

Research Article

Deletion of transcription factor binding motifs using the CRISPR/spCas9 system in the β -globin LCR

Yea Woon Kim and AeRi Kim

Department of Molecular Biology, College of Natural Sciences, Pusan National University, Busan 46241, Republic of Korea

Correspondence: AeRi Kim (kimaeri@pusan.ac.kr)



Transcription factors play roles in gene transcription through direct binding to their motifs in genome, and inhibiting this binding provides an effective strategy for studying their roles. Here, we applied the CRISPR (clustered, regularly interspaced, short palindromic repeat)/spCas9 (CRISPR-associated protein 9) system to mutate the binding motifs of transcription factors. Binding motifs for erythroid-specific transcription factors were mutated in the locus control region (LCR) hypersensitive sites (HSs) of the human β -globin locus. Guide RNAs targetting binding motifs were cloned into lentiviral CRISPR vector containing the spCas9 gene, and transduced into MEL/ch11 cells carrying human chromosome 11. DNA mutations in clonal cells were initially screened by quantitative PCR (qPCR) in genomic DNA and then clarified by sequencing. Mutations in binding motifs reduced occupancy by transcription factors in a chromatin environment. Characterization of mutations revealed that the CRISPR/spCas9 system mainly induced deletions in short regions of <20 bp and preferentially deleted nucleotides around the fifth nucleotide upstream of Protospacer adjacent motifs (PAMs). These results indicate that the CRISPR/Cas9 system is suitable for mutating the binding motifs of transcription factors, and, consequently, would contribute to elucidate the direct roles of transcription factors.

Introduction

Transcription factors regulate the transcription of target genes by binding to motifs present in regulatory regions of the genome. The roles of transcription factors have been elucidated by inhibiting their expressions in a cellular environment [1,2] or by mutating their binding motifs in extrachromosomal loci [3-7]. Such expressional inhibition can demonstrate their direct roles, but cannot exclude indirect effects because transcription factors often affect the expressions of other transcription factors. Thus, mutation at binding motifs can better reveal the direct roles of transcription factors. However such mutations have not been commonly tried in a genome context because of technical difficulties in genome editing. Furthermore, binding motifs for different transcription factors are closely located in regulatory regions such as enhancers or promoters.

Recently, an effective genome editing tool was developed from bacterial adaptive immune system type II, clustered, regularly interspaced, short palindromic repeats (CRISPR) [8,9]. In the bacterial CRISPR system, CRISPR RNA (crRNA) hybridizes with *trans*-activating crRNA (tracrRNA) [10]. The crRNA-tracrRNA duplex directs CRISPR-associated protein 9 (Cas9) endonuclease into target sequences by base-pairing crRNA with DNA [11]. DNA cleavage by Cas9 requires a Protospacer adjacent motif (PAM) next to target sequences [8,11,12]. In the CRISPR/Cas9 system for genome editing, the crRNA-tracrRNA duplex has been simplified to a single guide RNA (sgRNA) [8,13], which usually contains 20 nts for base pairing with target sequences. sgRNA-directed Cas9 has been shown to generate double-stranded breaks in the target sequences [8,11]. These breaks are repaired in a non-homologous end joining manner, resulting in mutations in target sequences [13-15]. The CRISPR/Cas9 system has been used for knocking out genes by disrupting coding regions in various cell lines and organisms [16-22].

Received: 23 June 2017
Revised: 18 July 2017
Accepted: 19 July 2017

Accepted Manuscript Online:
20 July 2017
Version of Record published:
2 August 2017

spCas9 derived from *Streptococcus pyogenes* uses 5'-NGG-3' sequences as PAMs [8,23]. These sequences occur every eight bases in the human genome [13], which suggests that almost all the genome sequences can be cleaved by the spCas9. Genome targeting using spCas9 gives rise to the deletion of <10 bp in most cases and the insertion or substitution of short sequences in low frequency [16,24]. Here, we applied the CRISPR/spCas9 system to mutate the binding motifs of transcription factors in a genome context, because these motifs are typically ~10 bp [25]. The locus control region (LCR) hypersensitive sites (HSs) of the human β -globin locus were used as model regulatory regions, which contains binding motifs for tissue-specific transcription factors such as GATA-1, TAL1, and KLF1 [26]. To clearly analyze changes by genome editing, we employed mouse MEL/ch11 cells that contain a human chromosome 11 where the β -globin locus is present [27]. We found that the CRISPR/spCas9 system can effectively mutate the binding motifs of transcription factors in a genome context and believe that it will contribute to studies on the direct roles of transcription factors.

Materials and methods

Cloning guide sequences of sgRNA into CRISPR vectors

Guide sequences of sgRNA were cloned into lentiCRISPRv2 vector (Addgene #52961) [18] as suggested by the manufacturer, but with modifications in the digestion and dephosphorylation steps. To use the BsmBI site for cloning in lentiCRISPRv2 vector, 'CACCG' were added to the 5' end of oligonucleotides for guide sequences, and 'AAAC' and 'C' were added to the 5' and 3' ends of oligonucleotides for complementary sequences, respectively. A pair of oligonucleotides (10 nM) was phosphorylated with T4 polynucleotide kinase (NEB M0201S) at 37°C for 1 h in 10 μ l of reaction volume, annealed by heating to 95°C for 5 min and cooling to 25°C at 0.5°C/min, and then diluted to 1:1000 in sterile water. One microgram of lentiCRISPRv2 vector was digested and dephosphorylated with FastDigest BsmBI (Thermo Scientific FD0454) and FastAP (Thermo Scientific EF0651) at 37°C for 30 min in 20 μ l of reaction volume. Fifty nanograms of digested vector and 1 μ l of diluted oligonucleotide duplex were ligated with Quick T4 DNA ligase (NEB M2200S) at 25°C for 20 min in 11 μ l of reaction volume. Ligated plasmid vectors were introduced into competent Stbl3 bacteria. For cloning into pLH-spsgRNA2 vector (Addgene #64114) [28], oligonucleotides were synthesized by adding 'ACCG' to the 5' end of target sequences and 'AAAC' to 5' end of complementary sequences. This pair of oligonucleotides was phosphorylated and annealed as described above, and then cloned into BbsI (NEB R0539S) site in pLH-spsgRNA2 vector. Ligated plasmid vectors were introduced into competent Stbl3 bacteria. All plasmid vectors were prepared using Plasmid Mini Kits (Qiagen). The EGFP sequences (5'-GGGCGAGGAGCTGTTACCG-3') were used as control guide sequences [18].

Cell culture and lentiviral transduction

MEL/ch11 and 293FT cells were cultured in DMEM medium (Gibco) supplemented with 10% FBS (Gibco) and 1% penicillin-streptomycin (Gibco). To produce lentivirus, 0.5 μ g lentiviral vector and 1.5 μ g packaging mix (pLP1, pLP2, and pLP/VSVG) were diluted in 125 μ l Opti-MEM (Gibco), and 6 μ l Lipofectamine 2000 (Invitrogen) in another 125 μ l Opti-MEM. After incubation for 5 min at room temperature, they were mixed and incubated for an additional 20 min. The DNA-Lipofectamine 2000 mixture was added to 1×10^6 293FT cells in six-well plate and medium containing the mixture was changed with complete medium the next day. At 72 h after transfection, lentiviral supernatants were harvested and filtered through a 0.45- μ m membrane. MEL/ch11 cells were infected with lentiviral supernatants in the presence of 6 μ g/ml polybrene (Sigma) and medium was replaced with fresh one the next day. Cells were selected using 2 μ g/ml puromycin (Sigma) for lentiCRISPRv2 vector or 500 μ g/ml hygromycin (Invitrogen) for pLH-spsgRNA2 vector at 72 h after transduction.

Screening of mutant cell clones

Clonal cells were grown in 96-well plates after antibiotics selection, and then transferred to 24-well plates when the cell numbers reached 1×10^5 per well. To screen clones with mutations in sequences targeted by sgRNA, genomic DNA was isolated using a QIAamp DNA Mini Kit (Qiagen) from 1×10^6 cells, and amplified by quantitative PCR (qPCR) using the following parameters: 20 ng genomic DNA, 4.5 μ M forward primer, 4.5 μ M reverse primer, and $1 \times$ Power SYBR Green PCR master mix (Applied Biosystems) in 10 μ l of reaction volume using the 7300 real-time PCR system (Applied Biosystems). Dissociation curves of qPCR were analyzed using qPCR program and PCR products were visualized on agarose gels. Mutations were clarified by DNA sequencing analysis. The sequences of primers used for PCR are presented in Supplementary Table S1.

ChIP

Clonal MEL/ch11 cells were cultured at 1.5×10^5 cells per ml in 5 mM HMBA (Sigma) for 72 h to activate the β -globin locus. To cross-link protein to genomic DNA, 1×10^7 cells were incubated in culture medium containing 1% formaldehyde (Sigma) at 25°C for 10 min with shaking, and then glycine (Invitrogen) was added to a final concentration of 0.125 M. After PBS washing, cross-linked cells were incubated in cell lysis buffer (10 mM Tris, 10 mM NaCl, 0.2% NP-40, pH 8.0) at 4°C for 10 min and centrifuged at 600 g for 5 min to isolate nuclei. The nuclei were treated with 100 U of MNase (Worthington) at 37°C for 15 min, incubated in nuclei lysis buffer (50 mM Tris, 10 mM EDTA, 1% SDS) at 4°C for 10 min and then sonicated to digest chromatin to mainly mononucleosomes. After pre-clearing with protein G agarose beads (Millipore), chromatin was made to react with antibodies at 4°C for 3 h, and protein–DNA complexes were collected using protein G agarose beads. DNA was purified by phenol extraction and ethanol precipitation and then dissolved in 150 μ l of Tris/EDTA buffer. DNA obtained by ChIP was analyzed by qPCR using TaqMan chemistry and the following parameters: 2 μ l DNA, 2 μ M TaqMan probe, 4.5 μ M primers, and 1 \times TaqMan universal master mix II (Applied Biosystems) in 10 μ l of reaction volume using the 7300 real-time PCR system (Applied Biosystems). Antibodies used for ChIP were normal goat IgG (sc-2028), GATA-1 (sc-1233), and TAL1 (sc-12984) from Santa Cruz Biotechnology, and KLF1 (ab2483) from Abcam. The sequences of probes and primers used for ChIP assay are presented in Supplementary Table S2.

Results and discussion

Application of the CRISPR/spCas9 system to the mutation of the binding motifs of transcription factors in the human β -globin LCR HSs

The human β -globin LCR HSs, which act as enhancers for the β -like globin genes, contain binding motifs for erythroid-specific transcription factors, GATA-1, TAL1, and KLF1 [26]. To mutate these motifs in the LCR HS2 and HS3, we obtained candidate guide sequences for sgRNA in the CRISPR/spCas9 system using an online tool (<http://crispr.mit.edu>) and chose that their putative cutting sites are located in binding motifs (Figure 1A,B). This tool provides scores for guide sequences that are inversely related to the possibility of binding to off-target DNA. The scores of the chosen guide sequences exceeded 70, except for sequences targeting the KLF1 motif in LCR HS3 (HS3_KL_R1) (Figure 1B). Next, we cloned the guide sequences into lentiCRISPRv2 vector and transduced them into MEL/ch11 cells using lentiviral system (Figure 1C), because suspension cells, like MEL cells, are inefficiently transfected using reagents such as Lipofectamine. Transduced cells were selected using puromycin, diluted, and grown in 96-well plates to obtain clonal cells.

Screening of mutations generated by the CRISPR/spCas9 system

It is important to screen cell clones that are desirably mutated in target sequences by the CRISPR/spCas9 system. For initial screening, genomic DNA was purified and analyzed by PCR. Primer sets for PCR were designed for guide sequences (A2F/R) and both side regions (A1F/R, A3F/R) (Figure 2A). Figure 2B shows the results of qPCR for wild-type MEL/ch11 cells and clones in which the LCR HS3 GATA-1 motif (HS3 Δ GA) or the TAL1 motif (HS3 Δ TA) was targeted by guided spCas9. Genomic DNA was similarly amplified using A1 or A3 primer sets in wild-type cells and HS3 Δ GA and HS3 Δ TA clones. However, A2 primer set did not amplify DNA in HS3 Δ GA and HS3 Δ TA clones, indicating that mutations occurred in regions annealed by A2 primers (Figure 2B and Supplementary Figure S1A).

To screen clones in which the LCR HS3 KLF1 motif was targeted, A2F-1 primer was designed in the outside region of target sequences because of high GC content in target sequences. Deletions in target sequences were detected by comparing sizes of PCR products (Figure 2C) and by comparing dissociation curves of qPCR (Supplementary Figure S1B). The reduction in PCR product sizes resulted in changes in dissociation curves. Finally mutated regions were identified by DNA sequencing using A4F/R primers. Sequencing results showed that 22 and 12 bp, which include a GATA-1 binding motif, were deleted in HS3 Δ GA #1 and #2 clones, respectively (Figure 2D). Mutations of the TAL1 and KLF1 motifs were sequenced in HS3 Δ TA and HS3 Δ KL clones. These results show that binding motifs for transcription factors can be mutated by sgRNAs of the CRISPR/spCas9 system in a genome context.

On the other hand, we found some clones in which genomic DNA was not amplified by the A1 or A3 primer set in addition to A2 primer set (Figure 2E). It indicates that large regions, including target sequences, were deleted by the attack of guided-spCas9. To determine the approximate lengths of deleted regions, we performed PCR using A4F/R primers, which cover large regions including A1, A2, and A3 amplicons, and ran PCR products in 0.8% agarose gel. It was found that deleted regions had lengths up to 400 bp. Using the same strategy, clones with mutations in the GATA-1 or KLF1 motifs of LCR HS2 were screened and sequenced (Supplementary Figure S2).

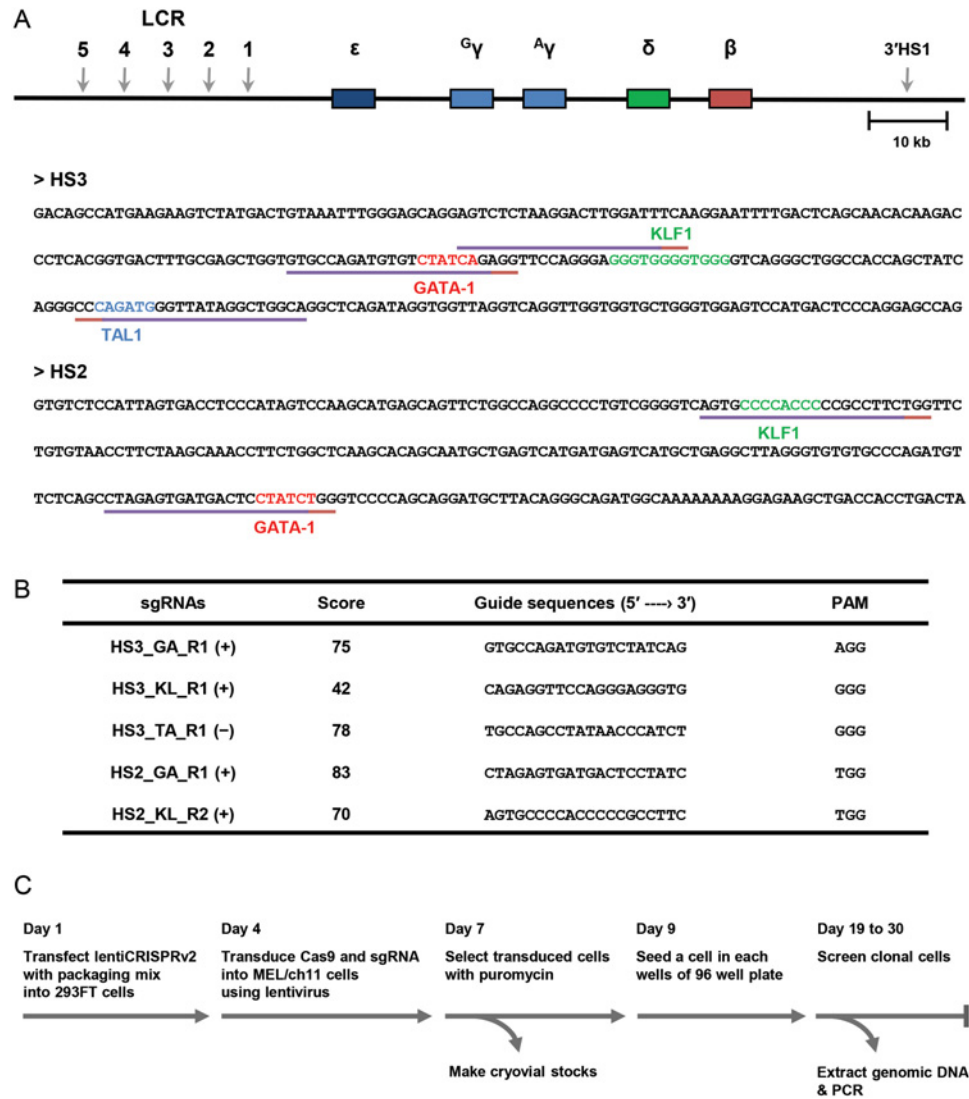


Figure 1. The CRISPR/spCas9 strategy for mutating binding motifs for transcription factors

(A) The human β -globin locus is presented. Vertical gray arrows indicate DNase I HSs in the LCR and 3'-HS1. The five globin genes, ϵ , $G\gamma$, $A\gamma$, δ , and β are represented by rectangle in sequence. DNA sequences are for the LCR HS3 and HS2, and the binding motifs of transcription factors are marked by colored bases; GATA-1 motif: red, KLF1 motif: green, and TAL1 motif: blue. Target sequences for sgRNA are indicated by purple lines with PAMs in red lines. (B) Guide sequences of sgRNA are listed with names, scores, and PAM sequences, where (+) and (-) mean guide sequences are positive or negative strand DNA, respectively. (C) Experimental procedure for mutating binding motifs for transcription factors in MEL/ch11 cells is illustrated and described in 'Materials and methods' section.

The mutations of target sequences can be analyzed using various methods. Mismatch cleavage assay is most commonly used to analyze genome edited by CRISPR/Cas9 [29]. This assay uses heteroduplex DNA that is formed by denaturing and annealing PCR products and digests it using Surveyor nuclease or T7E1 nuclease [29]. However, this assay cannot be used for haploid cells or single chromosomes, such as sex chromosomes, and does not provide information about mutated regions. In contrast with the mismatch cleavage assay, analysis using qPCR can provide information about the locations and approximate lengths of deletions. In addition, many clones can be screened and analyzed in one time by qPCR. Even deletion and/or insertion in a couple of base pair can be missed out in genome screening using qPCR and agarose gel running, this qPCR strategy still provides a simple and efficient means for screening mutant clones for transcription factor binding motifs.

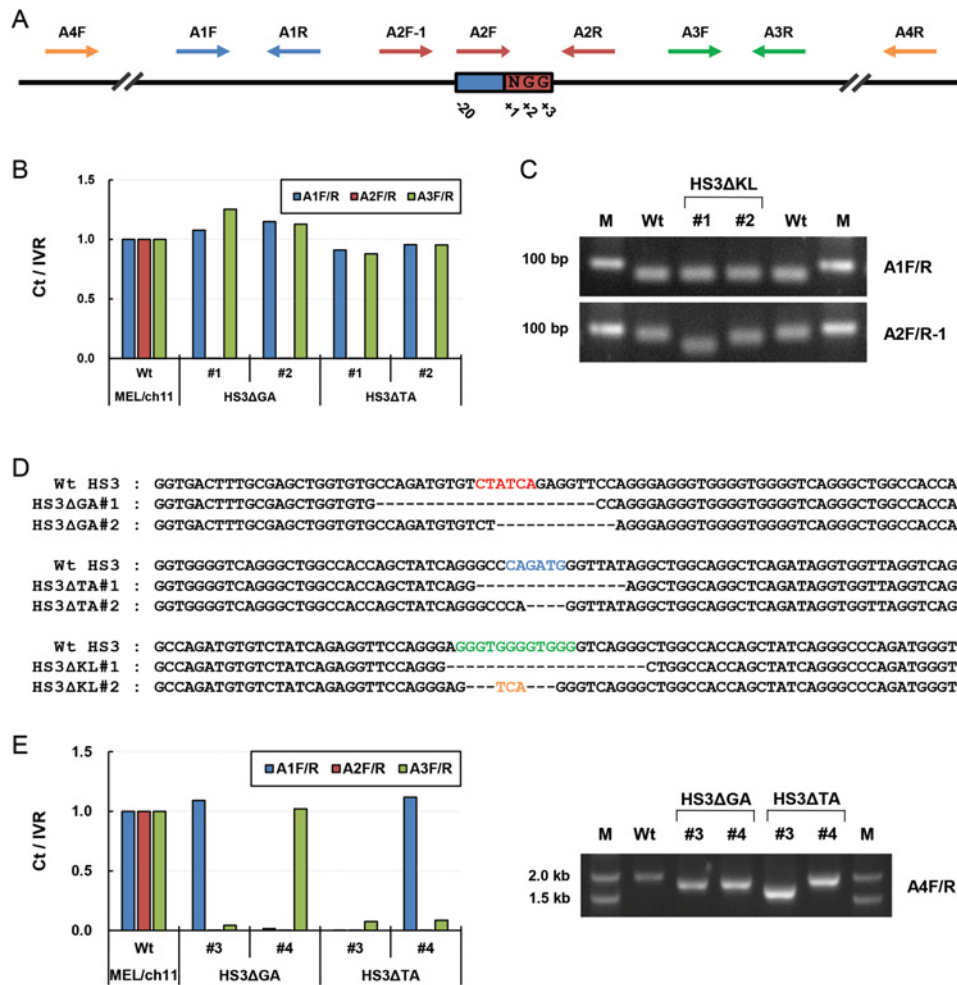


Figure 2. Screening of mutations generated by the CRISPR/spCas9 system

(A) The locations of primer sets for PCR are presented on DNA with respect to the target sequences of sgRNA. The target sequences are 20 nts in length, and 'NGG' nucleotides acting as PAMs are numbered as '+1', '+2', and '+3'. One primer in A2 amplicon (red arrows) locates on the target sequences, and A2R-1 primer is on next to the target sequences. A1 (blue arrows) and A3 primers (green arrows) locate in both flanking regions of the target sequences. A4F/R (yellow arrows) primers amplify large region including A1, A2, and A3 amplicons. (B) Genomic DNA from clones targeted by the CRISPR/spCas9 system, was amplified by qPCR using the three primer sets, A1, A2, and A3. PCR product amounts were compared using IVR amplicon for intervening region between the LCR and ϵ -globin gene as an internal control and then normalized compared with amounts in wild-type (Wt) MEL/ch11 cells. (C) In HS3 Δ KL clones, PCR products amplified by A1F/R and A2F/R-1 primers were visualized on an agarose gel, M is the DNA marker. (D) DNA sequences in the LCR HS3 are presented for Wt cells and clones with GATA-1, TAL1, or KLF1 motif deletions. The binding motifs of transcription factors are marked by colored bases as described in Figure 1. Black dashes represent deleted nucleotides and yellow bases are inserted nucleotides. (E) Genomic DNA of clones was amplified and quantitated as described above. PCR products amplified by A4F/R primers were visualized in an agarose gel.

Deletion of binding motifs for transcription factors inhibited their binding in a chromatin environment

To determine whether the deletion of binding motifs results in the loss of transcription factor binding, we carried out ChIP assays on three mutant clones, that are HS3 Δ GA#2, HS3 Δ TA#1, and HS3 Δ KL#2. The HS3 Δ GA#2 and HS3 Δ TA#1 clones showed decrease in approximately 50% in the occupancies of GATA-1 and TAL1 in the LCR HS3, respectively, as compared with control cells (Figure 3A,B). KLF1 occupancy was more strongly decreased in the HS3 Δ KL#2 clone (Figure 3C). The incomplete loss of occupancy for GATA-1 and TAL1 may have been due to other binding motifs near target motifs. Indeed, there are two or more GATA-1 and TAL1 motifs within 200 bp in the human β -globin LCR HS3, whereas only one cluster of KLF1 motifs in the LCR HS3, which was deleted in the

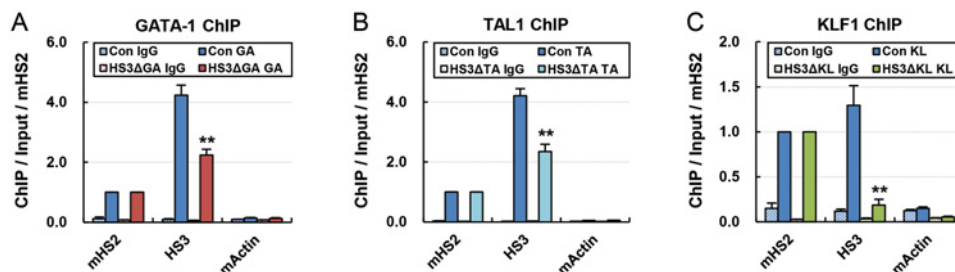


Figure 3. Occupancy of transcription factors in binding motif mutant clones

Control cells and clones with GATA-1 (A), TAL1 (B), or KLF1 (C) motif deletion were subjected to ChIP using antibodies specific for GATA-1, TAL1, or KLF1. Amounts of immunoprecipitated DNA were compared with input DNA and then normalized compared with amounts of immunoprecipitated DNA in the mouse β -globin LCR HS2. The mouse *Actin* gene served as an internal negative control and normal goat IgG (IgG) as an experimental negative control. Results are the means \pm S.E.M. of two independent experiments; ** $P < 0.01$.

HS3ΔKL#2 clone. These results indicate that binding motif deletion can inhibit the binding of transcription factor in a chromatin environment.

Double deletion of GATA-1 binding motifs in the human β -globin LCR HS3

To more strictly inhibit GATA-1 binding in the LCR HS3, we tried to mutate another GATA-1-binding motif located near the first GATA-1 motif. Guide sequences targetting the second GATA-1 binding motif (HS3_GA_R2) were cloned into pLH-spsgRNA2 vector, which has the hygromycin resistance gene as a selectable marker, and then transduced into the HS3ΔGA#2 clone using the procedure described in Figure 1C. After hygromycin selection, clones were analyzed by qPCR using the primer sets shown in Figure 4A. In clone #1 and #2 of HS3ΔΔGA, genomic DNA was not properly amplified by A2 primers (Figure 4B). DNA sequencing revealed that 16 and 37 bp including the second GATA-1 motif were deleted in HS3ΔΔGA #1 and #2, respectively, in addition to the deletion of 12 bp for the first GATA-1 motif (Figure 4C). Finally, we analyzed the occupancy of GATA-1 by ChIP. It was found GATA-1 occupancy was strongly decreased in the LCR HS3 (Figure 4D). These results show that two motifs for the same transcription factor can be sequentially deleted using CRISPR/Cas9 systems with different selection markers, and that this results in greater inhibition of factor binding.

Characterization of mutations generated by the CRISPR/spCas9 system

The CRISPR/Cas9 technique has evolved over several years and has been utilized to knockout gene expression in many studies [16-22]. To determine whether this genome editing technique using spCas9 is suitable for mutating binding motifs of transcription factors, we analyzed 94 clones that we obtained with six kinds of sgRNA. We could have missed out clones having mutations in a couple of base pairs because of the limited resolution of agarose gels in screening processes. Initially, clones were sorted by the type of mutation (Figure 5A). Sixty-eight clones were found to only exhibit deletion of nucleotides in and near target sequences. Insertion occurred in two clones without deletion and in eight clones with deletion. Sixteen clones had a deletion of a region exceeding 1 kb, as demonstrated by PCR using various pairs of primers. The number of inserted nucleotides was under 20 bp in six out of ten clones (Figure 5B). The lengths of deleted regions were under 20 bp in 48 of 76 clones (Figure 5C). These mutation patterns are similar to those reported previously, but the lengths of deleted regions was longer in our clones as compared with clones in that gene bodies were targetted [14,16]. This may have been due to the open-chromatin structure of the LCR HSs, which make them susceptible to attack by Cas9 endonuclease. In spite of high sensitivity in enhancer regions, our results show that the CRISPR/spCas9 system generates alterations mainly in shorter regions, that is, under 20 bp.

Next, we determined the locations of deleted nucleotides with respect to the guide sequences of sgRNA in 48 clones containing deletions of under 20 bp (Figure 5D). The fourth and fifth nucleotides upstream from PAMs were deleted in the clones over 90%, and several nucleotides near them were deleted in relatively high frequencies. These deletion patterns are consistent with those found in previous studies on plasmids or oligonucleotides, in which a putative cleavage site was identified between the third and fourth nucleotides upstream of PAMs [8,11]. Figure 5E shows the deletion pattern in 76 clones containing deletions of <1 kb. No directional preference was observed to the up or downstream regions from the highly deleted nucleotides between -1 and -6 from PAMs, although the downstream region was deleted further in a small number of clones. This deletion pattern provides guidelines regarding the choice of guide sequences for sgRNA.

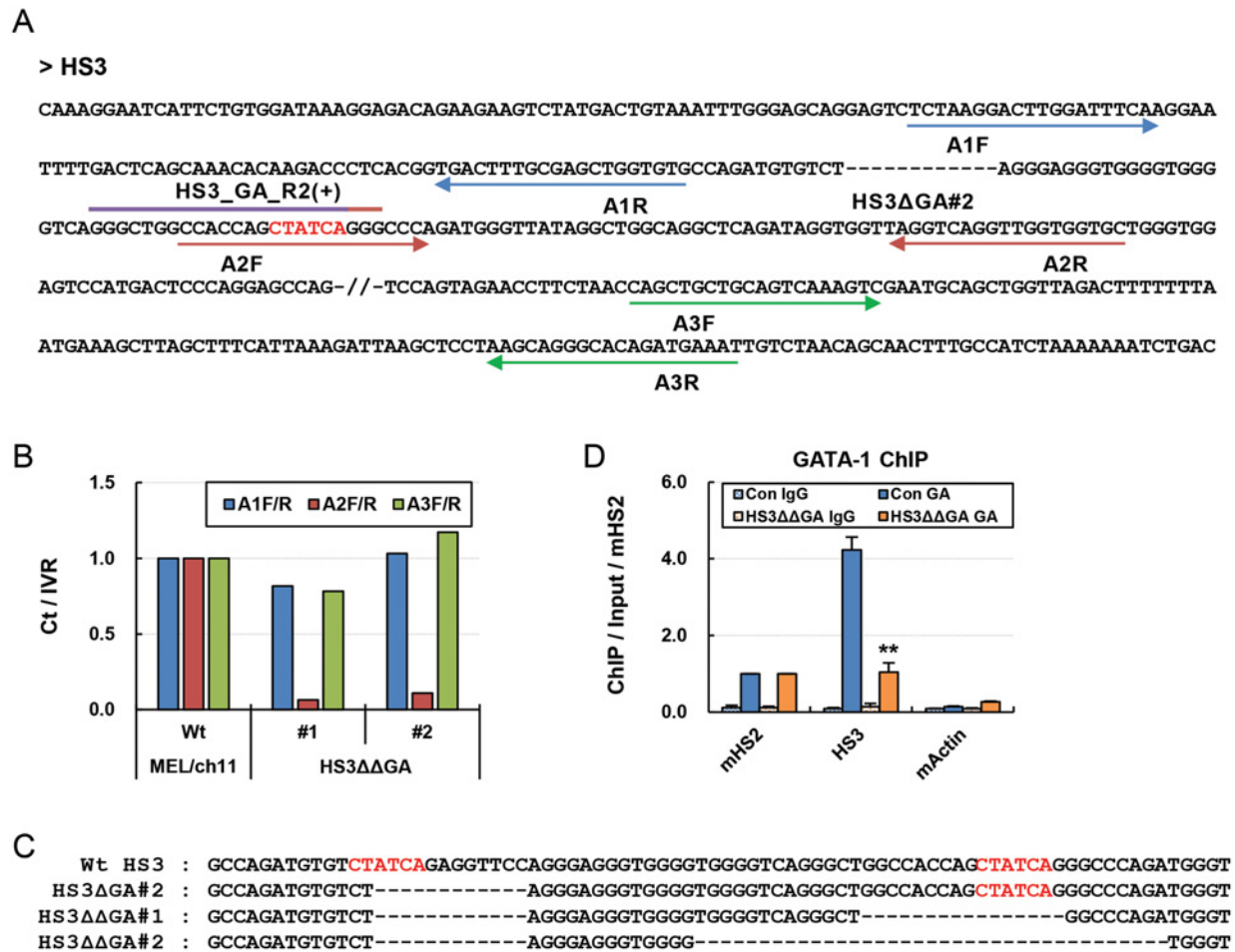


Figure 4. Double deletion of GATA-1-binding motifs in the human β -globin LCR HS3
 (A) DNA sequences of HS3 in the HS3 Δ GA#2 clone are presented with target sequences (purple line) for sgRNA. Another GATA-1 motif is indicated by six red bases. A2F primer (red arrows) is on the GATA-1 motif. (B) Genomic DNA was amplified by PCR in the HS3 Δ GA clones. Amounts of PCR products were measured as described in Figure 2. (C) DNA sequences in the LCR HS3 are presented for Wt, HS3 Δ GA#2 clone, and HS3 Δ GA#1 and #2 clones. (D) ChIP was performed with antibodies specific for GATA-1 in the HS3 Δ GA#1. Amounts of immunoprecipitated DNA were determined as described in Figure 3; ** $P < 0.01$.

In summary, we mutated the binding motifs of transcription factors using the CRISPR/spCas9 system in a genome context, and observed reductions in transcription factor occupancies in a chromatin environment. Our findings imply that the CRISPR/spCas9 system can be used to study the direct roles of transcription factors for target genes without disturbing the expressions of other transcription factors. Furthermore, DNA breaks caused by the CRISPR/spCas9 system resulted mainly in deletions or insertions of less than 20 nts, which indicates that this system is suitable for mutating short *cis*-elements, such as transcription factor binding motifs. Due to this restriction to short regions, other *cis*-elements near target sequences might be maintained during genome editing using the CRISPR/spCas9 system, which would allow study of the individual roles of *cis*-elements. In addition, the deletion of binding motifs for transcription factors using the CRISPR/Cas9 system would be useful for inhibiting the expressions of target genes without side effects, because the mutations of gene coding regions could result in the expression of abnormal polypeptides by nucleotide deletion or frameshift in translation. In recent studies, the *ALAS2* gene transcription was severely inhibited by a GATA-1 motif deletion generated using the CRISPR/Cas9 system [30,31]. Enhancer deletion of the *BCL11A* gene using the CRISPR/Cas9 system suggested the therapeutic possibility for the β -hemoglobinopathies by failing *BCL11A* expression and resulting in induction of the fetal γ -globin expression [32]. Overall, the CRISPR/Cas9 system is believed to effectively mutate the binding motifs of transcription factors in a genome context, and consequently the mutations can regulate the binding of transcription factors in chromatin and might regulate gene expression.

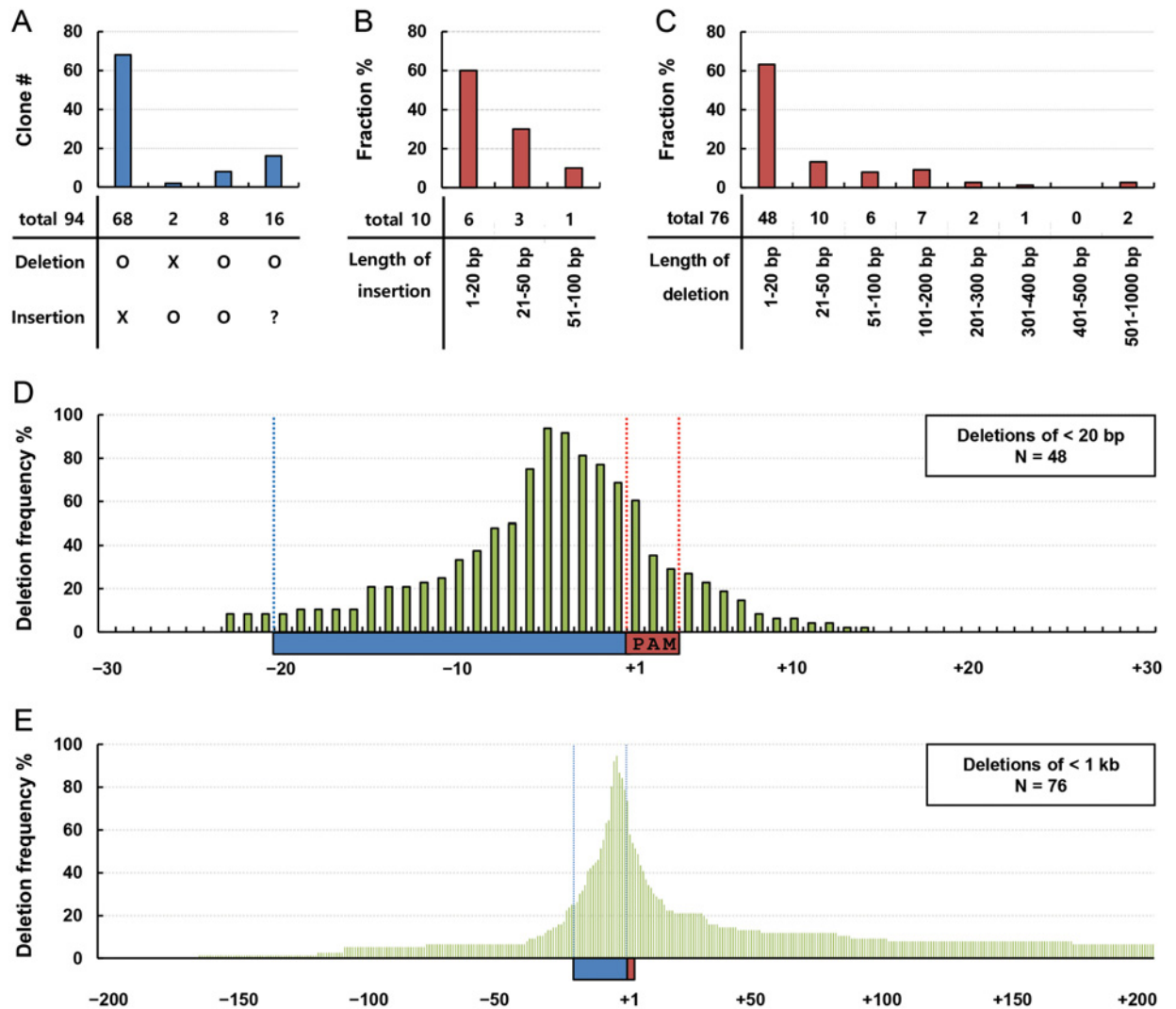


Figure 5. Characterization of mutations generated by the CRISPR/spCas9 system

(A) Mutations in 94 clones generated using six kinds of sgRNA were sorted by deletion and insertion. Mutations were analyzed by qPCR and/or DNA sequencing. (B) Ten clones with an insertion or an insertion and a deletion were sorted depending on the lengths of inserted nucleotides. (C) Seventy-six clones with a deletion of <1 kb were sorted depending on the lengths of deleted nucleotides. (D) Deletion frequencies were calculated at each nucleotide with respect to target sequences of sgRNA in 48 clones with deletions of <20 bp. The first nucleotide of PAM 'NGC' sequences in the red rectangle is indicated by '+1', and nucleotides upstream and downstream of PAMs are indicated by '-' and '+', respectively. (E) Using 76 clones with deletions of <1 kb, deletion frequencies were calculated at each nucleotide with respect to target sequences of sgRNA.

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Author contribution

Y.W.K. and A.R.K. designed the research. Y.W.K. performed the experiments and analysed the data. both the authors wrote the manuscript.

Funding

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning [grant number NRF-2014R1A2A1A11051702]; and the Ministry of Education [grant number NRF-2014R1A1A2009565].

Abbreviations

Cas9, CRISPR-associated protein 9; CRISPR, clustered, regularly interspaced, short palindromic repeat; crRNA, CRISPR RNA; DMEM, Dulbecco's modified eagle's medium; EGFP, Enhanced green fluorescent protein; GATA-1, GATA binding protein 1; HMBA, Hexamethylene bisacetamide; HS, hypersensitive site; IVR, Intervening region between the LCR and ϵ -globin gene; KLF1, Krueppel-like factor 1; LCR, locus control region; MEL/ch11, Murine erythroleukemia cell containing human chromosome 11; MNase, Micrococcal nuclease; PAM, Protospacer adjacent motif; qPCR, quantitative PCR; sgRNA, single guide RNA; TAL1, T-cell acute lymphoblastic leukemia 1; tracrRNA, trans-activating crRNA.

References

- 1 Kim, Y.W., Kim, S., Kim, C.G. and Kim, A. (2011) The distinctive roles of erythroid specific activator GATA-1 and NF-E2 in transcription of the human fetal γ -globin genes. *Nucleic Acids Res.* **39**, 6944–6955
- 2 Zhou, D., Liu, K., Sun, C.W., Pawlik, K.M. and Townes, T.M. (2010) KLF1 regulates BCL11A expression and γ - to β -globin gene switching. *Nat. Genet.* **42**, 742–744
- 3 Anderson, K.P., Crable, S.C. and Lingrel, J.B. (2000) The GATA-E box-GATA motif in the EKLf promoter is required for *in vivo* expression. *Blood* **95**, 1652–1655
- 4 Lee, W., Mitchell, P. and Tjian, R. (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**, 741–752
- 5 Maire, P., Wuarin, J. and Schibler, U. (1989) The role of *cis*-acting promoter elements in tissue-specific albumin gene expression. *Science* **244**, 343–346
- 6 Rudge, T.L. and Johnson, L.F. (2002) Synergistic activation of the TATA-less mouse thymidylate synthase promoter by the Ets transcription factor GABP and Sp1. *Exp. Cell Res.* **274**, 45–55
- 7 Li, D., Lin, Y., Liu, Z., Zhang, Y., Rong, Z. and Liu, X. (2012) Transcriptional regulation of human novel gene SPATA12 promoter by AP-1 and HSF. *Gene* **511**, 18–25
- 8 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821
- 9 Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S. et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712
- 10 Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A. et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607
- 11 Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586
- 12 Anders, C., Niewoehner, O., Duerst, A. and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573
- 13 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823
- 14 Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E. et al. (2013) RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826
- 15 Cho, S.W., Kim, S., Kim, J.M. and Kim, J.S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232
- 16 Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84
- 17 Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S. et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87
- 18 Sanjana, N.E., Shalem, O. and Zhang, F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784
- 19 Shen, B., Zhang, J., Wu, H., Wang, J., Ma, K., Li, Z. et al. (2013) Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res.* **23**, 720–723
- 20 Hai, T., Teng, F., Guo, R., Li, W. and Zhou, Q. (2014) One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell Res.* **24**, 372–375
- 21 Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F. et al. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918
- 22 Yang, D., Xu, J., Zhu, T., Fan, J., Lai, L., Zhang, J. et al. (2014) Effective gene targeting in rabbits using RNA-guided Cas9 nucleases. *J. Mol. Cell Biol.* **6**, 97–99
- 23 Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D. et al. (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202
- 24 Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T. et al. (2014) Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324
- 25 Stewart, A.J., Hannehalli, S. and Plotkin, J.B. (2012) Why transcription factor binding sites are ten nucleotides long. *Genetics* **192**, 973–985
- 26 Liang, S., Moghimi, B., Yang, T.P., Strouboulis, J. and Bungert, J. (2008) Locus control region mediated regulation of adult β -globin gene expression. *J. Cell Biochem.* **105**, 9–16
- 27 Kim, S., Kim, Y.W., Shim, S.H., Kim, C.G. and Kim, A. (2012) Chromatin structure of the LCR in the human β -globin locus transcribing the adult δ - and β -globin genes. *Int. J. Biochem. Cell Biol.* **44**, 505–513

- 28 Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S. and Pederson, T. (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3002–3007
- 29 Zischewski, J., Fischer, R. and Bortesi, L. (2017) Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol. Adv.* **35**, 95–104
- 30 Tanimura, N., Miller, E., Igarashi, K., Yang, D., Burstyn, J.N., Dewey, C.N. et al. (2016) Mechanism governing heme synthesis reveals a GATA factor/heme circuit that controls differentiation. *EMBO Rep.* **17**, 249–265
- 31 Wakabayashi, A., Ulirsch, J.C., Ludwig, L.S., Fiorini, C., Yasuda, M., Choudhuri, A. et al. (2016) Insight into GATA1 transcriptional activity through interrogation of *cis*-elements disrupted in human erythroid disorders. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4434–4439
- 32 Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O. et al. (2015) BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature* **527**, 192–197