Augmenting and directing long-range CRISPR-mediated activation in human cells

Y. Esther Tak1,2,3,4, Joy E. Horng1,2,3,7, Nicholas T. Perry1,2,3,7, Hayley T. Schultz1,2,3, Sowmya Iyer1, Qiuming Yao1,2,4,5, Luli S. Zou5,6, Martin J. Aryee1,2,4,5,6, Luca Pinello1,2,4,5 and J. Keith Joung1,2,3,4

Epigenetic editing is an emerging technology that uses artificial transcription factors (aTFs) to regulate expression of a target gene. Although human genes can be robustly upregulated by targeting aTFs to promoters, the activation induced by directing aTFs to distal transcriptional enhancers is substantially less robust and consistent. Here we show that long-range activation using CRISPR-based aTFs in human cells can be made more efficient and reliable by concurrently targeting an aTF to the target gene promoter. We used this strategy to direct target gene choice for enhancers capable of regulating more than one promoter and to achieve allele-selective activation of human genes by targeting aTFs to single-nucleotide polymorphisms embedded in distally located sequences. Our results broaden the potential applications of the epigenetic editing toolbox for research and therapeutics.

Epigenetic editing using aTFs with programmable DNA-binding domains enables tunable regulation of target gene expression and has a broad range of potential applications in basic research, synthetic biology and human therapeutics. To date, robust transcriptional activation using aTFs has been primarily accomplished by targeting these factors to promoter sequences (typically less than ±500 bp relative to the transcription start site (TSS)). However, more distally located regulatory sequences such as enhancers, which are enriched for disease-associated single-nucleotide polymorphisms (SNPs)1–4, are attractive targets for achieving more complex outcomes such as allele-specific gene activation. Although aTFs have been previously reported to induce activation from enhancers in heterotopic cell settings and other distal regulatory sequences, these efforts have not consistently resulted in efficient target gene activation, with fold-activation levels often much lower than what has been achieved by targeting aTFs to promoters5–14.

Here we show that long-range activation using CRISPR-based aTFs in human cells can be made more consistent and robust by concurrently directing an aTF to the target gene promoter of interest. Importantly, we illustrate how aTF-mediated activation can be used to influence target gene choice for an enhancer sequence known to regulate multiple promoters and provide a first proof of concept for allele-selective activation of human genes by targeting aTFs to SNPs embedded in distally located sequences. Our results improve the ability to effectively employ aTFs for directing long-range gene activation and thereby broaden the potential applications of the epigenetic editing toolbox.

Results

Artificial transcription factors do not consistently activate from enhancer sequences. Initially, we identified examples where recruitment of an aTF to a distal sequence does not yield robust activation of the expected target gene, despite these sequences acting as enhancers in other cell types (Fig. 1a). Previous studies have attempted to achieve such heterotopic cell-type activation of enhancers or other distal sequences but did not consistently yield target gene expression of fivefold or more and/or required the use of multiple guide RNAs (gRNAs) to recruit catalytically inactive or ‘dead’ Streptococcus pyogenes Cas9 (dCas9)-based aTFs5–14 (Supplementary Note). We targeted three endogenous genes (IL2RA, CD69 and MYOD1) that are not expressed at detectable levels as measured by RNA sequencing (RNA-seq; fragments per kilobase of transcript per million mapped reads (FPKM) values <2; Methods) in four human cell lines: U2OS, HEK293, HepG2 and K562 (with the exception of CD69, which is moderately expressed in K562 cells, and therefore not tested; Supplementary Table 1). We used a bipartite, small-molecule-inducible, dCas9-based aTF consisting of two components: (1) dCas9 fused to four DmrA domains (DmrA(×4)) and (2) DmrC fused to a nuclear factor-κB (NF-κB) activation domain (hereafter the bipartite p65 aTF)15. DmrA and DmrC domains are fragments of the FK506-binding protein (FKBP) and FKBP-rapamycin-binding protein, respectively, and interact only in the presence of a rapamycin analog known as the A/C heterodimerizer (Fig. 1b). The bipartite p65 aTF provides a robust activator that can recruit multiple copies of an activation domain using a single gRNA. For IL2RA, we designed gRNAs to direct this aTF to two sequences known to be functional enhancers in T cells14 that are located ~5kb upstream or ~10kb downstream of the TSS (Fig. 1c). These targeted sequences are present in inactive, closed chromatin in HEK293 and K562 cells and open in chromatin with histone 3 lysine 27 acetylation (H3K27Ac) marks in U2OS and HepG2 cells (Extended Data Fig. 1a). Testing of individual gRNAs targeted to each of these two regions (Fig. 1c) did not yield a significant increase in IL2RA transcription in any of the four cell lines (Fig. 1c and Methods). Similarly, we did not observe activation of CD69 in U2OS, HEK293 and HepG2 cells when we used the bipartite p65 aTF with single gRNAs targeting an upstream conserved noncoding sequence 2 (CNS2) known to be a stimulus-responsive enhancer in...
We speculated that the inability to con-
unlocks enhancer activity.

**Fig. 1** | Heterotopic activation of enhancer sequences by dCas9-based aTFs in multiple human cell lines. **a**, Schematics illustrating: (i) an enhancer X that activates promoter Y in cell type A, (ii) lack of enhancer X activity on promoter Y in a different cell type B, (iii) lack of enhancer X activity on promoter Y in cell type B when an aTF is recruited only to enhancer X, and (iv) robust enhancer X activity on promoter Y in cell type B when aTFs are recruited to both enhancer X and promoter Y. **b**, Architectures of dCas9-based bipartite and direct fusion aTFs used in this study. c–e, mRNA expression levels of the endogenous IL2RA, CD69 and MYOD1 genes in human cell lines in the presence of the bipartite p65 aTF and one or more gRNAs targeting enhancer (E1, E2, E3 or E4) or promoter (P) sequences. Relative expression of each gene was measured by quantitative PCR with reverse transcription (RT–qPCR), normalized to HPRT1 levels and calculated relative to that of a control sample (none) expressing a non-targeting gRNA. Open circles indicate biological replicates (n = 3), bars represent the mean of replicates, and error bars show the s.e.m. An asterisk indicates that the value is significantly different from the sample targeting only the promoter, *P < 0.05 (two-tailed Student’s t-test, assuming equal variance). The exact P values are available as source data.

T cells\(^h\) and present in closed chromatin in these three cell lines (Fig. 1d and Extended Data Fig. 1b). Additionally, we tested the bipartite p65 aTF with four single gRNAs targeted to a core enhancer (CE) located ~20 kb upstream of the MYOD1 TSS (Fig. 1e), previously shown to be active in myoblasts\(^i\), but that resides in inactive, closed chromatin in human HEK293, U2OS, HepG2 and K562 cell lines (Extended Data Fig. 1c). These experiments revealed only modest activation of MYOD1 (sixfold) with just one of the four gRNAs (E4) in HEK293 and U2OS cells and no significant activation with any of the four gRNAs in HepG2 and K562 cells (Fig. 1e).

**Concurrent artificial transcription factor promoter targeting unlocks enhancer activity.** We speculated that the inability to consistently and efficiently induce gene activation from distal enhancer elements with an aTF might be due to the inactive, closed state of the target gene promoter in these heterotopic cell settings (Fig. 1a) and therefore further envisioned that concurrent targeting of an aTF to both the distal element and the target promoter might yield more reliable and robust activation (Fig. 1a). Consistent with this idea, we were able to modestly activate MYOD1 from the CE enhancer sequence in U2OS and HEK293 cells (Fig. 1e), in which the promoter exhibited an open architecture and weak H3K27Ac marks (Extended Data Fig. 1c); by contrast, we could not activate MYOD1 with an aTF targeted to the CE enhancer in HepG2 and K562 cells (Fig. 1e), in which the promoter was in closed chromatin (Extended Data Fig. 1c), perhaps rendering it inert to any activating effects. To test this hypothesis, we coexpressed each of the enhancer-targeted gRNAs used in our experiments with IL2RA, CD69 and MYOD1 described above together with a promoter-targeted gRNA (Fig. 1c–e), thereby recruiting the bipartite p65 aTF to both sequences concurrently (Fig. 1a). In control experiments, we found that these promoter-targeted gRNAs each activated transcription of its target gene (ranging from 3- to 62-fold, 1- to 44-fold and 2- to 52-fold for IL2RA, CD69 and MYOD1, respectively) across the various cell lines.
tested (Fig. 1c–e). However, coexpression of enhancer-targeted and promoter-targeted gRNAs with the bipartite p65 aTF led to synergistically higher levels of target gene transcription (that is, greater levels of expression than the product of activation with each gRNA individually) for nearly all combinations of gRNAs (ranging from 5- to 224-fold, 6- to 160-fold and 14- to 496-fold for IL2RA, CD69 and MYOD1, respectively; Fig. 1c–e). This represents as much as an additional 10-, 8- and 32-fold upregulation in the expression of IL2RA, CD69 and MYOD1, respectively (Fig. 1c–e), that can be attributed to aTF binding to the distal enhancer sequence. Levels of activation with concurrent enhancer–promoter targeting were generally somewhat lower than the synergistic effect observed with two aTFs targeted to the promoter (Supplementary Note and Supplementary Fig. 1). In addition, RNA-seq experiments revealed that the transcriptome-wide specificity of activation with concurrent enhancer–promoter aTF targeting was dependent on the design of the gRNAs and the functional effects of the target genes themselves (Supplementary Note and Supplementary Fig. 2).

We additionally assessed whether bipartite aTFs harboring synthetic VPR or VP64 domains and direct fusions of dCas9 to p65, VPR, VP64 or p300 domains (Fig. 1b) could mediate activation from distal sequences with concurrent promoter targeting. For these experiments, we used the same pairs of enhancer–promoter gRNAs that we tested with the bipartite p65 aTF on IL2RA, CD69 and MYOD1 (Fig. 2a–c). In general, we found that bipartite p65 aTFs and direct fusion VPR aTFs functioned robustly at all three genes across all cell types tested (Fig. 2a–c). Bipartite VPR, bipartite VP64, direct fusion VP64 and direct fusion p300 aTFs could also each activate all three gene targets but did so with less consistency across cell lines (although each aTF activated all three target genes in at least one cell type; Fig. 2a–c). Bipartite VPR aTFs showed substantial toxicity in U2OS cells and therefore could not be reliably assessed for gene activation in that context (Fig. 2a–c). We speculate that toxicity observed with bipartite VPR aTFs may result from high-level expression of the relatively small-sized DmrC–VPR fusion and its oligomerization on dCas9–DmrA(×4), both of which
could contribute to potential transcriptional squelching in U2OS cells. Direct fusion p65 aTFs did not robustly activate any of the three genes in all four cell lines (Fig. 2a–c). In general, concurrent targeting worked in nearly every setting in which the aTF bound to the promoter alone stimulated gene expression, suggesting that efficient long-distance aTF activity is strongly dependent on active transcription from the target gene promoter.

**Directing promoter choice of multigene enhancers using aTFs.** We wondered whether our strategy might be used to direct the activity of an enhancer that is known to regulate multiple target genes in a different cell type. For example, the locus control region (LCR) enhancer sequentially and preferentially activates transcription in erythroid cells from the HBE1, HBG1/2 and HBB promoters during the embryonic, fetal and postnatal stages of human development, respectively (Supplementary Note and Supplementary Fig. 3). Finally, (Fig. 3c–e). In addition, the direct fusion VP64 aTF did not activate any genes in HEK293 cells (Fig. 3e) and (3) no notable toxicity in U2OS cells (Fig. 3d), (2) the direct fusion VP64 and direct fusion p300 aTFs also worked to direct LCR enhancer activity to target promoters with only a small number of exceptions: (1) the bipartite VPR aTF was not completely specific in its differential activation of HBG in HepG2 cells and again showed notable toxicity in U2OS cells (Fig. 3d), (2) the direct fusion VP64 aTF did not activate any genes in HEK293 cells (Fig. 3e) and (3) no activation of HBE1 was observed with some aTFs in certain cell lines (Fig. 3c–e). In addition, the direct fusion VP64 aTF did not activate even when paired with aTFs harboring heterologous activation domains (Supplementary Note and Supplementary Fig. 3). Finally, direct fusion VPR aTFs targeted to the LCR failed to activate when only dCas9 or dCas12a proteins (lacking activation domains) were targeted to the promoter, demonstrating the requirement for activation domains at both the enhancer and promoter (Supplementary Note and Supplementary Fig. 4).

To test whether a TF targeting could guide promoter choice of a different multigene enhancer, we used the human APO gene cluster, which includes an enhancer that regulates the expression of both APOA4 and APOC3 in hepatic cells. We designed an SpCas9 gRNA (named E0) that targets a site in the enhancer and gRNAs that target sites in the APOA4 or APOC3 promoter (named P_A4 and P_C3, respectively; Fig. 4a). Coexpression of the bipartite p65 aTF with only the E0 gRNA failed to activate either APOA4 or APOC3 but addition of either the P_A4 or P_C3 promoter-targeted gRNA led to dramatic and specific upregulation of each cognate gene that was substantially higher than that observed with only the P_A4 or P_C3 gRNA (Fig. 4b,c and Extended Data Fig. 3a,b).

**Allele-selective activation using SNPs in distal sequences.** Although native transcription factors have been shown to exert allele-selective gene activation in human cells24–26, no study has, to our knowledge, shown that aTFs can do so using SNPs in distal regulatory sequences (although a recent study showed allele-selective binding of an aTF to a 12-bp-inserted allele present in the TALI super-enhancer in Jurkat cells27). To perform a proof-of-principle experiment, we used the APOA4 and APOC3 genes in HEK293 cells, which we found were heterozygous for alleles (hereafter allele 1 and allele 2) distinguishable by SNPs within the coding sequences of each gene (in exon 2 of APOA4 and exon 3 of APOC3; Methods, Fig. 4a and Extended Data Fig. 4a–c). We also identified a distinct sequence (Fig. 4a and Extended Data Fig. 4a) that we hypothesized might function as a potential enhancer of both APOA4 and APOC3 based on previously defined H3K27Ac and open chromatin marks at this site in HepG2 cells (Extended Data Fig. 4a). Within this potential enhancer sequence in HEK293 cells, we identified target sites for six SpCas9 gRNAs (named E1–E6), each of which are heterozygous for a SNP that alters one of the two conserved guanines in the protospacer adjacent motif (PAM) sequence (Fig. 4a and Extended Data Fig. 4b,c). Using amplification sequencing of chromatin immunoprecipitation (ChIP) products, we confirmed that these gRNAs can each preferentially direct binding of the bipartite p65 aTF to the allele that bears an intact PAM relative to the other allele that has a disrupted PAM (that is, the E1, E2 and E4 gRNAs bind preferentially to their target sites on allele 1 over allele 2 and vice versa for the E3, E5 and E6 gRNAs; Extended Data Fig. 5). When tested with the bipartite p65 aTF, each of the six gRNAs (E1–E6) only activated APOA4 when the P_A4 promoter-targeted gRNA was also coexpressed (Fig. 4b and Extended Data Fig. 3a), verifying that binding to the potential enhancer can lead to long-range activation (Fig. 4b). cDNA sequencing of these activated APOA4 mRNA transcripts revealed unbalanced expression of the two APOA4 alleles with each of the E1–E6 gRNAs, in contrast to more equally balanced expression with the E0 gRNA (Fig. 4d). Combined expression of enhancer gRNAs targeted to the same allele (that is, E1 + E2 + E4 or E3 + E5 + E6) together with the P_A4 gRNA resulted in even greater increases in APOA4 expression (Fig. 4b) and further imbalances in relative expression of the two alleles (Fig. 4d). We were able to induce similar allele-selective expression of APOC3 with the enhancer gRNAs (E0, E1–E6), the promoter-targeted P_C3 gRNA and the bipartite p65 aTF in HEK293 cells (Fig. 4c and Extended Data Fig. 3b–d). We used the term ‘allele selective’ rather than ‘allele specific’ to describe the differential gene activation effects on different alleles, which are preferential but not absolute. Potential reasons for differences in the magnitude of imbalance observed for a TF binding versus target gene activation (Fig. 4d,c and Extended Data Fig. 5c) include the possibility that not all binding events of an aTF molecule might lead to activation and that the two methods used to measure these parameters have different sensitivities.

To further test the generalizability of our approach for allele-selective gene activation, we tested two additional genes in two other cell lines. In one case, we assessed HBB expression in U2OS cells using four gRNAs that target sites in the H54 LCR enhancer that are each heterozygous for a PAM-disruptive SNP (Extended Data Fig. 6a). In the second case, we examined MYOD1 expression in K562 cells and used allele-selective gRNAs targeting the distal regulatory region enhancer27 (Extended Data Fig. 7a). For both experiments, we tested these enhancer-targeted gRNAs with a gRNA targeting the target gene promoter and the bipartite p65 aTF. At both genes, we were able to leverage targeting of SNPs present in enhancer sequences to achieve robust, allele-selective gene activation (Extended Data Figs. 6b–c and 7b–c).

**Discussion**

The work described here defines a general strategy to more robustly and consistently access the gene activation capabilities of enhancers...
in heterotopic settings or other distal sequences by directing an aTF not only to these sequences (as done previously[8–12]) but also concurrently to the target gene promoter. The extent of gene activation we observed depended on cell type, perhaps due to differing expression of cofactors of the activation domains in the aTFs we used.

in another cell type. While all the distal and promoter sequences that we used lie within a single topologically associated domain conserved across multiple cell types (Supplementary Note and Supplementary Fig. 5), we found in preliminary studies that simultaneous targeting of aTFs to sequences outside the topologically associated domain in which the target gene lies can in some cases also lead to activation (Supplementary Note and Supplementary Fig. 6).

Our studies have potential implications for understanding normal enhancer function. The finding that enhancer activity is influenced by promoter status may impact how such sequences are identified using CRISPR activation screens. For example,
in heterotopic cell settings, an associated enhancer for an inactive target promoter might be missed without also activating that promoter. In addition, our studies can improve our understanding of how a single enhancer differentially regulates multiple promoters within a cluster. Our findings with the β-globin gene cluster suggest that enhancers might be redirected simply by upregulating or
downregulating different promoters. Consistent with this, hemoglobin gene switching studies have shown both an increase in the KLF1 activator at the HBB promoter and evicion of the NF-Y activator by the BCL11A repressor on the HBG promoter when LCR activity is redirected from HBG to HBB^13–15.

Finally, our method for robust heterotic activation of enhancers expands the utility and precision of CRISPR-based aTFs. Concurrent aTF targeting could enable differential increases in target gene expression when more than one promoter can potentially be upregulated, enabling the generation of more complex spatiotemporal gene expression patterns. This approach provides a more parsimonious solution to the challenge of robustly regulating different target genes in the same cluster because a single enhancer-targeted gRNA can be used with each promoter-targeted gRNA to activate individual genes instead of using multiple promoter-targeted gRNAs for each of those genes. In addition, our aTF strategy enables allele-selective gene expression by differentially targeting SNPs embedded in enhancer or other distal sequences. Our analysis using data from the 1000 Genomes Project and chromatin accessibility data from multiple cell lines found that SNPs that disrupt or create NGG PAM sequences for SpCas9 are greatly enriched genome wide in putative enhancers compared with promoters: ~2-fold and ~12-fold higher for SNP density and for total number of SNPs, respectively (Methods, Extended Data Fig. 8 and Supplementary Table 2). Allele-selective gene activation might provide a general therapeutic strategy for haplo-insufficient or dominant-negative diseases, enabling preferential upregulated expression of a wild-type allele over a mutant allele for therapeutic benefit^16–18. In sum, our robust strategy for enabling long-range activation should broaden the scope and range of research, synthetic biology and therapeutic applications of CRISPR-based aTFs.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of data and code availability are available at https://doi.org/10.1038/s41592-021-01224-1.

Received: 20 October 2020; Accepted: 28 June 2021; Published online: 5 August 2021

References
1. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. Nat. Rev. Mol. Cell Biol. 20, 490–507 (2019).
2. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods 13, 127–137 (2016).
3. Wang, H., La Russa, M. & Qi, L. S. CRISPR-Cas9 in genome editing and beyond. Annu. Rev. Biochem. 85, 227–264 (2016).
4. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).
5. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43–49 (2011).
6. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. Nature 593, 238–243 (2021).
7. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661 (2014).
8. Gao, X. et al. Comparison of TALE designer transcription factors and the CRISPR-Cas9 system for regulation of gene expression by targeting enhancers. Nucleic Acids Res. 42, e155 (2014).
9. Hilton, I. B. et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. 33, 510–517 (2015).
10. Koscu, C. et al. Temporal and spatial epigenome editing allows precise gene regulation in mammalian cells. J. Mol. Biol. 431, 111–121 (2019).
11. Li, K. et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. Nat. Commun. 11, 485 (2020).
12. Klann, T. S. et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. 35, 561–568 (2017).
13. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. 49, 1602–1612 (2017).
14. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature 549, 111–115 (2017).
15. Tak, Y. E. et al. Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods 14, 1163–1166 (2017).
16. Laguna, T. et al. New insights on the transcriptional regulation of CD69 gene through a potent enhancer located in the conserved noncoding sequence 2. Mol. Immunol. 66, 171–179 (2015).
17. Chen, J. C. J. & Goldhamer, D. J. The core enhancer is essential for proper timing of MyoD activation in limb buds and branchial arches. Dev. Biol. 265, 502–512 (2004).
18. Wienert, B., Martyn, G. E., Funnell, A. P. W., Quinlan, K. G. R. & Crossley, M. Wake-up sleepy gene: reactivating fetal globin for β-hemoglobinopathies. Trends Genet. 34, 927–940 (2018).
19. Muegstraten, S. T. & Hart, A. H. Modelling human haemoglobin switching. Blood Rev. 33, 11–23 (2019).
20. Sankaran, V. G. & Orkin, S. H. The switch from fetal to adult hemoglobin. Cold Spring Harb. Perspect. Med. 3, a011643 (2013).
21. Sankaran, V. G. et al. Human fetal hemoglobin expression is regulated by the myoglobin-repressing enhancer and the BCL11A locus. Science 322, 1839–1842 (2008).
22. Li, Q., Harji, S. & Petersen, K. R. Locus control regions: coming of age at a decade plus. Trends Genet. 15, 403–409 (1999).
23. Zannis, V. I., Kan, H. Y., Kritis, A., Zanni, E. E. & Kardassis, D. Transcriptional regulatory mechanisms of the human apolipoprotein genes in vitro and in vivo. Curr. Opin. Lipidol. 12, 181–207 (2001).
24. Cavalli, M. et al. Allele-specific transcription factor binding to common and rare variants associated with disease and gene expression. Hum. Genet. 135, 485–497 (2016).
25. Spisák, S. et al. CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. Nat. Med. 21, 1357–1363 (2015).
26. Bailey, S. D. et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat. Commun. 2, 6186 (2015).
27. Tapscott, S. J., Lassar, A. B. & Weintraub, H. A novel myoblast enhancer element mediates Myod transcription. Mol. Cell. Biol. 12, 4994–5003 (1992).
28. Jost, M. et al. Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. Nat. Biotechnol. 38, 355–364 (2020).
29. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461 (2014).
30. ENCODE Project Consortium. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710 (2020).
31. Zhou, D., Pawlik, K. M., Ren, J., Sun, C.-W. & Townes, T. M. Differential binding of erythroid Krupple-like factor to embryonic/fetal globin gene promoters during development. J. Biol. Chem. 281, 16052–16057 (2006).
32. Liu, N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. Cell 173, 430–442 (2018).
33. Liu, N. et al. Transcription factor competition at the γ-globin promoters controls hemoglobin switching. Nat. Genet. 53, 511–520 (2021).
34. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).
35. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum. Genet. 132, 1077–1130 (2013).
36. Veitia, R. A., Caburet, S. & Birchler, J. A. Mechanisms of Mendelian dominance. Clin. Genet. 93, 419–428 (2013).
37. Matharu, N. et al. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. Science 363, eaau0629 (2019).
38. Dang, V. T., Kassahn, K. S., Marcos, A. E. & Rangan, M. A. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. Eur. J. Hum. Genet. 16, 1350–1357 (2008).
39. Inoue, K. & Fry, E. A. Haploinsufficient tumor suppressor genes. Adv. Med. Biol. 118, 83–122 (2017).
40. Procampsaly, R. R., Horwitz, E. M., DiGirolamo, C. M., Stokes, D. S. & Prokopck, D. J. Correction of a mineralization defect by overexpression of a wild-type cDNA for COL1A1 in narrow stromal cells (MSCs) from a patient with osteogenesis imperfecta: a strategy for rescuing mutations that produce dominant-negative protein defects. Gene Ther. 12, 1119–1125 (2005).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021, corrected publication 2022
Methods

Plasmids and oligonucleotides. The list of plasmids and related sequences used in this study can be found in the Supplementary Note. SpCas9 gRNA and LbCas12a CRISPR RNA (crRNA) oligonucleotide sequences are available in Supplementary Table 3.

Human cell culture conditions. The American Type Culture Collection (ATCC) cell lines, authenticated by short tandem repeat analysis, were HEK293 (Invitrogen, similar to ATCC CRL-1573; a loss of two alleles, no. 9 in the DS8818 locus and the no. 11 at the CSF1PO locus), U2OS (gift of T. Cathomen, similar match to ATCC HTB-99 from no. 1 of the DS8818 locus), HepG2 (ATCC HB-8065) and K562 (ATCC CCL-243). HEK293, U2OS and K562 cells were authenticated on 8 November 2019, and HepG2 cells were authenticated on 14 December 2018. All cell culture reagents were obtained from Thermo Fisher unless otherwise specified. HEK293 and U2OS cells were grown in DMEM (11995073), HepG2 cells in EMEM (Welshers, 30-2033), 37 °C, 5% CO2 and K562 cells in RPMI 1640 medium (62870-127) with additional 2 mM GlutaMAX (35050061), supplemented with 10% heat-inactivated fetal bovine serum (16140-089) and 1% penicillin–streptomycin (15070-06).

Gene activation experiments. For direct fusion aTF experiments, HEK293, VPR plasmids (200 ng), and Cas9 gRNA/Cas12a crRNA plasmids (400 ng). dCas12a–DmrA(×) for heterotopic activation of enhancer sequences of cells and 37 °C, 5% CO2. Media supernatant was analyzed biweekly for any contamination at 95 °C for 20 s followed by 45 cycles of 95 °C for 3 s and 60 °C for 30 s for 30 cycles. Relative efficiency for each target was calculated by normalization to input control.

RNA sequencing. RNA libraries were prepared from 500 ng of total RNA treated with Ribogold zero to remove ribosomal RNA, using TruSeq Stranded Total RNA Library Prep Gold kit (Illumina, 20020399) and TruSeq RNA Single Indexes. The RNA libraries were sequenced with SE 75-cycle reads on an Illumina NextSeq 500 at the Broad Institute of Harvard and MIT. Reads were aligned to human reference genome hg19 using STAR (https://star.readthedocs.io/en/stable) and PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/). Reads aligning to ribosomal RNA were then filtered out of the alignment. Genomic coverage from filtered alignments was calculated by normalizing to sequencing depth using BEDTools. FPKM values were calculated using Cufflinks. Differential gene expression was performed using DESeq2 (v1.20.0).

Gene Ontology enrichment analysis. GO analysis was performed using the PANTHER website (http://pantherdb.org)53. A list of genes that showed differential expression after activation of MYOD1 identified by RNA-seq (Supplementary Fig. 2c) was used as the input for the analysis. The PANTHER Overrepresentation Test (released on 7 April 2020) was performed with the Gene Ontology (GO) database (released on 21 February 2020). Fisher’s exact test with false discovery rate correction was used, and GO Biological Process Complete was used as an annotation dataset.

ATAC sequencing. Open or closed status of the chromatin was determined using ATAC-seq. The ATAC–seq libraries were constructed following the protocol of Corces et al.54 and using Nextera DNA Flex Library Prep Kit (Illumina, FC-121-1030). The libraries were sequenced with paired-end (PE) 150-cycle reads on an Illumina NextSeq 500 system at the Broad Institute of Harvard and MIT. Reads were aligned to human reference genome hg19 using Burrows–Wheeler alignment and filtered to exclude PCR duplicates and processed as previously described55. Read start positions were shifted toward the 3′ end by 4 bp for reads aligning to the plus strand and toward the 5′ end by 5bp for reads aligning to the minus strand. Genomic coverage was calculated by counting reads in 150-bp sliding windows at 20-bp steps across the genome and then normalized to 10 million reads in each experimental using BEDTools.

Defining APOC3 enhancer sequences for SNP analysis. Known APOC3 enhancer sequences are located 500 to 890 bp upstream of the TSS56 and show open chromatin and H3K27Ac enrichment features in HepG2 cells in which APOC3 is highly expressed (UCSC Genome Browser (hg19); Supplementary Table 1). We identified potential enhancer sequences in the region encompassing −4.4 Kb to 2 Kb upstream of the TSS based on similar open chromatin and H3K27Ac enrichment features (Extended Data Fig. 4a).

Chromatin immunoprecipitation. Twenty-four hours before transfections, HEK293 cells seeded in 10-cm dishes and then transfected with 15 µg of plasmids (6 µg of dCas9–DmrA(×), 3 µg of DmrC–p65 and 6 µg of Cas9 gRNA) using 45 µl of TransIT-T293. Cells were trypsinized 72 h after transfection and ChIP assays were carried out as previously described with some modifications, using specific antibodies detailed below. Input DNA control samples were not treated with antibodies. Antibody–chromatin complexes were then pulled down with protein G-Dynabeads (Thermo Fisher, 10003D) and DNA was eluted in 20 µl of 2% SDS and 0.1 M NaCl and with magnetic beads as described previously and quantified using a Qubit 4 Fluorometer (Thermo Fisher, Q33226).

H3K27Ac ChIP sequencing. Active status of chromatin was determined by H3K27Ac levels using ChIP–seq. H3K27Ac CHIP assay was conducted with 5 µg of anti-H3K27Ac antibody (Active Motif, 39133) on the protocol described above. Sequencing libraries were prepared with 3 ng each of H3K27Ac ChIP DNA and input sample using SMARTer ThruPLEX DNA-seq kit (Takara, R00675). Libraries were sequenced on paired-end (SE) 75-cycle reads on an Illumina NextSeq 500 system at the Broad Institute of Harvard and MIT, and the reads were aligned to human reference genome hg19 using Burrows–Wheeler alignment. Genome-wide coverage was calculated after extending to 200 bases (approximate fragment size) and averaged over 25-bp windows using igvtools. Coverage was then normalized and scaled using Rseg (http://rseeg.sourceforge.net/#normalize-bigwig-py/). ChIP–seq peaks were called using MACS2 (2.0.10.201913).

ChIP-qPCR. dCas9 fused to DmrA(×) was pulled down using 5 µg anti-Cas9 antibody (Active Motif, 61757) for each ChIP assay as described above. The immunoprecipitated DNA was analyzed by qPCR using Fast SYBR Green Master Mix (Thermo Fisher, 4385612) with the primers listed in Supplementary Table 3 on a LightCycler 480 (Roche) using the following program: initial denaturation at 95 °C for 20 s followed by 45 cycles of 95 °C for 3 s and 60 °C for 30 s. Relative efficiency for each target was calculated by normalization to input control.

RNA sequencing. RNA libraries were prepared from 500 ng of total RNA treated with Ribogold zero to remove ribosomal RNA, using TruSeq Stranded Total RNA Library Prep Gold kit (Illumina, 20020399) and TruSeq RNA Single Indexes. The RNA libraries were sequenced with SE 75-cycle reads on an Illumina NextSeq 500 at the Broad Institute of Harvard and MIT. Reads were aligned to human reference genome hg19 using STAR (https://star.readthedocs.io/en/stable) and PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/). Reads aligning to ribosomal RNA were then filtered out of the alignment. Genomic coverage from filtered alignments was calculated by normalizing to sequencing depth using BEDTools. FPKM values were calculated using Cufflinks. Differential gene expression was performed using DESeq2 (v1.20.0).

Gene Ontology enrichment analysis. GO analysis was performed using the PANTHER website (http://pantherdb.org)53. A list of genes that showed differential expression after activation of MYOD1 identified by RNA-seq (Supplementary Fig. 2c) was used as the input for the analysis. The PANTHER Overrepresentation Test (released on 7 April 2020) was performed with the Gene Ontology (GO) database (released on 21 February 2020). Fisher’s exact test with false discovery rate correction was used, and GO Biological Process Complete was used as an annotation dataset.

ATAC sequencing. Open or closed status of the chromatin was determined using ATAC–seq. The ATAC–seq libraries were constructed following the protocol of Corces et al.54 and using Nextera DNA Flex Library Prep Kit (Illumina, FC-121-1030). The libraries were sequenced with paired-end (PE) 150-cycle reads on an Illumina NextSeq 500 system at the Broad Institute of Harvard and MIT. Reads were aligned to human reference genome hg19 using Burrows–Wheeler alignment and filtered to exclude PCR duplicates and processed as previously described55. Read start positions were shifted toward the 3′ end by 4 bp for reads aligning to the plus strand and toward the 5′ end by 5bp for reads aligning to the minus strand. Genomic coverage was calculated by counting reads in 150-bp sliding windows at 20-bp steps across the genome and then normalized to 10 million reads in each experimental using BEDTools.

Defining APOC3 enhancer sequences for SNP analysis. Known APOC3 enhancer sequences are located 500 to 890 bp upstream of the TSS56 and show open chromatin and H3K27Ac enrichment features in HepG2 cells in which APOC3 is highly expressed (UCSC Genome Browser (hg19); Supplementary Table 1). We identified potential enhancer sequences in the region encompassing −4.4 Kb to 2 Kb upstream of the TSS based on similar open chromatin and H3K27Ac enrichment features (Extended Data Fig. 4a).

Haplotype analysis. Primers flanking the APOA4 exon 2 SNP (rs5092) and enhancer site E6 (rs2071532) were used to amplify −4.3 kb to 7.2 kb of HEK293 genomic DNA (Supplementary Table 3). Amplicons were TOPO cloned using Zero Blunt TOPO PCR cloning kit (Thermo Fisher, 450031) and 100 colonies for each amplicon were analyzed by Sanger sequencing (Extended Data Fig. 4b).

Allele-selective binding of activators and gene expression experiments. Allele-selective binding of activators to gDNA identified by ChIP, allelic ratio in native gDNA and allele–selective gene expression were determined using next-generation amplicon sequencing. Libraries for sequencing were prepared in two steps by PCR. In the first step, amplicons were amplified by PCR primers that contain Illumina adaptor sequences. The PCR reactions contained 50 ng of gDNA, 5 µl of ChIP DNA or 5 µl of a 1:20 dilution of control DNA, 500 nM each of forward and reverse primer, 200 µM dNTP, 1 unit of Phusion Hot Start Flex

NATURE METHODS | www.nature.com/naturemethods
DNA Polymerase (NEB, M0535L) and 1× Phusion HF buffer in a total volume of 50 μl. The first PCR cycling conditions were 98 °C for 2 min followed by 25 cycles of 98 °C for 10 s, 65 °C for 12 s and 72 °C for 12 s, and a final 72 °C extension for 10 min. PCR products were purified using paramagnetic beads (0.7–1.2× beads to sample ratio) according to ampiclon size as described previously and quantified on a Quibit 4 Fluorometer (Thermo Fisher, Q33262) using 1× DNA high-sensitivity kit (Thermo Fisher, Q33231). Bead-purified ampiclons with Illumina adaptors from the first PCR (1–19 ng) were barcoded with Illumina indexes containing sequences complementary to the adaptor overhangs in a second PCR, using the cycling conditions of 98 °C for 2 min, 7 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 30 s followed by 72 °C for 10 min. The PCR products were purified as above and quantified using a Quibit 4 Fluorometer. Amplicon libraries were sequenced with PE 300-cycle reads on the Illumina MiSeq using the 300-cycle MiSeq Reagent Kit v2 (Illumina, M5-102-2002) or Micro Kit v2 (Illumina, MS-103-2002). Demultiplexed FASTQ files of the Illumina HiSeq data were obtained from https://github.com/FelixKrueger/TritonGalore/, FLASH2 (https://github.com/dstreetett/FLASH2/) and CRISPRESSO2 (ref. 38). Allele-selective expression of APOC3 in HEK293 cells was confirmed by RT-qPCR using allele-specific primers targeting the APOC3 exonic SNP (rs4520) designed according to work by Li et al. for mismatch amplification mutation assays (Extended Data Fig. 3e). All the primers used in the above reactions are listed in Supplementary Table 1. The specificity of the allele-specific primers was verified using U2OS cDNA in which the variant allele is not present (Extended Data Fig. 3e).

K562 Hi-C analysis. We used Juicer to extract high-throughput chromosome conformation capture (Hi-C) contacts from the K562 cell line at the HBG and MYOD1 loci windows at 25-kb resolution, summed them to determine the overall conformation capture (Hi-C) contacts from the K562 cell line at the HBG and MYOD1 loci, respectively.

Comparison of SNP densities at Cas9 PAM sequences in promoters and putative enhancers. For this analysis, promoters were defined as ±500 bp from the TSS, and putative enhancers were determined as DNase hypersensitivity sites (DHSs) excluding promoter sequences described above. NCBI RefSeq version GCF_000004105.25_GRC37.p13 was used for defining the TSS, and 83 DHS tracks of different cells and tissues from the ENCODE Roadmap Project (https://www.encodeproject.org) were combined for the analysis (Supplementary Table 4). All SNPs from the 1000 Genomes Project phase 3 were used for the analysis (https://www.internationalgenome.org/data/). SNP sites were classified into three distinct categories based on their activity on the PAM sites: PAM creation, PAM disruption and mixed (that is, creation and disruption at the same time but on different strands). Based on the overlapping counts of SNPs in promoters and putative enhancers, we defined the SNP density as the number of SNPs in each region divided by the length of each regulatory element; enhancer SNP density indicates the number of SNPs in each DHS divided by the peak size of each DHS, and promoter SNP density indicates the number of SNPs in each promoter divided by 1,000 bp.

Statistical analysis. Gene expression analyses were conducted using Student’s t-test (two-tailed, assuming equal variance) and comparison of SNP densities between promoter and enhancer using the Mann–Whitney U test. The results were considered statistically significant if the P value was less than 0.05.

Data availability
Datasets from amplicon sequencing have been deposited with the NCBI Sequence Read Archive (PRJNA578485). Datasets from Chip-seq, RNA-seq and ATAC–seq experiments have been deposited with the Gene Expression Omnibus repository (GEO) with GEO accession GSE139190. The GO database used in this study can be obtained from https://bioprotal.bioontology.org/ontologies/GO/). Source data are provided with this paper.

References
41. Liang, J. R., Lingeman, E., Ahmed, S. & Corn, J. E. Atlantins remodel the endoplasmic reticulum for selective autophagy. J. Cell Biol. 217, 3354–3367 (2018).
42. Bernstein, B. E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120, 169–181 (2005).
43. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946 (2012).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
45. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer: high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).
46. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
47. Trapnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515 (2010).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 530 (2014).
49. Mi, H. et al. Protocol v2: update for large-scale genome and gene function analysis with the PANTHER classification system (v14.0). Nat. Protoc. 14, 703–721 (2019).
50. Corces, M. R. et al. An improved ATAC–seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962 (2017).
51. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
52. Kitstaki, E., Lacorte, J.-M., Katakriki, N., Zannis, V. I. & Talanidis, I. Transcriptional regulation of the apolipoprotein A-IV gene involves synergy between a proximal orphan receptor response element and a distant enhancer located in the upstream promoter region of the apolipoprotein C-III gene. Nucleic Acids Res. 22, 4689–4696 (1994).
53. Clement, K. et al. CRISPRESSO2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. 37, 224–226 (2019).
54. Li, B., Kadura, L., Pa, A. & Watson, D. E. Genotyping with TaqMAMA. Genomics 83, 311–320 (2004).
55. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98 (2016).
56. Wang, Y. et al. SPIN reveals genome-wide landscape of nuclear compartmentalization. Genome Biol. 22, 36 (2021).
Extended Data Fig. 1 | Chromatin status at IL2RA, CD69, and MYOD1 determined by ATAC-seq and H3K27Ac ChIP-seq.  

**a,** IL2RA promoter was closed and inactive in all cell types, IL2RA enhancer regions were closed and inactive in HEK293 and K562 cells, but open and active in U2OS and HepG2 cells. E1, E2, E3, E4: IL2RA enhancer gRNA target sites, P: IL2RA promoter gRNA target site. The RBM17 locus which was open (transposase accessible chromatin) and active (enriched in H3K27Ac marks) in all cell types is shown for comparison.  

**b,** CD69 promoter was closed and inactive in all cell types, CD69 enhancer regions were closed in all cell types. E1, E2: CD69 enhancer gRNA target sites, P: CD69 promoter gRNA target site.  

**c,** Chromatin at MYOD1 promoter was open in U2OS and HEK293 cells but not in HepG2 and K562 cells. E1, E2, E3, E4: MYOD1 enhancer gRNA target sites, P: MYOD1 promoter gRNA target site.
Extended Data Fig. 2 | Chromatin status at the β-globin locus determined by ATAC-seq and H3K27Ac ChIP-seq. All promoters at the β-globin locus showed closed and inactive chromatin states in all cell types. HS2 enhancer region showed closed and inactive chromatin features in HEK293 cells, but open and active chromatin features in U2OS and HepG2 cells. E: HS2 enhancer gRNA target site, P_E: HBE1 promoter gRNA target site, P_G: HBG1/2 promoter gRNA target site, P_B: HBB promoter gRNA target site.
Extended Data Fig. 3 | Total activation of APOA4 and APOC3 and orthogonal confirmation of allele-selective expression via allele-specific RT-qPCR.

a-b, Total expression of APOA4 and APOC3 in HEK293 cells by bi-partite p65 aTF targeting the promoter (P) and various sites on the enhancers including SNP regions (E1 to E6) and non-SNP region (E0) determined by RT-qPCR, normalized to HPRT1 levels, calculated relative to sample with non-targeting gRNA (None). Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m.

c, Schematic of the location of RT-qPCR primers used for APOC3 allele-selective expression. Allele-specific primers detecting a SNP in APOC3 exon 3 have a common forward primer (Pf) which spans the exon 2 and exon 3 junction, and two different reverse primers which are specific for allele 1 (T at rs4520, Pr_1) or for allele 2 (C at rs4520, Pr_2) in exon 3, with a ‘T’ mismatch in the penultimate base at the 3’ for both primers.

d, Allele-selective expression of APOC3 in HEK293 cells by bi-partite p65 aTF targeting the promoter (P) and various sites on the enhancers including SNP regions (E1 to E6) and non-SNP region (E0) determined by RT-qPCR using the primers described in c, normalized to HPRT1 levels, calculated relative to sample with non-targeting gRNA. Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m. The apparent difference in allele-specific expression levels when compared to total expression in b is potentially due to the amplification of a smaller fragment of cDNA in the allele-specific reaction.

e, Validation of the specificity of allele-specific RT-qPCR primers used in d, with U2OS cells in which the variant T nucleotide is absent at rs4520 (only the C nucleotide is present at the same position). APOC3 expression was measured by RT-qPCR using the allele-specific primers used in d, in U2OS cells co-expressing the bi-partite p65 aTF and gRNAs targeting the promoter or non-SNP region of the enhancer (E0). Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m.
Extended Data Fig. 4 | Haplotype of potential APOA4 and APOC3 enhancer regions and allele ratios of target SNPs. **a,** The potential enhancer region was identified by its open and active chromatin features which are similar to the known enhancer, based on the DNase-seq and H3K27Ac data (UCSC genome browser) from HepG2 cells in which APOC3 is highly expressed. Genomic locations of SNPs in SpCas9 PAMs identified in the potential enhancer are shown. SNPs in exon 2 of APOA4 and exon 3 of APOC3 are shown. **b,** Sanger sequencing traces from TOPO cloned amplicons showing the SNPs in the potential enhancer and exonic regions of APOA4 and APOC3 in HEK293 cells. E1 to E6 are gRNA binding sites in the potential enhancer region which has SNPs in the PAM sequence. SNPs are exclusively associated with one another in two unique haplotypes. **c,** Allele ratios of target SNPs in the genomic DNA of HEK293 cells were determined by targeted genomic DNA amplicon sequencing and indicate a 1:1 ratio.
Extended Data Fig. 5 | Binding of bi-partite p65 aTF to APOA4 and APOC3 promoter and enhancer target sites in HEK293 cells. a, Genomic locations of the enhancer gRNAs, APOA4 promoter and APOC3 promoter gRNA. The regions amplified in ChIP-qPCR assays are shown as boxes. b, Binding activity of bi-partite p65 aTF at each gRNA target region in APOA4 and APOC3 loci determined by Cas9 ChIP-qPCR, expressed as a percentage of input DNA. Two sets of primers designed to amplify the human genome at locations other than the APOC3 and APOA4 loci were used as negative controls. Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m. c, Binding of the bi-partite p65 aTF to the potential upstream enhancer sequence in the presence of the E1-E6 gRNAs. E1, E2, and E4 are expected to bind selectively to Allele 1 (yellow); E3, E5, and E6 to Allele 2 (orange). Relative quantification (percent next-generation sequencing reads) of the two alleles in the DNA from ChIP experiments performed with an anti-Cas9 antibody are shown. Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m. In, Input DNA; Ch, Cas9 ChIP DNA.
Extended Data Fig. 6 | Allele-selective upregulation of HBB genes in U2OS and K562 cells using heterotopic activation of enhancer. 

**a.** Schematic of HBB gene and the three alleles present in U2OS cells. P indicates the binding site for the gRNA targeting the HBB promoter. E1–E4 indicate binding sites for gRNAs in the HS4 putative enhancer region, which are expected to target either all alleles (E1), selectively target one allele (E2, E3) or two alleles (E4) based on the PAM of the target site (black bold indicates a base that maintains an intact PAM site and gray bold indicates a base that is expected to disrupt the PAM). A SNP in exon 1 of HBB distinguishes between allele 1 (light purple) and allele 2/3 (pink).

**b.** Total expression of HBB in U2OS cells when the bi-partite p65 aTF was co-expressed with a gRNA targeting the promoter (P) and/or with one or more gRNAs targeting the HS4 enhancer region (E1–E4) was determined by RT-qPCR, normalized to HPRT1 levels, and calculated relative to sample with non-targeting gRNA (None). Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m.

**c.** Relative quantification (percent next-generation sequencing reads of cDNA) of the three alleles of HBB mRNA when the bi-partite p65 aTF was co-expressed with a gRNA targeting the promoter (P) alone or with one or more gRNAs targeting the HS4 enhancer region (E1–E4). Open circles indicate biological replicates (n = 3), bars the mean of replicates and error bars the s.e.m.
Extended Data Fig. 7 | Allele-selective upregulation of MYOD1 genes in U2OS and K562 cells, respectively, using heterotopic activation of enhancer.

**a**, Schematic of MYOD1 gene and the three alleles present in K562 cells. P indicates the binding site for the gRNA targeting the MYOD1 promoter. E1-E4 indicate binding sites for gRNAs in the known enhancer region termed the distal regulatory region (DRR), which are expected to selectively target allele 1 (E1, E3, E4) or all alleles (E2) based on the PAM of the target site (black bold indicates a base that maintains an intact PAM site and gray bold indicates a base that is expected to disrupt the PAM). A SNP in exon 3 of MYOD1 distinguishes between allele 1 and allele 2/3. **b**, Total expression of MYOD1 in K562 cells when the bi-partite p65 ATF was co-expressed with a gRNA targeting the promoter (P) and/or with one or more gRNAs targeting the DRR enhancer region (E1-E4) was determined by RT-qPCR, normalized to HPRT1 levels, and calculated relative to sample with non-targeting gRNA (None). Open circles indicate biological replicates (n=3), bars the mean of replicates and error bars the s.e.m. **c**, Relative quantification (percent next-generation sequencing reads of cDNA) of the three alleles of MYOD1 mRNA when the bi-partite p65 ATF was co-expressed with a gRNA targeting the promoter (P) alone or with one or more gRNAs expected to target the DRR enhancer region (E1-E4). Open circles indicate biological replicates (n=3), bars the mean of replicates and error bars the s.e.m.
Extended Data Fig. 8 | Distribution of SNP densities that create or disrupt NGG PAM sequences at putative enhancers and promoters. The density of SNPs is the number of SNPs divided by the base pair size of each regulatory element (promoter or enhancer) identified from the 1000 Genomes Project using DHS data from 83 different cell lines from ENCODE/Roadmap project. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. *** indicates p < 0.001 (Mann-Whitney U test (two-sided) with Bonferroni test for multiple comparison).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [ ] Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

**Data collection**

Illumina Base Space was used to collect high-throughput sequencing data from Nextseq sequencer

**Data analysis**

Alignment of RNA-seq reads to the human genome was performed with STAR 2.4.0h. Genomic coverage from filtered alignments were calculated by normalizing to sequencing depth using bedtools v.2.29.2. PCR duplicates were removed using Picard tools v.2.21.7. FPKMs were calculated using Cufflinks v2.2.1. Differential gene expression was performed using DESeq2 v.1.20.0. Alignment of ChIP-seq reads to the human genome was performed with bwa 0.6.2. Genome-wide coverage was calculated after extending to 200 bases and averaged over 25 bp windows using ıgtools 2.9. Coverage was then normalized and scaled using RSeqC v3.0.1 (http://rseqc.sourceforge.net/index.html). ChIP-seq peaks were called using MACS2 2.0.10.20120913. ATAC-seq reads were aligned to the human genome using bwa 0.7.12-r1039. Genomic coverage was calculated by counting reads in 150 bp sliding windows at 20 bp steps across the genome and then normalized to 10 million reads in each experiment using bedtools v2.21.7. Frequency, mean, and standard error of the mean were calculated using GraphPad Prism 8. For amplicon sequencing, demultiplexed FASTQ files were analyzed using TrimGalore v.0.6.5 (https://github.com/FelixKrueger/TrimGalore), FLASH2 2.2.0.0 (http://github.com/dstree1/FLASH2) and CRISPResso2. These are not custom codes.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Amplon sequencing data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA578485. Data sets from ChIP-seq, RNA-seq, and ATAC-seq experiments have been deposited with the GEO with the accession number GSE139190.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.
☐ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/mr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No statistical methods were used to predetermine sample size. Sample sizes were determined based on previous published literature (Supplementary Note) for epigenetic editing experiments |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data was excluded. |
| Replication | All experiments were repeated at least three times. All attempts were successful. |
| Randomization | All experiments were neither randomized nor blinded, as it is not relevant to the study. For each experimental group, all samples were prepared within the same batch to control. |
| Blinding | All experiments were neither randomized nor blinded, as it is not relevant to the study. For each experimental group, all samples were prepared within the same batch to control. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a | n/a |
| ☐ X Antibodies | ☐ X ChIP-seq |
| ☐ X Eukaryotic cell lines | ☐ X Flow cytometry |
| ☐ X Palaeontology | ☐ X MRI-based neuroimaging |
| ☐ X Animals and other organisms | |
| ☐ X Human research participants | |
| ☐ X Clinical data | |

Antibodies

| Antibodies used | Validation |
|-----------------|------------|
| H3K27Ac (Active motif, cat#39133, lot#10918010): Sug used without dilution, Cas9 (Active motif, cat#61757, lot#10216001): Sug used without dilution | H3K27Ac:https://www.activemotif.com/catalog/details/39133 Cas9:https://www.activemotif.com/catalog/details/61757/cas9-antibody-mab-clone-8c1-f10 |
### Eukaryotic cell lines

| Policy information about **cell lines** |
|----------------------------------------|
| **Cell line source(s)**                |
| HEK293 (Invitrogen, similar to ATCC CRL-1573), a loss of two alleles, #9 at the DSS818 locus and the #11 at the CSF1PO locus), U2OS (gift of Dr. Toni Cathomen, similar match to ATCC HTB-96, gain of no. 8 allele at the DSS818 locus), HepG2 (ATCC HB-8065), K562 (ATCC CCL-243) |
| **Authentication**                      |
| Cells were authenticated by using STR analysis from ATCC. |
| **Mycoplasma contamination**           |
| All cell lines tested negative for mycoplasma. |
| **Commonly misidentified lines**      |
| (See [ICLAC register](https://www.iclacr.com)) | None |

### ChiP-seq

#### Data deposition

- **Confirm that both raw and processed data have been deposited in a public database such as** [GEO](https://www.ncbi.nlm.nih.gov/geo/)
- **Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.**

#### Data access links

- **May remain private before publication.**
- **To review GEO accession GSE139190:**
  - Go to [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139190](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139190)
  - Enter token iqvdkwwqtkbnvmd into the box
- **Files in database submission**
  - YET163_rep1_U2OS_H3K27Ac.bw, YET163_fastq.gz, YET207_rep2_U2OS_H3K27Ac.bw, YET207_fastq.gz, YET159_rep1_HEK293_H3K27Ac.bw, YET159_fastq.gz, YET160_rep2_HEK293_H3K27Ac.bw, YET160_fastq.gz, YET165_rep1_HepG2_H3K27Ac.bw, YET165_fastq.gz, YET166_rep2_HepG2_H3K27Ac.bw, YET166_fastq.gz, YET161_rep1_K562_H3K27Ac.bw, YET161_fastq.gz, YET162_rep2_K562_H3K27Ac.bw, YET162_fastq.gz, YET168_input_U2OS.bw, YET168_fastq.gz, YET167_input_HEK293.bw, YET167_fastq.gz, YET170_input_HepG2.bw, YET170_fastq.gz, YET169_input_K562.bw, YET169_fastq.gz
- **Genome browser session**
  - [https://genome.ucsc.edu/s/Fasthe/H3K27Ac_Chip%2Dseq_HEK293_K562_U2OS_K562](https://genome.ucsc.edu/s/Fasthe/H3K27Ac_Chip%2Dseq_HEK293_K562_U2OS_K562)

#### Methodology

- **Replicates**
  - YET163 53466, YET207 56343, YET159 20646, YET160 16248, YET165 55324, YET166 36642, YET161 37038, YET162 40045, YET163 53466, YET207 56343, YET159 20646, YET160 16248, YET165 55324, YET166 36642, YET161 37038, YET162 40045
- **Sequencing depth**
  - All samples are single-end, 75 bp reads; Sample name, total number of reads, uniquely mapped reads; YET163, 14127383, 13191110; YET207, 18101844, 16862963; YET159, 13211176, 12156447; YET160, 11055103, 10872947; YET165, 12057690, 12099886; YET166, 4008400, 350249; YET161, 13362338, 12374343; YET162, 13522567, 12546004; YET168, 19971814, 17801191; YET167, 17088778, 15363436; YET170, 37906836, 33768171; YET169, 26926209, 24012059
- **Antibodies**
  - H3K27Ac [Active motif, cat#3933, lot#10918010] Car9 [Active motif, cat#61757, lot#10216001]
- **Peak calling parameters**
  - `macs2 callpeak --nomodel -B --SPMR -g hs -q 0.001 -t <path to treated bam> -c <path to control bam> -n <sample name>`
- **Data quality**
  - Sample name, the number of peaks that are at FDR 5% and above 5-fold enrichment; YET163 53466, YET207 56343, YET159 20646, YET160 16248, YET165 55324, YET166 36642, YET161 37038, YET162 40045, YET163 53466, YET207 56343, YET159 20646, YET160 16248, YET165 55324, YET166 36642, YET161 37038, YET162 40045
- **Software**
  - bwa-0.6.2, picard-tools-1.84, macs2 2.0.10.20120913