Identification of physical interactions between genomic regions by enChIP-Seq

1

16

E-mail: hodaka@biken.osaka-u.ac.jp

2 3 Toshitsugu Fujita, Miyuki Yuno, Yutaka Suzuki, Sumio Sugano, and Hodaka Fujii* 4 5 Toshitsugu Fujita, Miyuki Yuno, Hodaka Fujii 6 Chromatin Biochemistry Research Group, Combined Program on Microbiology and Immunology, 7 Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan 8 9 Yutaka Suzuki, Sumio Sugano 10 Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University 11 of Tokyo, Kashiwa, Chiba, Japan 12 Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University 13 of Tokyo, Minato-ku, Tokyo, Japan 14 15 *Corresponding author

ABSTRACT

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

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

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

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

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

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

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

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

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

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

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

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

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

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 ($\pm 10 \text{ kbp}$) 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.

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

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

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

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

editing and gene regulation [36]. 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.

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

serum (FCS).

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 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

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

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

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

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 ul) 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.

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

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 *β-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 β -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 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.

REFERENCES

355

- 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
- specific genomic regions. J Biosci Bioeng. 2009;108(5):446-9. doi:
- 369 10.1016/j.jbiosc.2009.05.005.
- 7. Fujita T, Fujii H. Direct idenification of insulator components by insertional chromatin
- immunoprecipitation. PLoS One. 2011;6(10):e26109. doi: 10.1371/journal.pone.0026109.
- 8. Fujita T, Fujii H. Efficient isolation of specific genomic regions by insertional chromatin
- 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.
- 9. Fujita T, Fujii H. Efficient isolation of specific genomic regions retaining molecular
- 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.

- 378 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:
- 380 10.1371/journal.pone.0116579.
- 381 11. Fujita T, Fujii H. Biochemical analysis of genome functions using locus-specific chromatin
- immunoprecipitation technologies. Gene Regul Syst Bio. 2015;in press.
- 383 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.
- 386 2013;439:132-6. doi: 10.1016/j.bbrc.2013.08.013.
- 387 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.
- 390 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:
- 392 10.1371/journal.pone.0103084.
- 393 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.
- 395 2015;10:e0123387. doi: 10.1371/journal.pone.0123387.
- 396 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.
- 399 17. Fujii H, Fujita T. Isolation of specific genomic regions and identification of their associated
- 400 molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation
- 401 (enChIP) using the CRISPR system and TAL proteins. Int J Mol Sci. 2015;16:21802-12.

- 402 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.
- 404 19. Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys2his2 zinc finger poroteins.
- 405 Annu Rev Biochem. 2001;70:313-40.
- 406 20. Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting.
- 407 Science. 2011;333(6051):1843-6.
- 408 21. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing
- 409 CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell.
- 410 2013;152:1173-83. doi: 10.1016/j.cell.2013.02.022.
- 411 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).
- 413 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.
- 415 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.
- 418 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.
- 422 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.

- 425 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.
- 427 28. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, et al. Chromosome
- 428 Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping
- interactions between genomic elements. Genome Res. 2006;16:1299-309.
- 430 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.
- **432** 2011;286:17870-8.
- 433 30. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the
- 434 CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32:670-6. doi:
- 435 10.1038/nbt.2889.
- 436 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:
- 438 10.1038/nbt.2916.
- 439 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.
- 441 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.
- 444 2015;43:3389-404. doi: 10.1093/nar/gkv137.
- 445 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
- genome organization: contrasting views from chromosome conformation capture and
- 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
- 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
- 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
- characterization of transcriptional start sites in humans by integrative transcriptome analysis.
- 459 Genome Res. 2011;21:775-89.
- 40. Seki M, Masaki H, Arauchi T, Makauchi H, Sugano S, Suzuki Y. A comparison of the Rest
- complex binding patterns in embryonic stem cells and epiblast stem cells. PLoS One.
- 462 2014;9(4):e95374.

- 41. Fujita T, Ryser S, Tortola S, Piuz I, Schlegel W. Gene-specific recruitment of positive and
- negative elongation factors during stimulated transcription of the MKP-1 gene in
- 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.

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

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

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

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 (\pm 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 \pm 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.

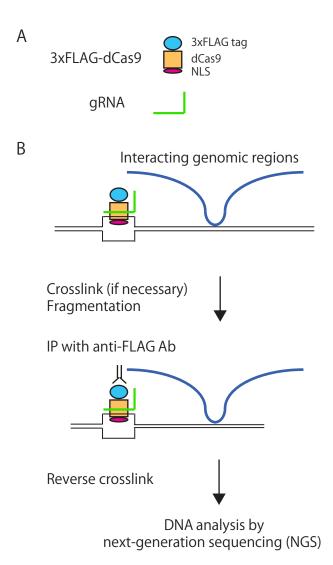
527

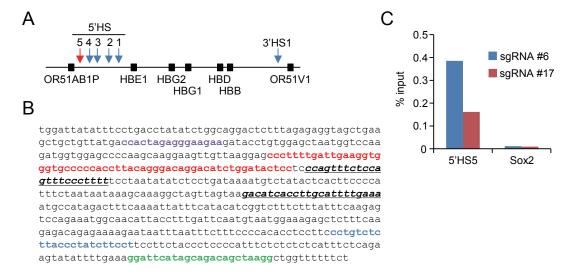
sgRNA	Condition	Fold enrichment (compared with input gDNA)	Number of tags
#6	minus (mock)	120	655
#0	NaB	168	783
Д17	minus (mock)	65	273
#17	NaB	122	217

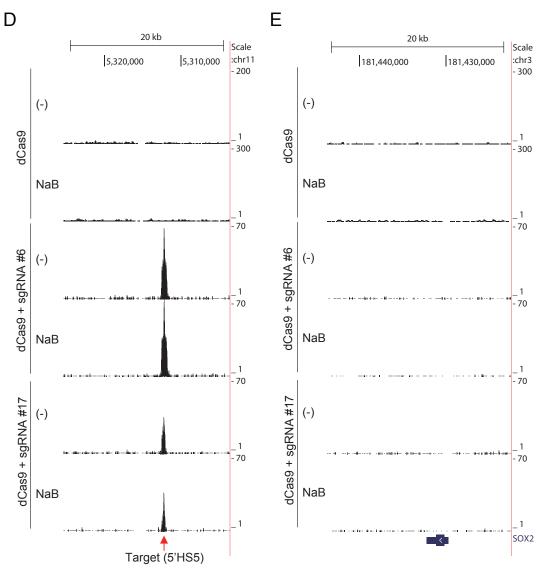
Table 2. List of peak positions detected in common by both sgRNA #6 and #17.

			sgRNA #6	6 (NaB)									
	chr	start	end	length	tags	fold enrichment	chr	start	end	length	tags	fold enrichment	status
Ī	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
5	30												

Figure 1







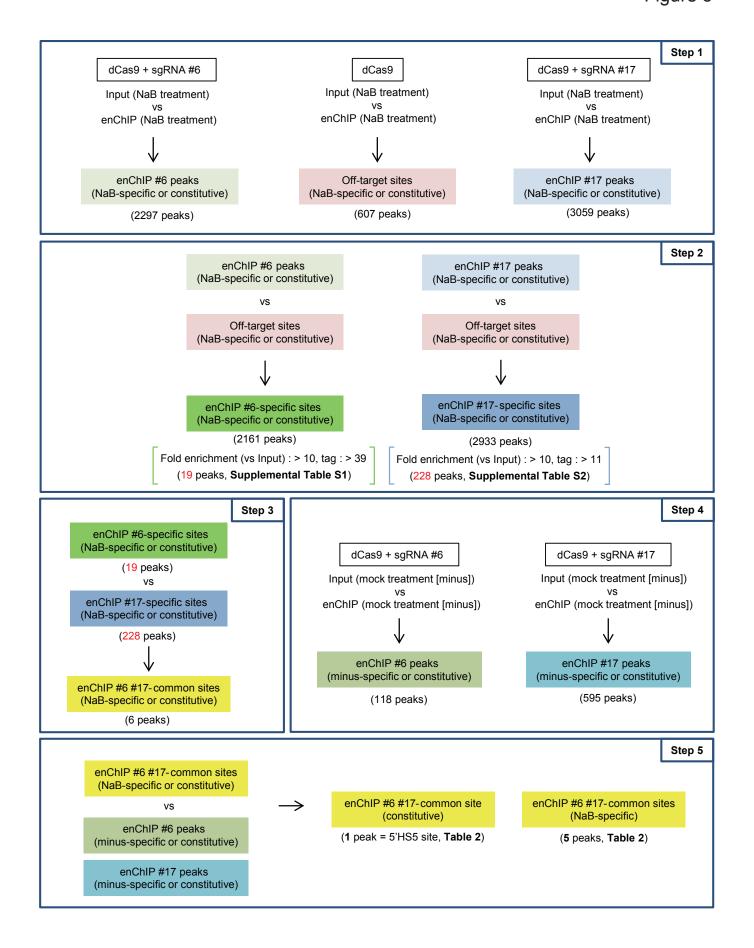


Figure 4

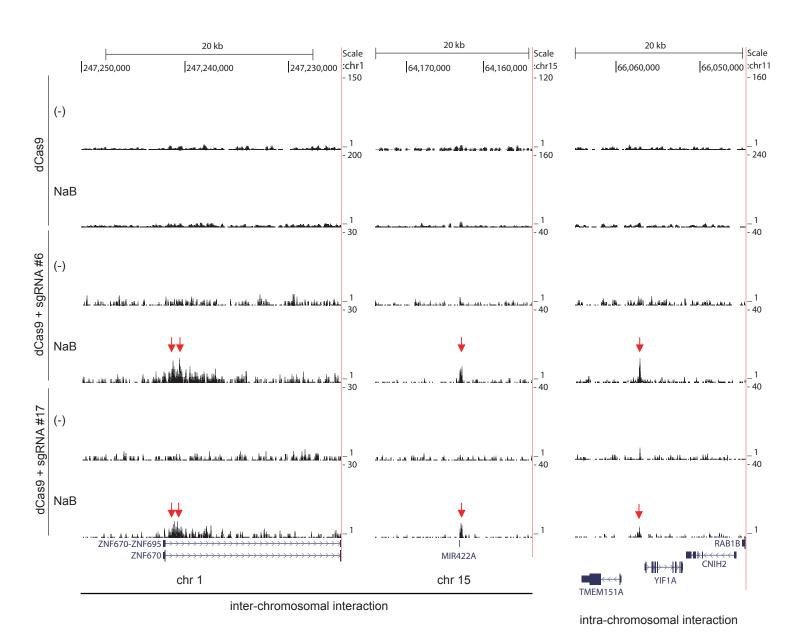


Figure 5

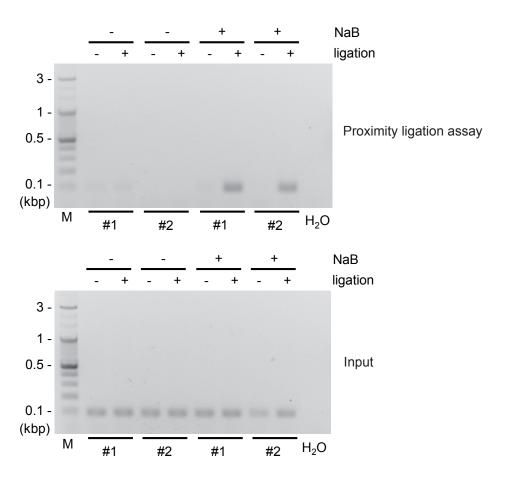


Figure 6

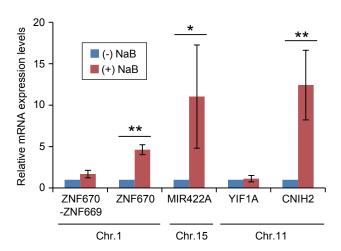


Table S1. List of genomic regions detected in the differentiated state (sgRNA #6) by enChIP-Seq.

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

Common (sgRNA #6 and #17)
Target site

1	certified by pe chr¹	11919466	/I@W) I	5-th 201	e∂l	1 th (or/funder, V	Mho.has 21.5	-gra	nted bioRxiv a license to display the preprint in perpetuity. It is made ava -BY 4.0 International license. [575,888,002521.98PPB]
1	chr1			151 115	5	1 11	61.36 61.65			
1966	chr1	16840452	16840625	174	12	9 28	105.1	12.77	0.61	
20	chr1	26744588	26744690	103	5	6 11	90.74	26.45	0.81	-7319:NM_024674:LIN28A .
	chr1	47201499	47201650	152	6	4 12	87.34	19.65	1.04	
	chr1	65583464	65583638	175	3	B 13	79.36	13.49	0.98	
	chr1 chr1)
	chr1								0.74	
	chr1	115300292	115300460	169	7	7 13	81.86	10.92		
1. 1. 1. 1. 1. 1. 1. 1.	chr1	118448119	118448224	106	3.	2 11	83.6	18.24		
Column C	chr1	119098643	119098812	170	11-	4 16	88.27	17.54	0.72	
	chr1 chr1								1.62	
	chr1								1.15	-215.NM_001184891,NM_001184892.NM_025207,NM_201398:FLAD1 1d 9145.NM_001130040,NM_001130041,NM_001202859,NM_003029,NM_183001:SHC1
Column	chr1	165366980	165367170	191	9	6 14	83.95	16.03		
Control	chr1	182156036	182156175	140	6	6 12	84.73	25.79		
1	chr1	190281875	190281997	123	5.	2 12	79.99	16.05		
Page	chr1	206600619	206600731	113	2	9 12	98.07	20.98		-
Process	chr1 chr1									
Column	chr1								1.23	
Page	chr1	224659181	224659306	126	6	0 12	82.98	17.44	1.17	
Section Sect	chr1	230233766	230233849	84	3	0 15	97.35	14.12	1.02	
Part	chr1 chr1	233764914	233765193	280	17	1 23	117.81	15.28	0	
201	chr1		242123574	123	3				1.07	
1965	chr1	247239197	247239370	174	7:	9 12	61.3	11.08	1.52	-2744:NM_001204220.NM_033213:ZNF670
Series 1970	chr1	247240971	247241324	354	6	7 28	95.69	10.33	0.98	-790-NM_001204220,NM_033213:ZNF670
Series 1962	chr10	51531252	51531379	128	4	7 11	71.68	17.07	1.04	
	chr10	73341962	73342129	168	3.	2 12	75.86	11.54	0.85	
1906 1908 19	chr10 chr10									
1908	chr10 chr10								0	. 281:NM_018294:CWF19L1
1965 1965	chr11	2306298	2306439	142	6	6 12	70.13	13.02	0.92	
March Marc	chr11	10773231	10773367	137	4	3 11	80.96	22.29		
	chr11	47747635	47747752	118	7	1 16	86.11	18.34	0.99	
	chr11 chr11									
1971 1979 1979 1979 1979 19 18 18 70 8 18 19 19 19 19 19 19 19 19 19 19 19 19 19	chr11									
1985 1985	chr11	118797135	118797270	136	5	B 13	81.54	15.28		
	chr12	53135306	53135428	123	4	1 11	78.22	24.28	0.93	
	chr12	57042458	57042561	104	6	3 13	118.99	32.75	0	
	chr12 chr12								0.64	
179900 179800 179800 179800 179800 18 18 18 18 19 19 19 19	chr12 chr12								0.95	
Section Sect	chr12	117295952	117296132	181	10.	2 14	111.08	28.66	0.7	
	chr14	90961717	90961859	143	6	3 12	105.15	39.07	0.61	
1966 1969	chr15	64162762	64163061	300	16	9 31	237.77	27.51	0	-99/3/3ML/UU024/SERPINAS)
Page	chr16 chr16		12059000	128	4	4 11	89.81	21.49		91.NM_001192:TNFRSF17 .
Property	chr16 chr16									
1	chr16									
1897 1898	chr17	5144346	5144393	48	2-	4 12	220.54	52.09	0	. 6311:NM_207103:SCIMP
1987 1980 1980 1980 10 10 10 10 10 10 10	chr17	28456193	28456317	125	9.	5 11	87.68	16.73	1.06	
Sect	chr17 chr17									
Part	chr17 chr17									
Charle 77,200668 77,20068	chr17	56197642	56197819	178	9:	3 13	78.45			
Part	chr17	77290668	77290823	156	8	B 13	101.06	25.04	1.1	
Description 1985	chr18	21027513	21027636	124	7	6 11	55.2	11.46	1.72	5274:NM_003831:RIOK3 9712:NM_032933:TMEM241
	chr19 chr19	2865928	2866018	91	2	6 11	80.31	18.78	1.02	1405:NM_024967:ZNF556 .
Section Sect	chr19 chr19									
September Sept	chr19	13552812	13553009	198	5	8 15	111.98	15.28	0.72	
Decision	chr19	51899706	51899815	110	2	7 12	108.45	28.44	0.65	8606:NM_001161748.NM_030657:LIM2[5988:NM_001193623:LOC147646]
	chr19	58313749	58313906	158	3	3 17	145.97	28.98	0	
Second	chr2 chr2	19649935	19650067	133	9	1 11	97.24	21.71	1.02	
chr2 4978545 4978645 39 69 11 75.33 20.06 0.84 c	chr2 chr2				4	5 13		25.3	1.09	
chr2 4900652 4900758 35 35 11 80.31 2.476 102	chr2	48796345	48796483	139	6	9 11	75.33	20.06	0.84	
ber 2 657967 65796 4579204 6579204 6579204 6579204 6579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 4579204 657920 457	chr2	49006652	49006786	135	3	5 11	80.31	24.76	1.02	
che2 67481988 67482074 137 7 12 90.33 25.30 0.81 che2 14672400 1467250 14782400 1478220 1478220 14782400 1479220 12 85.41 14.11 0.96 che2 178624470 14685400 170 90 12 85.41 14.11 0.96 che2 178664937 17865400 12 15 86.13 16.18 1 . che2 29823160 28832300 152 80 20 80.38 14.55 107 che2 3497305 3497316 150 70 11 78.02 16.83 0.92 che2 449840 4409730 3497305 3497316 15 66.85 10.00 1.46 che2 449840 4499210 450 15 66.9 10.25 169 che2 53482817 5842917 10 73 14 50.36 10.53 169	chr2	62765971	62766119	149	7	1 12	89.85	23.6	0.77	
1467240	chr2 chr2	67451938	67452074		5	7 12	90.53			
chr2 17868148 212 151 15 15 88 13 16.18 1	chr2	144672490	144672592	103	8	0 12	103.96	21.4	0.6	
Debruik Debrui	chr2	176684937	176685148	212	15	1 15	86.13	16.18	1	
phr20 4 1607420 4 1607551 132 70 11 6 165 10.09 1.46	chr20	29833148	29833299	152	8	9 20	80.83	14.95	1.07)
chr20 4498 Mg 4498 228 Mg 128	chr20 chr20	41087420	41087551	132	7	0 11	61.65	10.09	1.46	
chr20 6298875 6298874 70 33 12 143.55 47.75 0 c	chr20 chr20			236	4	7 15				
chv21 10780850 10781320 471 181 91 191.42 10.35 0 chv21 10781967 10781888 292 72 44 111.33 10.66 0.7 chv21 10820644 108206802 159 68 16 55.99 10.69 1.58 chv21 3843969 38434067 99 66 11 93.38 20.98 0.34	chr20 chr21	62958875	62958944	70	3	3 12	143.55	47.75	0	
chr21 10820644 10820802 159 68 16 55.99 10.69 1.58	chr21	10780850	10781320	471	18	1 91	191.42	10.35	0	
	chr21	10820644	10820802	159	6	8 16	55.99	10.69	1.58	
chr21 40721397 40721547 151 78 11 75.74 18.84 0.85 501:NM_004965:HMGN1	chr21	35433969 40721397						20.98 18.84	0.94	

"BioRxiv prepr	rint döj	 	5110532	:// <mark>5</mark>	oi.ö	 g/	10.1101/03 6	160 g	his"	ersion posted January 7, 2016	The preprint in perpetuity. It is made ava
certified by pe	eerste	vie:	W)84b	s th	e ai	the	or/funderw	ho has	qra	nted bioRxiv a license to displ	ay the preprint in perpetuity. It is made ava
chr3	18516350) 18	8516461	112	28	13	97.61	19.31	2(1.0	C-BY 4.0 International license.	· · · · · · · · · · · · · · · · · · ·
chr3	20334686		0334787	102				24.32		0.	
chr3	20464141		0464268	128				16.38			
hr3	22765046 37179355		2765174 7179507	129 153				18.88			•
hr3	38398191		8398470	280				17.12		z. D =9940:NM_005108:XYLB	•
hr3	39641045		9641181	137				20.06			•
hr3	43838051		3838241	191				20.38			•
hr3	128734050		8734200	151				30.57			•
r3	187387531		7387687	157				18			-513:NM_001048:SST
r4	16942427		6942548	122				16.78			
r4	49114113		9114357	245				14.16		3.	
r4	63348398	63	3348547	150			91.42	26.05	0.8	4 .	
r4	90013787	7 90	0013966	180	122	12		14.33	1.0	7.	
4	99533935	5 99	9534152	218	132	23	132.43	13.99		0.	
4	114745360		4745505	146				26.05			
r4	123125633		3125737	105				14.59			
4	129691953		9692053	101				23.16			
4	129692062		9692229	168				13.48	1.1		
1	135285183		5285325	143				26.23	0.7		
4	166395801 173415022		6395913 3415155	113				20.69 25.95	0.9		
4	173415022		3415155 3604277	134				25.95 39.07			
1	183604156		7615429	122				39.07 10.79			
5	1247105		1247251	147				13.48			
5	19122156		9122284	129				19.01	0.9		
5	19893738		9893901	164				24.69	1.0		
5	35873182		5873268	87				15.96	0.7		
5	42907147		2907258	112				13.69			
5	78838334		8838461	128			90.53	23.48			
5	82283114	\$ 82	2283242	129	33	12	102.14	18.49	1.1	4 .	
5	96979018	96	6979189	172	87	12	95.49	30.39	0.9	7.	
5	130413821		0413936	116	82	- 11		15.28	0.9		
5	139780313		9780495	183				19.37		9 1086:NM_020690NM_001197030,NM_017747,NM_017978,NM_024668:AI	
5	149507482		9507609	128				21.4			
r5	172145919		2146093	175				12.07			•
r5	174098577		4098697	121				26.2		8.	•
r6 r6	339643 26215728		339807 6215919	165 192				14.72		9 9655:NM_003532:HIST1H3E[1420:NM_021052:HIST1H2AE]	- -952:NM_003518:HIST1H2BG
6	33480560		3480703	144				18.16			-332:NM_003318:RIS11R2BQ
6	36199960		6200049	90				20.96			•
r6	37471332		7471475	144				17.61			- 3776:NM_138493:CCDC167
r6	38944243		8944442	200				14.23			
r6	40987638		0987756	119				13.49		1.	
6	43023483		3023619	137				13.49		6 3889:NM_138343,NM_201521,NM_201522,NM_201523:KLC4	1937:NM_001168370,NM_014780:CUL7 -3622:NM_015950:MRPL2
6	47227648		7227794	147				11.23		7.	
6	49082920		9083027	108				11.55	1.4	7.	
6	49435653	3 49	9435778	126		12	62.29	13.1	1.4	4 -4557:NM_018132:CENPQ	4738:NM_000255:MUT
r6	50513516		0513842	327				21.72		0.	
r6	57034776		7034906	131				14.02		2 2328:NM_004282:BAG2	
ır6	84559590		4559726	137				18.88		3 9780:NM_016230:CYB5R4 3395:NM_001009994:RIPPLY2	
r6	90605137		0605297	161				27.17		1 -949:NM_032602:GJA10	
r6	90611205		0611336	132				10.59		3 -7017:NM_032602:GJA10	
r6	135462571		5462749	179				13.68			
r6	142410064		2410285	222	165			20.84		8.	350:NM_002511:NMBR
r7 r7	30711536 64126544		0711787 4126681	252 138				25.01 12.66	0.0	0 . 4 –33:NM_001013746,NM_016220:ZNF107	
7	72882751		2882897	147				20.38			
7	100701825		0701994	170				20.38	1.0		
7	105750346		5750463	118				19.65	0.7		- -2593:NM_006754,NM_182715:SYPL1
7	108247705		8247875	171	77			11.96	1.1		
7	120055328		0055450	123				16.92			
7	120750945		0751200	256				14.11			
7	128687683		8687824	142				13.02			-7402:NM_001191028,NM_012470:TNPO3
7	128934494	1 128	8934639	146				19.11			
1	130495379		0495587	209	29	14		11.98	1.0	5.	
1	132001929		2002385	457	90	50	366.81	22.29		0	
1	132750586		2750723	138			68.4	15.08			
7	138085294		8085386	93				29.67			
7	138085441		8085593	153				10.52			
7	147700410		7700564	155				24.45			
8	37855469		7855620	152				21.26			07074NA 001000017 NA 014701 DD1 001
8	53624207		3624318	112				12.07	0.9		-2707:NM_001083617,NM_014781:RB1CC1
8	80795754		0795871	118				27.91	0.6		
8	96224042 106891294		6224186 6891455	145 162				11.08 29.71	1.4		
3	106891294		0350678	162							
	120350535		5691830					21.71			
8	125691717		5691830 5692133	114 234		11		15.68			
)	71335971		1336255	285				14.79		0	
<u> </u>	71879957		1880100	144				30.39		6.	
9	94547018		4547253	236				12.62			
)	101715656		1715844	189		13		12.74		7 –9518:NM_001855:COL15A1	
9	114345859		4345945	87				22.93			
19	116310179		6310329	151				20.06			
	127799707		7799787	81				22.93			
			8368214	110				15.13			
r9								11.14			
r9 r9	128368105 132552715		2553119	405	208	44	100.08	11.14			
r9 r9 r9	128368105	132	2553119 4674817	405 141				16.18			
r9 r9 r9	128368105 132552715	5 132 7 134			86	13 15	91.05 111.98		0.8	2 .	
9 9 9 9 9	128368105 132552715 134674677 134989573 138276599	5 132 7 134 3 134 9 138	4674817 4989704 8276819	141 132 221	86 33 189	13 15 26	91.05 111.98 223.01	16.18 15.28 36.39	0.8	2 . 2 .	
	128368105 132552715 134674677 134989573 138276599 138276995	5 134 7 134 3 134 9 138 5 138	4674817 4989704 8276819 8277204	141 132 221 210	86 33 189 180	13 15 26 17	91.05 111.98 223.01 116.83	16.18 15.28 36.39 19.26	0.8	2	
-9 -9 -9 -9 -9 -9 -9 -9	128368105 132552715 134674677 134989573 138276599	5 134 7 134 3 134 9 138 5 138	4674817 4989704 8276819	141 132 221	86 33 189 180 30	13 15 26 17	91.05 111.98 223.01 116.83 83.49	16.18 15.28 36.39	0.8 0.7	2	

Common (sgRNA #6 and #1

Taid Rxiv Preprint doi:histpas//doi.org/10.1101/036160; this version posted January 7, 2016. The copyright holder for this preprint (which was not confided by peer review) is the author/typeer(s/who) has granted bioRxiv a license to display the preprint in perpetuity. It is made available under

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	gagtettggaetegggetea	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	agettetgttgetetgggacet	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	aagettggeteagggacagag	RT-PCR in Figure 6 (MIR422A)
27782	hTMEM151A-Ex2-F	cttagatgcccgagtgattgttg	RT-PCR in Figure 6 (TMEM151A)
27783	hTMEM151A-Ex2-R	cttaggettecaetteeteeaaa	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

