136	66	11	80.31	21.83	1.02	.	.
chr1	215441144	215441251	108	33	15	158.35	42.46	0	.	.
chr1	224278844	224278981	138	75	11	63.66	16.18	1.23	.	.
chr1	224659181	224659306	126	60	12	82.98	17.44	1.17	.	.
chr1	227498113	227498241	129	61	12	70.13	11.83	0.92	.	−7584:NM_003607,NM_014826:CDC42BPA|
chr1	230233766	230233849	84	30	15	97.35	14.12	1.02	.	.
chr1	231064213	231064322	110	83	12	100.86	21.54	1.09	.	.
chr1	233764914	233765193	280	171	23	117.81	15.28	0	.	.
chr1	235028303	235028404	102	24	12	88.24	17.89	0.72	.	.
chr1	242123452	242123574	123	30	15	66.9	11.28	1.07	.	.
chr1	247239197	247239370	174	79	12	61.3	11.08	1.52	.	−2744:NM_001204220,NM_033213:ZNF670|
chr1	247240610	247240804	195	110	18	91.65	12.84	0.85	.	−1310:NM_001204220,NM_033213:ZNF670|
chr1	247240971	247241324	354	67	28	95.69	10.33	0.98	.	−790:NM_001204220,NM_033213:ZNF670|
chr10	38811337	38811453	117	87	23	116.01	11.62	0	.	.
chr10	51531252	51531379	128	47	11	71.68	17.07	1.04	.	.
chr10	62735764	62735872	109	32	11	77.69	21.27	0.91	.	.
chr10	73341962	73342129	168	32	12	75.86	11.54	0.85	.	.
chr10	73555129	73555287	159	56	14	52.85	10.21	1.47	.	.
chr10	85609352	85609539	188	118	14	107.74	20.11	0.65	.	.
chr10	102027577	102027717	141	89	15	139.52	37.52	0	.	281:NM_018294:CWF19L1|
chr10	103124592	103124695	104	80	24	141.5	14.31	0	.	.
chr11	2306298	2306439	142	66	12	70.13	13.02	0.92	.	.
chr11	5311878	5312599	722	379	217	2485.12	122.27	0	.	.
chr11	10773231	10773367	137	43	11	80.96	22.29	1.08	−420:NM_014633:CTR9|	.
chr11	43598639	43598790	152	71	16	106.97	20.91	0.62	.	.
chr11	47747635	47747752	118	70	11	86.11	18.34	0.99	.	.
chr11	66057120	66057297	178	122	13	72.58	14.48	1.09	2253:NM_153266:TMEM151A|	660:NM_020470:YIF1A|
chr11	77516946	77517101	156	78	12	89.81	21.16	0.77	.	.
chr11	101073559	101073677	119	61	12	93.06	25.79	0.91	.	.
chr11	112021013	112021234	222	115	24	157.94	16.89	0	.	.
chr11	118797135	118797270	136	58	13	81.54	15.28	1.1	.	.
chr12	8868954	8869095	142	58	11	80.59	26.2	1.06	.	.
chr12	53135306	53135428	123	41	11	78.22	24.28	0.93	.	.
chr12	56203742	56203871	130	79	16	134.36	26.82	0	8064:NM_014182:ORMDL2|	−7668:NM_033082:SARNP|
chr12	57042458	57042561	104	63	13	118.99	32.75	0	.	2710:NM_001686:ATP5B|
chr12	85985227	85985323	97	46	12	123.66	37.52	0	.	.
chr12	98812996	98813098	103	39	11	107.33	28.92	0.64	.	.
chr12	105549728	105549905	178	103	15	122.99	28.57	0	.	.
chr12	108608436	108608645	210	44	13	78.91	12.84	0.95	.	.
chr12	117295952	117296132	181	102	14	111.08	28.66	0.7	.	.
chr14	33091150	33091253	104	59	11	99.11	24.69	1.07	.	.
chr14	90961717	90961859	143	63	12	105.15	39.07	0.61	.	.
chr14	95057304	95057394	91	29	13	52.85	10.01	1.47	−9573:NM_000624:SERPINA5|	.
chr15	64162762	64163061	300	169	31	237.77	27.51	0	.	.
chr16	11211397	11211556	160	70	13	93.99	22.6	0.95	.	.
chr16	12058873	12059000	128	44	11	89.81	21.49	0.77	91:NM_001192:TNFRSF17|	.
chr16	21144723	21144826	104	24	11	61.65	11.13	1.46	.	.
chr16	29805602	29805714	113	85	11	80.31	20.73	1.02	−3568:NM_001256269,NM_001256270,NM_007317:KIF22|	.
chr16	33905884	33906025	142	43	11	55.65	11.37	1.61	.	.
chr16	83050094	83050214	121	29	11	81.85	17.19	1.13	.	.
chr17	5144346	5144393	48	24	12	220.54	52.09	0	.	6311:NM_207103:SCIMP|
chr17	5144554	5144702	149	82	11	91.74	26.05	0.86	.	6620:NM_207103:SCIMP|
chr17	28456193	28456317	125	95	11	87.68	16.73	1.06	.	.
chr17	35198187	35198353	167	74	13	92.02	26.02	0.88	.	.
chr17	37005338	37005503	166	61	13	78.91	15.22	0.95	.	7862:NM_001080465:C17orf98|−4549:NM_000978:RPL23|
chr17	50624202	50624445	244	181	14	80.58	16.38	1.06	.	.
chr17	56081214	56081363	150	90	12	63.35	11.96	1.21	.	−3343:NM_001078166,NM_006924:SRSF1|
chr17	56197642	56197819	178	93	13	78.45	13.49	0.94	.	.
chr17	72667558	72667729	172	97	14	111.01	27.2	0.69	−302:NM_001006638,NM_001163989,NM_001163990,NM_175738:RAB3	.
chr17	77290668	77290823	156	88	13	101.06	25.04	1.1	.	.
chr18	2767358	2767467	110	85	13	78.91	11.77	0.95	.	.
chr18	21027513	21027636	124	76	11	55.2	11.46	1.72	5274:NM_003831:RIOK3|	9712:NM_032933:TMEM241|
chr19	2740205	2740339	135	43	13	96.09	16.98	0.99	.	266:NM_144564,NM_213568:SLC39A3|
chr19	2865928	2866018	91	26	11	80.31	18.78	1.02	1405:NM_024967:ZNF556|	.
chr19	10172764	10172860	97	31	11	107.69	31.35	0.64	.	.
chr19	10806567	10806701	135	86	11	87.25	21.49	1.03	5545:NM_031209:QTRT1|	.
chr19	13552812	13553009	198	58	15	111.98	15.28	0.72	.	.
chr19	51163523	51163627	105	72	12	106.58	26.2	0.62	.	.
chr19	51899706	51899815	110	27	12	108.45	28.44	0.65	.	8606:NM_001161748,NM_030657:LIM2|5988:NM_001193623:LOC147646|
chr19	55728405	55728551	147	71	19	167.56	29.48	0	.	7678:NM_001161440,NM_002842:PTPRH|
chr19	58313749	58313906	158	33	17	145.97	28.98	0	.	.
chr2	2501326	2501472	147	85	11	92.39	21.71	0.88	.	.
chr2	19649935	19650067	133	91	11	97.24	21.71	1.02	.	.
chr2	33805048	33805185	138	45	13	101.03	25.3	1.09	.	.
chr2	41050089	41050214	126	32	14	111.89	26	0.71	.	.
chr2	48796345	48796483	139	69	11	75.33	20.06	0.84	.	.
chr2	48796571	48796702	132	48	11	84.05	19.65	1.25	.	.
chr2	49006652	49006786	135	35	11	80.31	24.76	1.02	.	.
chr2	55339054	55339165	112	45	11	61.65	11.98	1.46	.	.
chr2	62765971	62766119	149	71	12	89.85	23.6	0.77	.	.
chr2	65579204	65579296	93	48	12	110.42	21.83	0.67	.	.
chr2	67451938	67452074	137	57	12	90.53	25.13	0.81	.	.
chr2	144672490	144672592	103	80	12	103.96	21.4	0.6	.	.
chr2	176214740	176214909	170	90	12	85.41	14.11	0.96	.	.
chr2	176684937	176685148	212	151	15	86.13	16.18	1	.	.
chr20	29828167	29828268	102	31	15	60.74	11.74	1.61	.	.
chr20	29833148	29833299	152	89	20	80.83	14.95	1.07	.	.
chr20	34973025	34973174	150	74	11	78.02	16.38	0.92	.	.
chr20	41087420	41087551	132	70	11	61.65	10.09	1.46	.	.
chr20	44991894	44992129	236	47	15	66.9	10.82	1.07	.	−934:NM_015945,NM_173073,NM_173179:SLC35C2|
chr20	58342812	58342917	106	73	14	50.36	10.53	1.69	.	.
chr20	62958875	62958944	70	33	12	143.55	47.75	0	.	.
chr21	10086992	10087071	80	24	12	90.53	16.51	0.79	.	.
chr21	10780850	10781320	471	181	91	191.42	10.35	0	.	.
chr21	10781597	10781888	292	72	44	111.33	10.66	0.7	.	.
chr21	10820644	10820802	159	68	16	55.99	10.69	1.58	.	.
chr21	35433969	35434067	99	66	11	93.59	20.98	0.94	.	.
chr21	40721397	40721547	151	78	11	75.74	18.84	0.85	.	501:NM_004965:HMGN1|

chr3	8584015	8584182	68	102	16	89.06	18.36	0.75		1305:NM_185560:HERC1|
chr3	15116410	15116532	123	30	11	73.03	29.31	1.0		371:NM_022467:MAP4K3|
chr3	18515787	18515945	159	96	11	73.89	29.31	1.0		
chr3	18516350	18516461	112	28	13	97.61	19.31	0.97		
chr3	20334686	20334787	102	32	13	123.48	24.32	0		
chr3	20464141	20464268	128	68	11	85.52	16.38	0.96		
chr3	22765046	22765174	129	72	11	75.99	18.88	0.85		
chr3	37179355	37179507	153	66	13	77.86	13.76	0.92		
chr3	38398191	38398470	280	209	26	141.35	17.12	0	−9940:NM_005108:XYLB|	
chr3	39641045	39641181	137	68	12	85.83	20.06	0.98		
chr3	43838051	43838241	191	84	17	107.29	20.38	0.63		
chr3	128734050	128734200	151	61	13	106.59	30.57	0.62		
chr3	187387531	187387687	157	69	12	90.53	18	0.79		−513:NM_001048:SST|
chr4	16942427	16942548	122	87	12	90.53	16.78	0.79		
chr4	49114113	49114357	245	149	42	91.16	14.16	0.83		
chr4	63348398	63348547	150	65	11	91.42	26.05	0.84		
chr4	90013787	90013966	180	122	12	72.13	14.33	1.07		
chr4	99533935	99534152	218	132	23	132.43	13.99	0		
chr4	114745360	114745505	146	73	11	92.72	26.05	0.9		
chr4	123125633	123125737	105	51	12	83.16	14.59	1.18		
chr4	129691953	129692053	101	46	12	90.53	23.16	0.79		
chr4	129692062	129692229	168	86	13	63.92	13.48	1.14		
chr4	135285183	135285325	143	65	12	90.53	26.23	0.79		
chr4	166395801	166395913	113	41	11	93.97	20.69	0.94		
chr4	173415022	173415155	134	69	11	96.71	25.95	1.01		
chr4	183604156	183604277	122	55	13	126	39.07	0		
chr4	187615264	187615429	166	87	15	60.1	10.79	1.53		
chr5	1247105	1247251	147	85	13	78.91	13.48	0.96		
chr5	19122156	19122284	129	42	12	79.38	19.01	0.98		
chr5	19893738	19893901	164	85	12	87.25	24.69	1.03		
chr5	35873182	35873268	87	32	12	90.53	15.96	0.79		
chr5	42907147	42907258	112	29	12	70.13	13.69	0.92		
chr5	78838334	78838461	128	57	12	90.53	23.48	0.79		
chr5	82283114	82283242	129	33	12	102.14	18.49	1.14		
chr5	96979018	96979189	172	87	12	95.49	30.39	0.97		
chr5	130413821	130413936	116	82	11	78.28	15.28	0.93		
chr5	139780313	139780495	183	152	15	96.34	19.37	0.99	1086:NM_020690NM_001197030,NM_017747,NM_017978,NM_024668:A4	
chr5	149507482	149507609	128	80	12	92.63	21.4	0.89		
chr5	172145919	172146093	175	117	12	65.13	12.07	1.25		
chr5	174098577	174098697	121	70	12	99.14	26.2	1.08		
chr6	339643	339807	165	97	15	86.25	14.72	1		
chr6	26215728	26215919	192	88	15	64.86	11.89	1.19	9655:NM_003532:HIST1H3E|1420:NM_021052:HIST1H2AE|	−952:NM_003518:HIST1H2BG|
chr6	33480560	33480703	144	85	13	80.56	18.16	1.05		
chr6	36199960	36200049	90	40	11	91.71	20.96	0.86		
chr6	37471332	37471475	144	91	13	78.91	17.61	0.95		3776:NM_138493:CCDC167|
chr6	38944243	38944442	200	46	22	110.78	14.23	0.68		
chr6	40987638	40987756	119	27	11	79.77	13.49	1		
chr6	43023483	43023619	137	65	11	57.96	13.49	1.76	3889:NM_138343,NM_201521,NM_201522,NM_201523:KLC4|	1937:NM_001168370,NM_014780:CUL7|−3622:NM_015950:MRPL2|
chr6	47227648	47227794	147	87	11	61.65	11.23	1.47		
chr6	49082920	49083027	108	33	11	61.65	11.55	1.47		
chr6	49435653	49435778	126	56	12	62.29	13.1	1.44	−4557:NM_018132:CENPQ|	4738:NM_000255:MUT|
chr6	50513516	50513842	327	232	27	146.43	21.72	0		
chr6	57034776	57034906	131	81	12	70.13	14.02	0.92	2328:NM_004282:BAG2|	
chr6	84559590	84559726	137	84	11	73.19	18.88	1.13	9780:NM_016230:CYB5R4|3395:NM_001009994:RIPPLY2|	
chr6	90605137	90605297	161	87	13	96.44	27.17	1	−949:NM_032602:GJA10|	
chr6	90611205	90611336	132	88	12	56.37	10.59	1.63	−7017:NM_032602:GJA10|	
chr6	135462571	135462749	179	26	12	70.13	13.68	0.92		
chr6	142410064	142410285	222	165	14	90.03	20.84	0.78		350:NM_002511:NMBR|
chr7	30711536	30711787	252	178	20	154.39	25.01	0		
chr7	64126544	64126681	138	67	11	70.44	12.66	0.94	−33:NM_001013746,NM_016220:ZNF107|	
chr7	72882751	72882897	147	58	13	101.49	20.38	1.14		
chr7	100701825	100701994	170	62	13	80.91	16.18	1.08		
chr7	105750346	105750463	118	33	13	89.23	19.65	0.76		−2593:NM_006754,NM_182715:SYPL1|
chr7	108247705	108247875	171	77	13	64.78	11.96	1.18		
chr7	120055328	120055450	123	59	11	80.31	16.92	1.02		
chr7	120750945	120751200	256	66	27	139.21	14.11	0		
chr7	128687683	128687824	142	80	12	70.13	13.02	0.92		−7402:NM_001191028,NM_012470:TNPO3|
chr7	128934494	128934639	146	62	11	70.8	19.11	0.99		
chr7	130495379	130495587	209	29	14	71.77	11.98	1.05		
chr7	132001929	132002385	457	90	50	366.81	22.29	0		
chr7	132750586	132750723	138	104	11	68.4	15.08	1.15		
chr7	138085294	138085386	93	65	12	102.75	29.67	1.15		
chr7	138085441	138085593	153	79	12	56.37	10.52	1.63		
chr7	147700410	147700564	155	31	17	135.03	24.45	0		
chr8	37855469	37855620	152	64	11	73.87	21.26	1.16		
chr8	53624207	53624318	112	64	11	77.39	12.07	0.91		−2707:NM_001083617,NM_014781:RB1CC1|
chr8	80795754	80795871	118	87	12	110.82	27.91	0.68		
chr8	96224042	96224186	145	76	11	61.65	11.08	1.46		
chr8	106891294	106891455	162	70	17	116.83	29.71	0		
chr8	120350535	120350678	144	95	11	93.39	21.71	0.93		
chr8	125691717	125691830	114	50	11	80.31	15.68	1.02		
chr8	125691900	125692133	234	70	15	77.21	14.79	0.9		
chr9	71335971	71336255	285	216	25	160	14.33	0		
chr9	71879957	71880100	144	51	12	104.79	30.39	0.6		
chr9	94547018	94547253	236	84	20	92.25	12.62	0.88		
chr9	101715656	101715844	189	96	13	72.27	12.74	1.07	−9518:NM_001855:COL15A1|	
chr9	114345859	114345945	87	20	11	111.81	22.93	0.71		
chr9	116310179	116310329	151	66	12	80.89	20.06	1.07		
chr9	127799707	127799787	81	33	11	101.17	22.93	1.12		
chr9	128368105	128368214	110	36	11	61.65	15.13	1.46		
chr9	132552715	132553119	405	209	44	163.59	11.14	0		
chr9	134674677	134674817	141	86	13	91.05	16.18	0.82		
chr9	134989573	134989704	132	33	15	111.98	15.28	0.72		
chr9	138276599	138276819	221	189	26	223.01	36.39	0		
chr9	138276995	138277204	210	180	17	116.83	19.26	0		
chr9	138809239	138809348	110	30	11	83.49	18.88	1.2		
chr9	140779125	140779271	147	77	13	63.92	11.8	1.14	−6884:NM_000718,NM_001243812:CACNA1B|	
chrX	48829583	48829716	134	67	12	83.87	18.88	1.23		1466:NM_004979:KCND1|
chrY	13837603	13837715	113	68	11	78.48	14.95	0.95		

Common (sgRNA #6 and #17)
Target site

Number	Name	Sequence (5' to 3')	Description
27420	hHS5-TAL-Target-F	ccagtttctccagtttccctttt	enChIP assays in Figure 2C (5'HS5)
27421	hHS5-TAL-Target-R	ttttcaaaatgcaaggtgatgtc	enChIP assays in Figure 2C (5'HS5)
27222	hSox2-prom-F	attggtcgctagaaacccatttatt	enChIP assays in Figure 2C (Sox2)
27223	hSox2-prom-R	ctgccttgacaactcctgatacttt	enChIP assays in Figure 2C (Sox2)
27915	h5'HS5-3D-3'-ISceI-F	(Biotin)ccggTAGGGATAACAGGGTAATttgagaaggtagggttgcatgag	1 st PCR in 3C-based assay in Figure 5 (Upper cases: I-Sce I site)
27880	hZNF670-3D-F2	gagtcttggactcgggctca	1 st PCR in 3C-based assay in Figure 5
27863	hZNF670-3D-F	atctttggggtgaagttcccttt	Nested PCR in 3C-based assay in Figure 5
27864	hZNF670-3D-R	ccaagagatctggctgctaaaca	Nested PCR in 3C-based assay in Figure 5
27771	hZNF670-ZNF695-Ex1_2-F	agcttctgttgctctgggacct	RT-PCR in Figure 6 (ZNF670-ZNF695)
27772	hZNF670-ZNF695-Ex1_2-R	cacatccctgaatgccaatagtc	RT-PCR in Figure 6 (ZNF670-ZNF695)
27779	hZNF670-Ex4-F2	gttccccaggaggatgtatgttc	RT-PCR in Figure 6 (ZNF670)
27780	hZNF670-Ex4-R2	acgagatgcagttaccacagagc	RT-PCR in Figure 6 (ZNF670)
27773	hMIR422A-F	gacttagggtcagaaggcctgag	RT-PCR in Figure 6 (MIR422A)
27774	hMIR422A-R	aagcttggctcagggacagag	RT-PCR in Figure 6 (MIR422A)
27782	hTMEM151A-Ex2-F	cttagatgcccgagtgattgttg	RT-PCR in Figure 6 (TMEM151A)
27783	hTMEM151A-Ex2-R	cttaggcttccacttcctccaaa	RT-PCR in Figure 6 (TMEM151A)
27777	hYIF1A-Ex1_2-F	agcagctgcctcattagtattcg	RT-PCR in Figure 6 (YIF1A)
27778	hYIF1A-Ex1_2-R	gaataaccaccgcttgtgtcatc	RT-PCR in Figure 6 (YIF1A)
27784	hCNIH2-Ex1_4-F	gcctccctcatcttctttgtcat	RT-PCR in Figure 6 (CNIH2)
27785	hCNIH2-Ex1_4-R	agtattctgggaccaccagcttc	RT-PCR in Figure 6 (CNIH2)

Figure S1

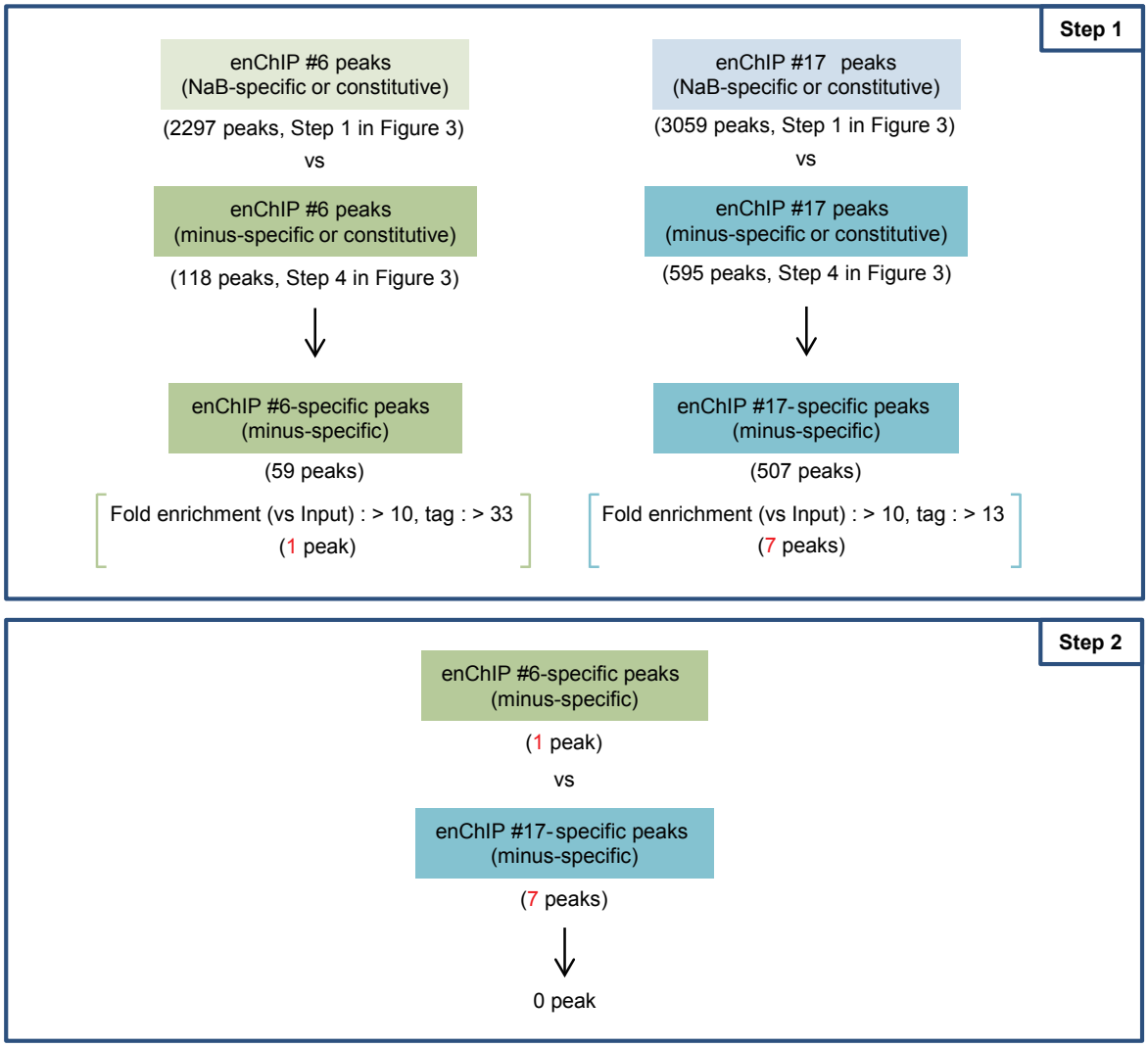


Figure S2