

Targeted transcriptional modulation with type I CRISPR–Cas systems in human cells

Adrian Pickar-Oliver^{1,2}, Joshua B. Black^{1,2}, Mae M. Lewis^{1,2}, Kevin J. Mutchnick^{1,2}, Tyler S. Klann^{1,2}, Kylie A. Gilcrest^{1,2}, Madeleine J. Sitton^{1,2}, Christopher E. Nelson^{1,2}, Alejandro Barrera³, Luke C. Bartelt^{2,3}, Timothy E. Reddy^{1,2,3,4}, Chase L. Beisel^{5,6,7}, Rodolphe Barrangou⁸ and Charles A. Gersbach^{1,2,9*}

Class 2 CRISPR–Cas systems, such as Cas9 and Cas12, have been widely used to target DNA sequences in eukaryotic genomes. However, class 1 CRISPR–Cas systems, which represent about 90% of all CRISPR systems in nature, remain largely unexplored for genome engineering applications. Here, we show that class 1 CRISPR–Cas systems can be expressed in mammalian cells and used for DNA targeting and transcriptional control. We repurpose type I variants of class 1 CRISPR–Cas systems from *Escherichia coli* and *Listeria monocytogenes*, which target DNA via a multi-component RNA-guided complex termed Cascade. We validate Cascade expression, complex formation and nuclear localization in human cells, and demonstrate programmable CRISPR RNA (crRNA)-mediated targeting of specific loci in the human genome. By tethering activation and repression domains to Cascade, we modulate the expression of targeted endogenous genes in human cells. This study demonstrates the use of Cascade as a CRISPR-based technology for targeted eukaryotic gene regulation, highlighting class 1 CRISPR–Cas systems for further exploration.

The ability to modulate and perturb genetic information is indispensable for studying gene function and elucidating biological mechanisms. Targetable DNA-binding proteins that modify genomes at specific loci have led to substantial advances in science, biotechnology and medicine¹. Specifically, the development of genome engineering tools based on class 2 CRISPR–Cas systems, which use a single effector protein such as Cas9 or Cas12, has revolutionized the field due to the ease of use of this technology in a vast range of applications². In fact, bioinformatics analyses have further revealed a diverse range of CRISPR–Cas systems, and the most recent classification encompasses six major types (I–VI)³. However, single-effector class 2 systems (types II, V and VI) have most commonly been used for nucleic acid targeting in eukaryotes, despite multi-subunit class 1 systems (types I, III and IV) comprising about 90% of all identified systems across bacteria and archaea⁴. Continued efforts to discover and develop single-component class 2 CRISPR effectors beyond Cas9-based type II systems have resulted in new technologies with specific advantages or applications. For example, the Cas12-based type V⁵ and Cas13-based type VI⁶ CRISPR–Cas systems have distinct targeting and editing mechanisms. Here, we describe the development of type I systems, which account for more than 50% of all identified CRISPR–Cas loci, for use as programmable transcriptional activators and repressors in mammalian cells.

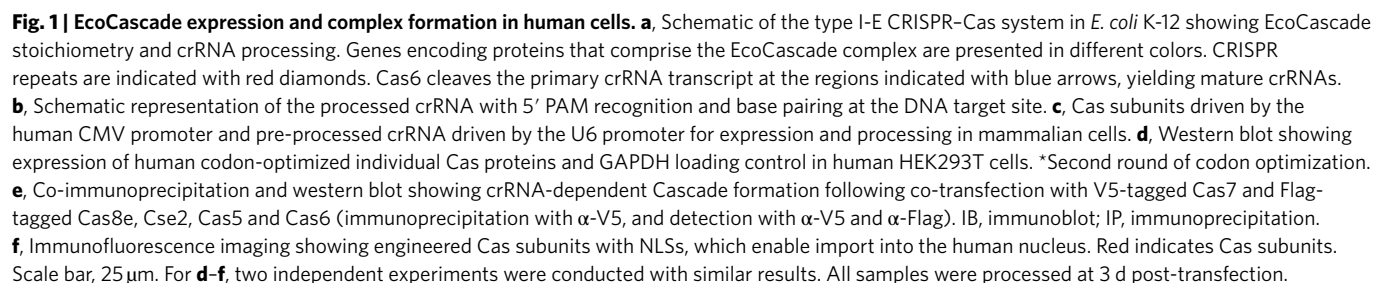
Type I CRISPR–Cas systems use the signature Cas3 nuclease–helicase to eliminate invading DNA, and are further divided into eight subtypes (I–A to I–G and I–U) on the basis of related, but subtype-specific, accessory *cas* genes^{3,4,7}. The well-studied

prototypical type I–E system of *Escherichia coli* K12 consists of eight *cas* genes and a downstream CRISPR array^{8–10}. Following transcription of the CRISPR array, the 29-bp repeat sequences flanking the variable spacer sequences are cleaved during CRISPR RNA (crRNA) biogenesis by the Cas6 endoribonuclease¹⁰. Together, five protein subunits (Cas8e, Cse2, Cas7, Cas5 and Cas6, previously referred to as CasA, CasB, CasC, CasD and CasE, respectively)⁴ and the mature 61-nucleotide crRNA form the 405 kDa CRISPR-associated complex for antiviral defense (Cascade)^{11–13} (Fig. 1a). Unlike type II CRISPR systems, there is no trans-activating crRNA (tracrRNA) required for effector complex formation. For type I CRISPR systems, Cas6 processes the CRISPR array and targeting relies on a single crRNA. To bind to a target, Cascade surveys the DNA to find a protospacer-adjacent motif (PAM) upstream of a target sequence with complementarity to the crRNA spacer sequence (Fig. 1b). We sought to harness this prokaryotic immune system for genome targeting in eukaryotes.

Results

Expressing the Cascade complex in human cells. To repurpose Cascade for use in mammalian cells, we used a cytomegalovirus (CMV) promoter to express each Cascade subunit of the *E. coli* K12 system (EcoCascade). Based on available structural information^{13–16}, N-terminal Flag epitope tags and nuclear localization signals (NLSs) were attached to each EcoCascade construct. We used the RNA polymerase III U6 promoter to express target spacers flanked by full repeat sequences for crRNA processing by Cas6 (Fig. 1c)¹⁷. Heterologous expression of all EcoCascade constructs

¹Department of Biomedical Engineering, Duke University, Durham, NC, USA. ²Center for Advanced Genomic Technologies, Duke University, Durham, NC, USA. ³Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA. ⁴University Program in Genetics and Genomics, Duke University, Durham, NC, USA. ⁵Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA. ⁶Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Centre for Infection Research (HZI), Würzburg, Germany. ⁷Medical Faculty, University of Würzburg, Würzburg, Germany. ⁸Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, USA. ⁹Department of Surgery, Duke University Medical Center, Durham, NC, USA. *e-mail: charles.gersbach@duke.edu



Programmable transcriptional activation by EcoCascade-p300.

Next, we sought to repurpose EcoCascade for CRISPR-based programmable transcriptional activation in mammalian cells. Transcriptional

modulation using class 2 CRISPR systems has been achieved by introducing point mutations into the endonucleolytic domains of single-component effectors to maintain binding but not cleavage of the target DNA¹⁹. For example, the nuclease-deficient Cas9 (dCas9) can be engineered to function as a synthetic transcriptional activator by genetic fusion to a transactivation or epigenome-modifying domain^{20–24}. In the natural type I CRISPR–Cas immune system, target site recognition by Cascade leads to recruitment of the Cas3 nuclease to eliminate target DNA. However, deletion of *cas3* from the endogenous type I-E system in *E. coli* has been used for a CRISPR interference strategy in bacteria by permitting Cascade to bind to target DNA and block transcription without DNA degradation²⁵. Therefore, we proposed that Cascade can be repurposed as a programmable DNA-binding technology in eukaryotes by neglecting to express *cas3*.

To repurpose EcoCascade as a programmable transcriptional activator, we explored the potential of various Cas-effector subunits for tethering of the activation domain. We previously demonstrated robust endogenous gene activation with dCas9 fused to the catalytic core domain of the human acetyltransferase p300 (ref. ²³). The EcoCascade system contains five Cas subunits available at various stoichiometries (Fig. 1a), providing versatile options for synthetic fusions to transcriptional regulatory domains and other modular engineering strategies. Following heterologous expression of EcoCascade with p300 fused to Cas8e, Cse2, Cas5 or Cas6, we confirmed EcoCascade complex formation by co-immunoprecipitation (Fig. 2a). The ability of each of these subunits to accommodate the fusion of the large p300 core catalytic domain (72 kDa) without abrogating complex formation suggests that the modular Cascade complex could be particularly useful for multiplexed targeting of regulatory domains at specific loci.

To test programmable endogenous gene activation in human cells, a panel of crRNAs was generated tiling the endogenous *IL1RN* promoter at protospacer targets downstream of known PAMs (5'-AAG, AGG, ATG, GAG, TAG-3')^{10,26–28} (Fig. 2b and Supplementary Table 1). Co-transfection of HEK293T cells with plasmids encoding EcoCascade with Cas6–p300 and individual crRNAs revealed robust *IL1RN* activation with many crRNAs, including more than 3,000-fold *IL1RN* activation with cr26 ($P < 0.001$, Fig. 2c). Importantly, cr26 with EcoCascade lacking a p300 domain, or cr26 alone, did not activate *IL1RN* (Supplementary Fig. 1), suggesting target-specific activation by EcoCascade–p300. On the basis of EcoCascade stoichiometry (Fig. 1a) and relative Cas construct expression (Fig. 1d), we optimized the relative masses of transfected plasmids to maximize gene activation (Supplementary Fig. 2). Additionally, the transactivation potential of all Cas–p300 fusions was explored with cr26. Relative to heterologous expression with a crRNA targeted to a control locus, EcoCascade containing Cas8e–p300 or Cas6–p300 achieved significant transactivation of *IL1RN* ($P < 0.05$, Fig. 2d).

Versatility of EcoCascade for targeted gene activation. To investigate EcoCascade–p300 interactions at the target locus, we

performed chromatin immunoprecipitation with an anti-Flag antibody followed by quantitative PCR (ChIP–qPCR) of two amplicons adjacent to the target site (Fig. 2b). We observed significant enrichment of the target regions in EcoCascade–p300 samples co-transfected with cr25 or cr26 compared with samples transfected with a control crRNA ($P < 0.001$, Fig. 2e). These results confirm EcoCascade as a programmable DNA-binding platform for efficient targeting of specific loci in the human genome. Intriguingly, we observed enhanced enrichment of *IL1RN* signal in samples treated with Flag-tagged EcoCascade–p300 and an *IL1RN*-targeted crRNA compared with samples treated with Flag-tagged dCas9–p300 and an *IL1RN*-targeted single-guide RNA (sgRNA) (Supplementary Fig. 3). These findings indicate increased occupancy by EcoCascade relative to dCas9 but could also be the result of differences in epitope avidity or presentation. Targeted endogenous *IL1RN* activation was also achieved by tethering Cas6 to the tripartite activator VP64–p65–Rta (VPR)²⁴, although both p300 and VPR tethered to Cas6 led to reduced activation levels compared with fusion to dCas9 (Supplementary Fig. 4). To assess activation of other endogenous targets in the human genome, we targeted the *HBG* promoter with EcoCascade–p300 (Fig. 2f) and observed robust transactivation (Fig. 2g). The natural role of Cascade in processing crRNAs suggests the possibility of using arrayed spacers for multiplexed genome engineering. By generating a CRISPR array containing multiple crRNA spacers that target both *IL1RN* and *HBG*, we demonstrated multiplexed activation of endogenous genes (Supplementary Fig. 5). Together, these results demonstrate the potential for repurposing type I-E Cascade systems as versatile programmable transcriptional activators in mammalian cells.

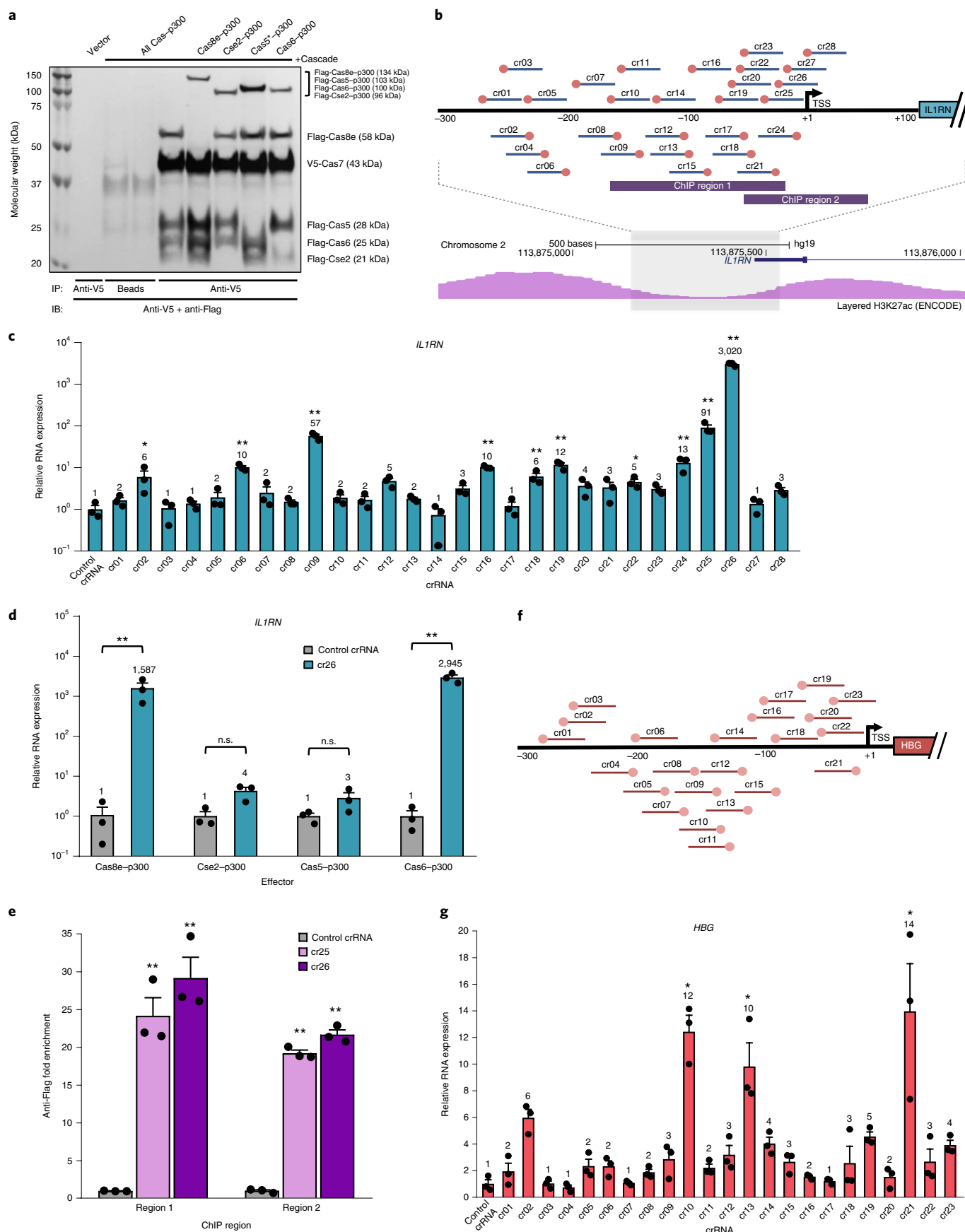
Highly specific crRNA-dependent EcoCascade–p300 targeting.

Rigorously assessing the specificity of genome and epigenome engineering tools is essential for their successful application in basic research and medicine. To date, there have been varied reports regarding the specificity of Cas9-based genome- and epigenome-editing technologies^{29–31}, which has led to the development of a variety of strategies to improve specificity³². To quantify the genome-wide binding specificity of EcoCascade–p300, we performed chromatin immunoprecipitation followed by sequencing (ChIP–seq) using the Flag epitope fused to the N terminus of each of the EcoCascade subunits. Binding of EcoCascade–p300 was highly enriched at the *IL1RN* promoter when targeting with cr25 and cr26, with no detectable binding observed with a *HBE1*-targeting control crRNA (Fig. 3a). Interestingly, the strength of binding signal was comparable between cr25 and cr26, although they differed substantially in their induction of *IL1RN* transcription (Fig. 2c). With a genome-wide false discovery rate (FDR) of less than 0.001, a few off-target differential binding sites were observed when comparing cr25 and cr26 to control crRNA (Fig. 3b,c). These sites all had substantially weaker binding signals than the signal at the *IL1RN* locus, with the exception of one genomic window that was significantly enriched with

Fig. 2 | EcoCascade activates transcription of endogenous genes in human cells. **a**, Co-immunoprecipitation showing EcoCascade formation following co-transfection with plasmids encoding the crRNA, Flag-tagged EcoCascade subunits with V5-tagged Cas7 and various Cas–p300 fusions. Immunoprecipitation performed with α -V5 and immunoblot performed with α -V5 and α -Flag. Two independent experiments were conducted with similar results. *Second round of codon optimization. **b**, Schematic of the *IL1RN* locus along with tiled *IL1RN* crRNA target sites. Enrichment of H3K27 acetylation (H3K27ac) from the ENCODE Consortium is shown with the vertical range set to 400 to indicate regulatory regions. The two *IL1RN* ChIP–qPCR amplicons are shown in corresponding locations. **c**, Relative *IL1RN* expression following co-transfection of plasmids encoding individual crRNAs and EcoCascade with Cas6–p300. Three biological independent samples, mean $\pm$ s.e.m. **d**, Relative *IL1RN* expression following co-transfection of control crRNA or cr26 and EcoCascade with various Cas–p300 effectors. Three biological independent samples, mean $\pm$ s.e.m. n.s., non-significant. **e**, ChIP–qPCR enrichment following co-transfection of individual crRNAs with EcoCascade and Cas6–p300. Immunoprecipitation performed with α -Flag and qPCR performed with primers for the amplicon regions designated in part **b**. Three biological independent samples, mean $\pm$ s.e.m., bars indicate mean fold enrichment. **f**, Schematic of the *HBG* locus along with *HBG* crRNA target sites. **g**, Relative *HBG* expression following co-transfection of individual crRNAs and EcoCascade with Cas6–p300. Three biological independent samples, mean $\pm$ s.e.m. All samples were processed at 3 d post-transfection. Tukey test following log transformation, ** $P < 0.001$ and * $P < 0.05$ compared to control crRNA. Numbers above bars indicate mean relative expression. TSS, transcription start site.

cr26 only, located on chromosome 6 about 5 kilobases (kb) from *TAAR4P* (Fig. 3c and Supplementary Fig. 6a). However, we could not readily detect a seed sequence complementary to cr26 within

this region. A site near the *UBB* locus was enriched in the control crRNA-treated sample relative to both cr25- and cr26-treated samples, indicating a possible off-target site for this crRNA (Fig. 3b,c).



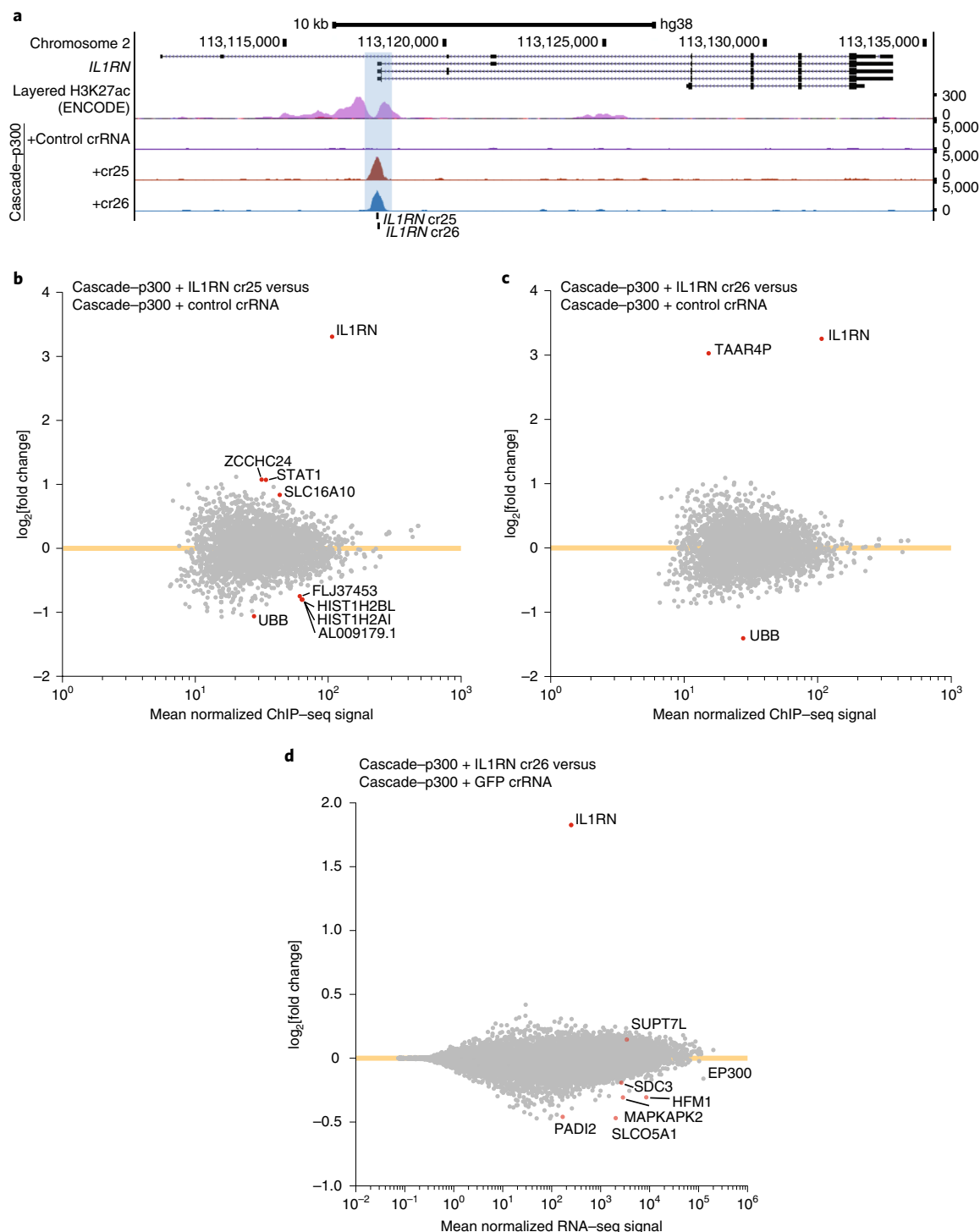


Fig. 3 | Genome-wide specificity of EcoCascade-p300 targeted to *IL1RN*. **a**, ChIP-seq tracks for binding of Flag-tagged EcoCascade-p300 targeted to the *IL1RN* promoter with cr25 and cr26, compared with binding of EcoCascade-p300 with control crRNA. An ENCODE H3K27ac track is included to highlight the regulatory regions. **b,c**, MA plots for the differential analyses of global binding activity including comparisons of EcoCascade-p300 targeted by cr25 (**b**) and cr26 (**c**) versus EcoCascade-p300 with *HBE1*-targeting control crRNA in HEK293T cells. Red data points indicate FDR < 0.001 by differential DESeq2 analysis using Wald test *P* values. **d**, MA plots for differential expression analyses comparing EcoCascade-p300 targeted by cr26 versus *GFP*-targeting control crRNA in HEK293T cells. Red data points indicate FDR < 0.01 by differential expression analysis using Wald test *P* values. Average of three biological independent samples.

To evaluate the specificity of crRNA-mediated endogenous gene activation with EcoCascade-p300, we performed RNA sequencing (RNA-seq) to quantify transcriptome-wide changes when targeting *IL1RN* with cr26 or with a *GFP*-targeting control crRNA (Fig. 3d).

Targeted gene activation was highly specific to the target gene when EcoCascade-p300 was co-expressed with cr26, with detection of a modest change in the activation of only six other genes (Fig. 3d). However, we observed that addition of the p300 domain to

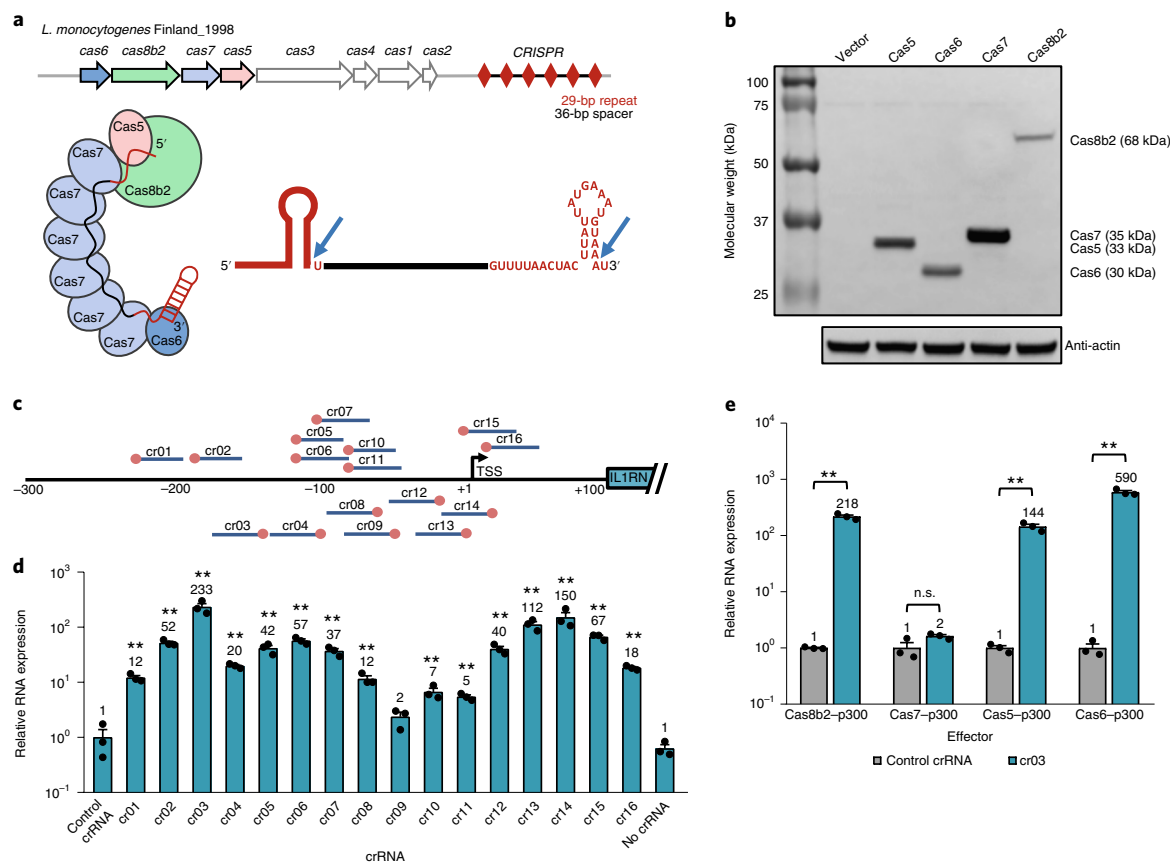


Fig. 4 | LmoCascade activates transcription of *IL1RN* gene in human cells. **a**, Schematic of the type I-B CRISPR-Cas system in *L. monocytogenes* Finland_1998 showing predicted LmoCascade stoichiometry and crRNA processing. Genes encoding proteins that comprise the LmoCascade complex are presented in different colors. CRISPR repeats are indicated with red diamonds. The primary crRNA transcript is predicted to be cleaved at the regions indicated with blue arrows, yielding mature crRNAs. **b**, Western blot showing expression of human codon-optimized individual Cas proteins and actin loading control in human HEK293T cells. Two independent experiments were conducted with similar results. **c**, Schematic of the *IL1RN* locus along with tiled crRNA target sites. TSS, transcription start site. **d**, Relative *IL1RN* expression following co-transfection of individual crRNAs and LmoCascade with Cas6-p300. **e**, Relative *IL1RN* expression following co-transfection of control crRNA or cr03 and LmoCascade with various Cas-p300 effectors. All samples were processed at 3 d post-transfection. Tukey test following log transformation, ** $P < 0.001$ compared to control crRNA. Three biological independent samples. Error bars, s.e.m. Numbers above bars indicate mean relative expression.

EcoCascade resulted in significant off-target transcriptional changes compared to EcoCascade alone (Supplementary Fig. 6b). Given the highly specific DNA targeting by EcoCascade (Fig. 3b,c) and cr26-dependent activation of *IL1RN* (Fig. 3d), these results indicate non-specific crRNA-independent effects of overexpression of the p300 acetyltransferase fused to Cas6. Collectively, this genome-wide specificity analysis demonstrates highly specific crRNA-dependent targeting of EcoCascade in mammalian cells.

Expanding the Cascade toolbox with LmoCascade. Beyond the well-characterized EcoCascade system, bioinformatic analyses have revealed a plethora of additional type I CRISPR-Cas systems. To explore the potential for repurposing other Cascade complexes, we extended our results with the model type I-E EcoCascade to repurpose the type I-B CRISPR-Cas system of *Listeria monocytogenes* Finland_1998 (LmoCascade) (Fig. 4a). Expression of all subunits was confirmed in HEK293T cells following human codon optimization with N-terminal Flag epitope tags and NLSs attached to each LmoCascade construct (Fig. 4b). To repurpose LmoCascade as a programmable transcriptional activator, we fused the catalytic core domain of p300 to Cas6, the predicted EcoCascade Cas6 ortholog. To test programmable endogenous gene activation in human cells, a panel of crRNAs with a predicted spacer length of 36 nucleotides was generated by tiling the endogenous *IL1RN* promoter at

protospacer targets downstream of the known PAM (5'-CCA-3')³³ (Fig. 4c). Co-transfection of HEK293T cells with plasmids encoding LmoCascade with Cas6-p300 and individual crRNAs revealed robust *IL1RN* activation with most crRNAs ($P < 0.001$, Fig. 4d). Additionally, the transactivation potential of all Cas-p300 fusions was explored with cr03. Relative to heterologous expression with a control crRNA, LmoCascade containing Cas8b2-p300, Cas5-p300 or Cas6-p300 achieved significant transactivation of *IL1RN* ($P < 0.001$, Fig. 4e). In contrast to the panel of *IL1RN*-targeting crRNAs for EcoCascade (Fig. 2c), almost all of the LmoCascade crRNAs, as well as three of the four Cas-p300 effector fusions, achieved significant *IL1RN* activation (Fig. 4d,e). To further assess crRNA-dependent activation of other endogenous targets in the human genome, we targeted the *HBG* promoter with LmoCascade-p300 and observed robust transactivation (Supplementary Fig. 7). These results demonstrate the potential to broaden our fundamental knowledge of type I CRISPR systems as we expand the CRISPR engineering toolbox by repurposing less characterized systems for use in mammalian cells.

Targeted gene repression by stable LmoCascade-KRAB expression. In addition to harnessing type I CRISPR systems for programmable transcriptional activation, we sought to take advantage of steric hindrance by the large Cascade complex, the strong binding of

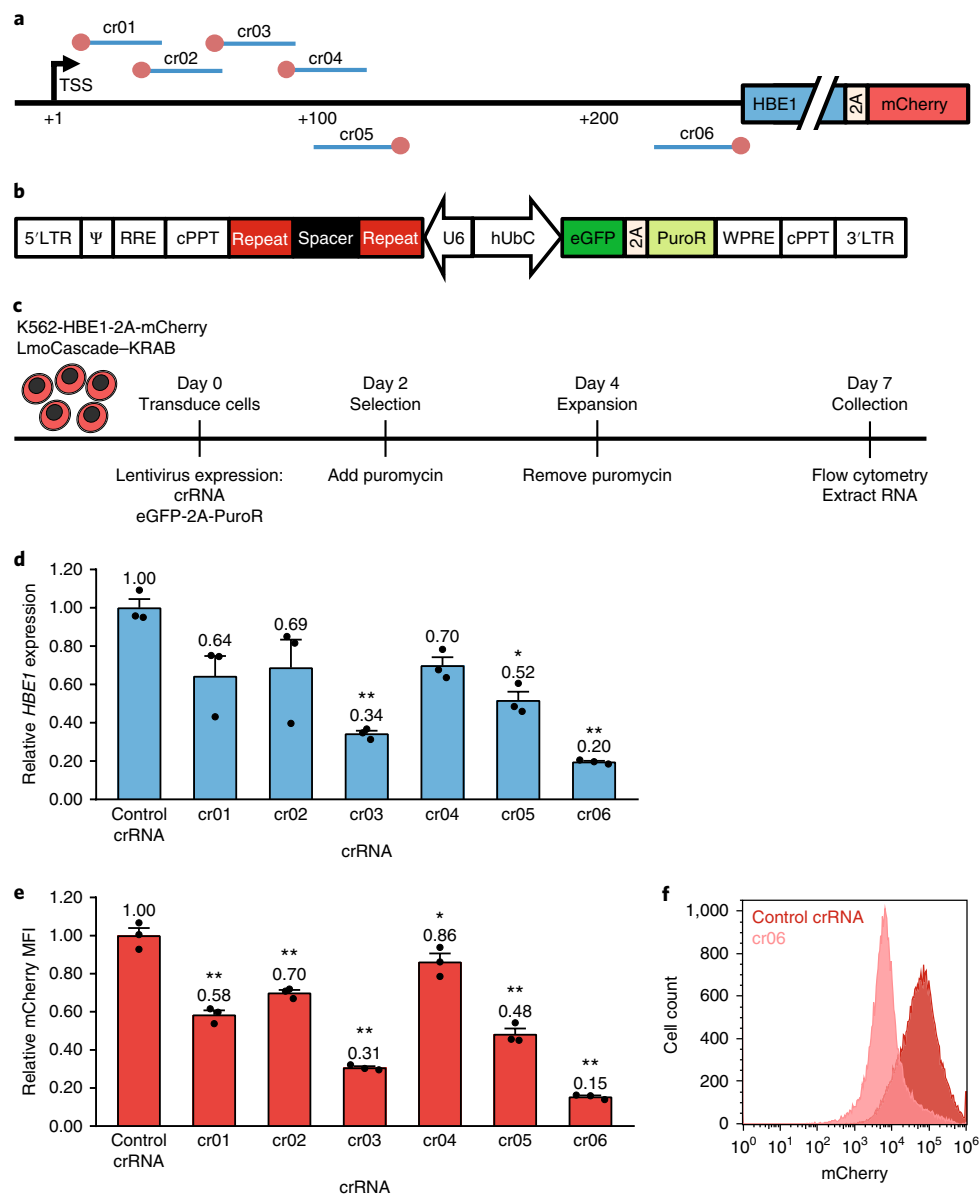


Fig. 5 | LmoCascade represses transcription of *HBE1* gene in human cells. **a**, Schematic of the *HBE1* locus along with tiled *HBE1* crRNA target sites. TSS, transcription start site. **b**, Schematic of lentiviral expression constructs with dual promoters for co-expression of crRNAs with eGFP and Puromycin^R (PuroR). **c**, Experimental timeline for transduction of monoclonal K562-HBE1-2A-mCherry cells with stable LmoCascade-KRAB expression. Cells were transduced with individual crRNAs, selected for 2 d with puromycin and harvested on day 7. **d**, Relative *HBE1* expression measured by qPCR. **e**, Relative mean fluorescence intensity (MFI) of mCherry measured by flow cytometry. **f**, Representative flow cytometry histogram. For **d–f**, two independent experiments were conducted with similar results. Tukey test, ** $P < 0.001$ compared to control crRNA. Three biological independent samples. Error bars, s.e.m. Numbers above bars indicate mean relative expression.

Cascade to target DNA³⁴ and tethering of transcriptional repressor domains such as Krüppel-associated box (KRAB)^{20,31,35} to repurpose LmoCascade for targeted transcriptional repression in mammalian cells. To achieve stable expression of LmoCascade-KRAB in human cells, we generated lentivirus expression constructs for each Cascade subunit, including the addition of a T2A-Blasticidin^R sequence downstream of Cas6-KRAB (Supplementary Fig. 8a). Following co-transduction of a K562-HBE1-mCherry endogenous gene reporter cell line³⁶, LmoCascade-KRAB-expressing cells were selected with blasticidin S, followed by clonal isolation and expansion. Clone no. 2 was selected following confirmation of LmoCascade-KRAB expression by western blot analysis (Supplementary Fig. 8b).

To test programmable endogenous gene repression in human cells, a panel of crRNAs was generated, tiling the endogenous

5'-untranslated region of *HBE1* (Fig. 5a). Lentiviral expression constructs were generated for stable, independent expression of a crRNA and an eGFP-2A-Puro^R selection cassette (Fig. 5b). Transduction, selection and expansion of K562-HBE1-mCherry cells expressing LmoCascade-KRAB (Fig. 5c) revealed *HBE1* transcriptional repression with all crRNAs (Fig. 5d). To further assess the repressive capacity of LmoCascade-KRAB, protein expression was evaluated by flow cytometry analysis of the HBE1-mCherry reporter (Fig. 5e,f). Significant reduction in mCherry fluorescence was observed for all crRNA targets ($P < 0.05$, Fig. 5e), including robust repression in five of six crRNAs ($P < 0.001$, Fig. 5e). These results demonstrate the potential for repurposing type I CRISPR systems as programmable transcriptional repressors in mammalian cells.

Discussion

In summary, these results show that Cascade from type I-E and type I-B CRISPR systems can be reprogrammed for RNA-guided transcriptional modulation in human cells. This new class of genome engineering tools has potential benefits that expand the CRISPR engineering toolbox. For example, the promiscuous PAM recognition of type I-E EcoCascade (5'-AAG, AGG, ATG, GAG, TAG-3'), and the additional PAM recognition of type I-B LmoCascade (5'-CCA-3'), located 5' of the spacer in contrast to the 3' PAM of type II systems^{16,37}, provides a larger set of available CRISPR target sequences.

By tiling crRNAs along endogenous promoters, our study revealed PAM-independent (Supplementary Table 1) variation in the transactivation potential of crRNAs (Figs. 2b,c,f,g, 4c,d). The reasons for this remain to be elucidated, but these differences in transactivation potential are similar to those observed among guide RNAs (gRNAs) and crRNAs for class 2 systems. Ineffective class 2 effectors can gain activity when co-targeting dCas9 to proximal sites³⁸, suggesting that dCas9 increases the accessibility of these sites. Similarly, we observed 16-fold enhanced transactivation with EcoCascade-p300 and cr22 when dCas9 was co-targeted to *IL1RN* (Supplementary Fig. 9). These results suggest a need for careful crRNA screening and selection for Cascade targeting to the mammalian genome; however, this selective targeting by particular crRNAs may also serve as a mechanism to increase targeting specificity.

Additionally, the preservation of complex formation observed after effector tethering suggests opportunities to use the stoichiometry of the Cascade complex to explore the synergistic activities of multiple effector domains. For dCas9, combinatorial targeting by tethering KRAB and DNA methyltransferases has been used to achieve heritable transcriptional silencing³⁹. Furthermore, the stoichiometry and architecture of Cascade have been tuned in bacteria by altering crRNA protospacer length^{25,40}. The several *cas* genes involved and their various corresponding Cas proteins also present individual opportunities to append molecules and functional domains with increased flexibility.

Beyond repurposing type I CRISPR systems for targeted transcriptional modulation, we also anticipate that Cascade subunits can be tethered to endonuclease effectors, such as the catalytic domains of the homodimeric FokI and monomeric I-TevI^{41–43} endonucleases, for programmable editing via generation of double-stranded breaks or single-stranded nicks in genomic DNA. In fact, we have observed indels characteristic of double-strand break repair following delivery of Cascade-I-TevI fusions in preliminary experiments (Supplementary Fig. 10). Alternatively, Cascade can be co-expressed with the Cas3 helicase-nuclease to generate a spectrum of long-range chromosomal deletions⁴⁴.

Targeted transcriptional modulation is important for perturbing gene function, designing gene regulatory networks, investigating the function of distal regulatory elements, manipulating cellular and organismal phenotypes, and inducing therapeutic changes to gene expression. Cascade complexes from type I CRISPR-Cas systems provide a novel and widely applicable RNA-guided platform for targeting DNA sequences in eukaryotes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41587-019-0235-7>.

Received: 25 April 2019; Accepted: 23 July 2019;

Published online: 23 September 2019

References

- Gaj, T., Gersbach, C. A. & Barbas, C. F. 3rd ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
- Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **34**, 933–941 (2016).
- Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
- Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- Cox, D. B. T. et al. RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
- Majumdar, S. et al. Three CRISPR-Cas immune effector complexes coexist in *Pyrococcus furiosus*. *RNA* **21**, 1147–1158 (2015).
- Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- Mulepati, S. & Bailey, S. *In vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* **288**, 22184–22192 (2013).
- Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- Jore, M. M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
- Wiedenheft, B. et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).
- Jackson, R. N. et al. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479 (2014).
- Zhao, H. et al. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 (2014).
- Hayes, R. P. et al. Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* **530**, 499–503 (2016).
- Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & Macmillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **18**, 688–692 (2011).
- Beloglazova, N. et al. CRISPR RNA binding and DNA target recognition by purified Cascade complexes from *Escherichia coli*. *Nucleic Acids Res.* **43**, 530–543 (2015).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
- Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
- Maeder, M. L. et al. CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
- 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).
- Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
- Luo, M. L., Mullis, A. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43**, 674–681 (2015).
- Westra, E. R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**, 595–605 (2012).
- Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E. & Doudna, J. A. DNA targeting by a minimal CRISPR RNA-guided Cascade. *Mol. Cell* **63**, 840–851 (2016).
- Leenay, R. T. et al. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* **62**, 137–147 (2016).
- Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).
- Thakore, P. I. et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
- Bolukbasi, M. F., Gupta, A. & Wolfe, S. A. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat. Methods* **13**, 41–50 (2016).
- Di, H. et al. Comparative analysis of CRISPR loci in different *Listeria monocytogenes* lineages. *Biochem. Biophys. Res. Commun.* **454**, 399–403 (2014).

34. Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P. R. & Lundgren, M. Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.* **43**, 237–246 (2015).
35. Margolin, J. F. et al. Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl Acad. Sci. USA* **91**, 4509–4513 (1994).
36. 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).
37. Sinkunas, T. et al. *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* **32**, 385–394 (2013).
38. Chen, F. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.* **8**, 14958 (2017).
39. Amabile, A. et al. Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell* **167**, 219–232.e214 (2016).
40. Gleditsch, D. et al. Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res.* **44**, 5872–5882 (2016).
41. Kleinstiver, B. P. et al. Monomeric site-specific nucleases for genome editing. *Proc. Natl Acad. Sci. USA* **109**, 8061–8066 (2012).
42. Beurdeley, M. et al. Compact designer TALENs for efficient genome engineering. *Nat. Commun.* **4**, 1762 (2013).
43. Kleinstiver, B. P. et al. The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3 (Bethesda)* **4**, 1155–1165 (2014).
44. Dolan, A. E. et al. Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using Type I CRISPR-Cas. *Mol Cell* **74**, 936–950 (2019).

Acknowledgements

We thank M. Oliver, I. Hilton and D. D. Kocak for technical assistance and helpful comments and D. Ousterout for valuable discussions. This work was supported by Locus Biosciences, an Allen Distinguished Investigator Award from the Paul G. Allen Frontiers Group, the Thorek Memorial Foundation, US National Institutes of Health (NIH) grant R01DA036865, an NIH Director's New Innovator

Award (DP2OD008586), a US National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award (CBET-1151035), NSF grant DMR-1709527, and an NSF Emerging Frontiers in Research and Innovation (EFRI) Award (EFMA-1830957). A.P.-O. was supported by a Pfizer-NCBiotech Distinguished Postdoctoral Fellowship. J.B.B. was supported by an NIH Biotechnology Training Grant (T32GM008555) and Predoctoral Fellowship (F31NS105419). R.B. was supported by internal funds from NC State University. C.L.B. was supported by an NIH Maximizing Investigator's Research Award (1R35GM119561-01).

Author Contributions

A.P.-O., J.B.B. and C.A.G. designed the experiments. A.P.-O., J.B.B., M.M.L., K.J.M., T.S.K., K.A.G., M.J.S. and L.C.B. performed the experiments. A.P.-O., J.B.B., C.E.N., A.B., T.E.R. and C.A.G. analyzed the data. C.L.B. and R.B. provided Cascade sequences. A.P.-O. and C.A.G. wrote the manuscript with input from all authors.

Competing interests

C.A.G., A.P.-O., R.B. and C.L.B. have filed patent applications related to genome engineering with type I CRISPR systems. T.S.K. is a co-founder of, and advisor to, Element Genomics. C.A.G. is a co-founder of, and advisor to, Locus Biosciences and Element Genomics, and an advisor to Sarepta Therapeutics. R.B. is a co-founder and Scientific Advisory Board member of Locus Biosciences and Intellia Therapeutics. C.L.B. is a co-founder and Scientific Advisory Board member of Locus Biosciences.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-019-0235-7>.

Correspondence and requests for materials should be addressed to C.A.G.

Reprints and permissions information is available at www.nature.com/reprints.

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. 2019

Methods

Mammalian expression plasmid construction. *E. coli* K-12 Cascade sequences were originally codon optimized by human codon usage tables using integrated DNA technology, synthesized as gene blocks and integrated into expression plasmids containing a CMV-driven cassette by Gibson cloning strategies. DNA2.0 (ATUM) synthesized the second-round human codon-optimized constructs using proprietary methods. See Supplementary Text for gene sequences of *E. coli* K-12 Cascade constructs. For crRNA expression, a cloning vector was constructed (pAPcrRNA_Eco) with a U6-driven cassette and digested with SacII and XhoI. To insert repeat-spacer pairs, oligonucleotides encoding the palindromic repeat and crRNA spacers were annealed, 5' phosphorylated with T4 polynucleotide kinase (NEB) and ligated into digested pAPcrRNA_Eco. See Supplementary Fig. 11 for an illustration of the cloning scheme. The control crRNA used throughout this study targets Cascade to an irrelevant control locus at *HBE1*. See Supplementary Table 3 for Cas9 gRNA target sequences²³.

L. monocytogenes Finland_1998 Cascade sequences were synthesized by DNA2.0 (ATUM) as human codon-optimized constructs using proprietary methods. See Supplementary Text for the gene sequences of *L. monocytogenes* Finland_1998 Cascade constructs. For crRNA expression, a cloning vector was constructed (pAPcrRNA_Lmo) with a U6-driven cassette and digested with SacII and AgeI. To insert repeat-spacer pairs, oligonucleotides encoding the palindromic repeat and crRNA spacers were annealed, 5' phosphorylated with T4 polynucleotide kinase (NEB) and ligated into digested pAPcrRNA_Lmo. See Supplementary Fig. 12 for an illustration of the cloning scheme. The control crRNA used throughout this study targets Cascade to an irrelevant control locus at *HBE1*. The plasmids used throughout this study are available from Addgene (plasmids 106270–106276, 126481–126494 and 126501).

Cell culture and transfections. HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (Invitrogen) with 10% FBS (Sigma-Aldrich) and 1% penicillin–streptomycin (Gibco). K562 cells were maintained in RPMI-1640 medium (Invitrogen) with 10% FBS (Sigma-Aldrich) and 1% penicillin–streptomycin (Gibco). Cells were incubated at 37°C with 5% CO₂. Lentivirus was produced in HEK293T cells using Lipofectamine 3000 (Invitrogen). All other transfections were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol.

Immunofluorescence staining. Cells were passaged and transfected with 100 ng plasmid DNA on coverslips in 24-well plates. At 3 d post-transfection, cells were washed with PBS and fixed with 4% paraformaldehyde (Sigma-Aldrich). Cells were incubated with blocking buffer (5% goat serum, 0.2% Triton X-100 in PBS) then incubated with mouse anti-Flag (1:200 dilution, Sigma-Aldrich, M2 clone), followed by incubation with goat anti-mouse Alexa Fluor 647 (1:200 dilution, Life Technologies, A21236), and DAPI nucleic acid stain (Invitrogen). Cells were imaged with a Leica DMI 3000 B fluorescence microscope. Exposure time was set by fluorescence of the lowest expressed construct and maintained for all samples.

Western blot and co-immunoprecipitation. For protein analysis, HEK293T cells were transfected with 2 µg of individual *Cas* constructs in 6-well plates. After 3 d, cells were lysed in RIPA buffer (Sigma-Aldrich) with a proteinase inhibitor cocktail (Roche). Samples were centrifuged at 12,000 r.p.m. for 5 min and the supernatant was isolated and quantified using a bicinchoninic acid assay (BCA) protein standard curve (Thermo Fisher Scientific) on a BioTek Synergy 2 Multi-Mode microplate reader. Mixed with NuPAGE loading buffer (Invitrogen) and 5% β-mercaptoethanol, 25 µg protein was heated at 100°C for 10 min. Samples were loaded into 10% NuPAGE Bis–Tris gels (Invitrogen) with MES buffer (Invitrogen) and electrophoresed for 70 min at 200 V on ice. Protein was transferred to nitrocellulose membranes (Bio-Rad) for 1 h in 1 × tris–glycine transfer buffer containing 10% methanol and 0.01% SDS at 4°C at 400 mA. The blot was blocked at room temperature (23°C) for 30 min in 5% milk–TBST (50 mM Tris, 150 mM NaCl and 0.1% Tween-20) and incubated with mouse anti-Flag (1:1,000 dilution, Sigma-Aldrich, M2 clone) in 5% milk–TBST at 4°C overnight. Blots were then washed in TBST and incubated with goat anti-mouse-conjugated horseradish peroxidase (HRP) (1:2,500 dilution, Sigma-Aldrich) in 5% milk–TBST for 45 min at room temperature. Blots were washed in TBST then visualized using Western-C ECL substrate (Bio-Rad) on a ChemiDoc XRS+ system (Bio-Rad). Blots were stripped with Restore PLUS western blot stripping buffer (Thermo Fisher Scientific), blocked, and re-blotted with rabbit anti-GAPDH (1:1,000 dilution, Cell Signaling, 14C10) or anti-actin (1:1,000 dilution, Sigma-Aldrich, A2066) and goat anti-rabbit-conjugated HRP (1:2,500 dilution, Sigma-Aldrich). Blots were visualized again using the methods described above.

For co-immunoprecipitation analysis, co-transfections were completed using a V5-Cas7 construct. HEK293T cells were co-transfected in 6-well plates with 425 ng of each *Cas* construct and crRNA for 2.25 µg total plasmid DNA per condition. At 3 d post-transfection, cells were lysed with IP lysis buffer (Thermo Fisher Scientific) with a proteinase inhibitor cocktail (Roche). Samples were centrifuged at 12,000 r.p.m. for 5 min and the supernatant was isolated and subjected to immunoprecipitation using goat anti-V5-agarose conjugate (10 µl, Abcam, ab1229) at 4°C overnight. The immunoprecipitation products were washed three times

with IP lysis buffer, mixed with NuPAGE loading buffer and 5% β-mercaptoethanol and heated at 100°C for 10 min. Samples were loaded into 10% NuPAGE Bis–Tris gels, and resolved as described above. Blots were blocked, incubated with mouse anti-Flag (1:1,000 dilution, Sigma-Aldrich, M2 clone) and mouse anti-V5 (1:40,000 dilution, Abcam, SV5-Pk1 clone), then with goat anti-mouse-conjugated HRP (1:2,500 dilution, Sigma-Aldrich). Blots were visualized as described above.

RNA analysis. For qPCR, HEK293T cells were co-transfected with individual crRNAs (100 ng) and EcoCascade (50 ng Cas8e, 100 ng Cse2, 50 ng Cas7, 250 ng Cas5 and 50 ng Cas6–p300) or LmoCascade (150 ng Cas8b2, 50 ng Cas7, 75 ng Cas5 and 150 ng Cas6–p300) in 24-well plates. After 3 d, total RNA was isolated using QIAshredder and QIAGEN RNeasy kits (Qiagen). Reverse transcription was carried out using 500 ng total RNA per sample in a 10 µl reaction using the SuperScript VILO Reverse Transcription Kit (Invitrogen). Per qPCR reaction, 1.0 µl of cDNA was used with Perfecta SYBR Green Fastmix (Quanta Biosciences) and run using the CFX96 Real-Time PCR Detection System (Bio-Rad). All sequences for qPCR primers can be found in Supplementary Table 3. All qPCR data are presented as fold change in RNA normalized to *Gapdh* expression and relative to samples targeting Cascade with a crRNA targeted to an irrelevant control locus at *HBE1*.

RNA sequencing. HEK293T cells were co-transfected with 3 µg total plasmid in 6-well plates. After 3 d, cells were washed twice with PBS and 350 µl of a 1:10 mixture of β-mercaptoethanol and Buffer RLT (Qiagen) was added to each well. While on ice, cells were lysed, nucleic acid was quantified using a Nanodrop instrument (Thermo Fisher Scientific) and RNA quality was assessed using an Agilent TapeStation 2200 with RNA ScreenTape (Agilent). Using 1 µg of total RNA input, stranded mRNA sample preparation was performed with the Illumina TruSeq Stranded mRNA Library Prep Kit (Illumina) following the manufacturer's protocol except that the enzymatic fragmentation time was reduced from 8 min to 1 min. Libraries were sequenced at the Duke GCB Sequencing Core as 51 cycles of paired-end runs (51PE), in an Illumina HiSeq 4000 platform. Reads were aligned against the human reference genome GRCh38 using the aligner STAR v2.4.1a⁴⁵ following the proposed two-pass strategy to first identify a splice junction database to improve the overall mapping quality. Gene counts were estimated with featureCounts from the subread package v1.4.6-p6⁴⁶, using gene annotations from Refseq⁴⁷ and allowing for multimapping reads (parameters -M and -fraction). Differential expression analyses were performed using the DESeq2 package⁴⁸ filtering out non-expressed genes and fitting a negative binomial generalized linear model to find significantly differentially expressed genes.

Chromatin immunoprecipitation qPCR. HEK293T cells were transfected with 40 µg total plasmid in 15 cm dishes. After 3 d, cells were fixed in 1% formaldehyde for 10 min at room temperature. The reaction was quenched with 0.125 M glycine and the cells were lysed using Farnham lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl and 0.5% NP-40) with a protease inhibitor cocktail (Roche). Nuclei were collected by centrifugation at 2,000 r.p.m. for 5 min at 4°C and lysed in RIPA buffer with a protease inhibitor cocktail (Roche). Chromatin was sonicated using a Biorupter Sonicator (Diagenode, model XL) and immunoprecipitated using an anti-Flag antibody (Sigma-Aldrich, M2 clone). The formaldehyde crosslinks were reversed by heating overnight at 65°C and genomic DNA fragments were purified using a spin column. For qPCR, 500 pg of immunoprecipitated DNA was used per reaction. qPCR was performed as described above. The data are presented as fold change gDNA normalized to a region of the β-actin locus and relative to samples targeting Cascade with the control crRNA mentioned above. All sequences for qPCR primers can be found in Supplementary Table 4.

Chromatin immunoprecipitation sequencing. Reverse-crosslinked immunoprecipitated DNA was cleaned using PCR purification columns (Qiagen). DNA concentration was determined using a Qubit dsDNA High Sensitivity and Broad Range assay kit (Invitrogen). To prepare sequencing libraries for Illumina sequencing, 7 ng input of immunoprecipitated DNA was used with the Hyper Prep kit (Kapa Biosystems). After library preparation, samples were barcoded with Illumina TruSeq indexes and normalized to 10 nM. Final libraries were pooled and run on a HiSeq 4000 to generate approximately 20 million, 50-bp single-end reads per sample.

Sequences for TruSeq Illumina adapters were removed from the raw reads using Trimmomatic v0.32⁴⁹. Then, reads were aligned using Bowtie v1.0.0⁵⁰, reporting the best alignment with up to two mismatches (parameters --best --strata -v 2). Duplicates were removed using Picard MarkDuplicates v1.130, and low mappability or blacklisted regions identified by the ENCODE project⁵¹ were filtered out from the final BAM files. Using the sequenced input controls, binding regions were identified using the callpeak function in MACS2 v2.1.0.20151222⁵². For the differential binding analysis, first, a union peakset was computed by merging individual peak calls using bedtools2 v2.25.0⁵³. Then, reads in peaks were estimated using featureCounts and the difference in binding was assessed with DESeq2.

For the genomic window of 562 bp on chromosome 6 near the *TAAR4P* pseudogene displaying significant differential binding for cr25 compared with control crRNA, we searched for both global and local alignments of the 32-bp

cr26 sequence. Using the water program in the EMBOSS v6.6.0 package⁵⁴, which implements the Smith–Waterman algorithm for local alignment, we were able to map only 8 nucleotides (26.6% of the cr26 sequence). When looking for end-to-end alignments with the needle program implementing the Needleman–Wunsch algorithm, the best alignment contained 5 nucleotides on the 5′ end of the cr26 protospacer sequence and 16 mismatches (50% sequence identity). By contrast, when aligning cr26 to the *IL1RN* gene, a nearly perfect continuous match (30 of 32 bp, 94% sequence identity) starting at the 5′ end occurred in the genomic location chr2:113117890–113117919. This difference suggests an alternative mode of binding for this potential off-target site.

Lentiviral transduction. K562–HBE1–mCherry reporter cells³⁶ were co-transduced with lentivirus expressing LmoCascade subunits and Cas6–KRAB–2A–BlastR. Transduced cells were selected with blasticidin S (Invitrogen) at a concentration of 10 µg ml^{−1}, then clonally isolated and expanded. For protein analysis, cells were lysed in RIPA buffer (Sigma–Aldrich) with a proteinase inhibitor cocktail (Roche) and a western blot was completed using 25 µg protein and mouse anti-Flag antibody (1:1,000 dilution, Sigma, M2 clone) using the methods described above.

LmoCascade–KRAB clone-two cells were transduced with lentivirus expressing individual crRNAs and selected with puromycin (Sigma–Aldrich) at a concentration of 1 µg ml^{−1}. After 7 d, cells were collected, washed twice with 2 mM EDTA (Thermo Fisher Scientific) and 0.5% BSA (Sigma–Aldrich) in PBS. Flow cytometry was done with a SH800 Cell Sorter (Sony Biotechnology). Total RNA was isolated and used for qPCR analysis using the methods described above. The qPCR data are presented as fold change in RNA normalized to *Gapdh* expression and relative to samples targeting Cascade with a crRNA targeted to an irrelevant control locus at *IL1RN*.

Deep sequencing and indel analysis. HEK293T cells were co-transfected with 600 ng total plasmid in 24-well plates. After 4 d, genomic DNA was isolated using QIAGEN DNeasy Blood and Tissue kit (Qiagen). In a first-round PCR, 100 ng of genomic DNA was amplified with genome-specific primers. A second round of PCR was used to add experimental barcodes and Illumina flow cell binding sequences. The resulting sequence libraries were diluted to 2 nM, pooled and sequenced with 150-bp paired-end reads on an Illumina MiSeq instrument. Samples were demultiplexed and analyzed for insertions and deletions with a local distribution of CRISPResso⁵⁵ with default parameters and a 30-bp window around the predicted I–TeVI 5′–CNNNG–3′ cut sites. All primers for genomic DNA amplification can be found in Supplementary Table 5.

Statistical analysis. All data analyzed had two to three biological replicates and are presented as mean ± s.e.m. Logarithmic transformations were completed before statistical analysis where indicated. All *P* values were calculated by global one-way analysis of variance (ANOVA) with Tukey post hoc tests ($\alpha = 0.05$).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus^{56,57} and are accessible through GEO Series Accession number [GSE114859](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114859). All other relevant raw data are available from the corresponding author upon request.

Code availability

Custom scripts used for ChIP–seq and RNA–seq analysis are available upon request.

References

- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
- O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
- Pinello, L. et al. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).
- Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
- Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

► Experimental design

1. Sample size

Describe how sample size was determined.

Sample size was determined by the minimum required samples that permitted statistical analysis for in vitro analysis.

2. Data exclusions

Describe any data exclusions.

Data were not excluded from the analyses.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Samples were not randomized as the data did not require randomization.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Samples were not blinded as the data did not require blinding.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ A statement indicating how many times each experiment was replicated
- ☐ ☒ 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 any assumptions or corrections, such as an adjustment for multiple comparisons
- ☐ ☒ Test values indicating whether an effect is present
Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Microsoft Excel v365 ProPlus, JASP v0.9.2.0, and GraphPad PRISM v8 were used to analyze the data in this study. For RNA sequencing analysis, STAR v2.4.1a was used for alignment, gene counts were estimated with featureCounts from the subread package v1.4.6-p6, and differential expression analyses were performed using the DESeq2 package. For ChIP-seq, Trimmomatic v0.32 was used for removing TruSeq Illumina adapters, reads were aligned using Bowtie v1.0.0, duplicates were removed using Picard MarkDuplicates v1.130, binding regions were identified using the callpeak function in MACS2 v2.1.0.20151222, a union peakset was computed merging individual peak calls using bedtools2 v2.25.0, reads in peaks were estimated using featureCounts, the difference in binding was assessed with DESeq2, and global and local alignments were searching using the water program in the EMBOSS v6.6.0 package. For indel analysis, samples were demultiplexed and analyzed with a local distribution of CRISPResso v1.0.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

Plasmids are available through Addgene. All other materials are available upon reasonable request to the corresponding authors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

For HEK293T cell studies: anti-Flag (Sigma, F3165, M2 clone, 1:1000, Lot#SLBT6752), anti-GAPDH (Cell Signaling, clone 14C10, Rabbit mAB #2118, 1:1000, Lot#10), anti-Actin (Sigma, A2066, 1:1000), anti-Alexa Fluor 647 (Invitrogen, A-21236, goat anti-Mouse IgG, 1:200, Lot#1740028), anti-V5-agarose conjugate (Abcam, ab1229, 10ul per IP, Lot#GR3186941-4), anti-V5 (Abcam, ab27671, SV5-Pk1 clone, 1:40,000, Lot#GR242588-15), goat anti-mouse IgG-HRP (Santa Cruz, A8924, 1:2500), goat anti-rabbit-IgG-HRP (Santa Cruz, A6154, 1:2500).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

HEK293T cells and K562 cells were obtained from the American Tissue Collection Center (ATCC) via the Duke University Cancer Center Facilities.

b. Describe the method of cell line authentication used.

Cell lines were validated via morphological inspection.

c. Report whether the cell lines were tested for mycoplasma contamination.

Cell lines were not tested for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly misidentified cells were used. All cells displayed homogeneous characteristic morphology.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No research animals were used.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human research participants were used.