



# HHS Public Access

Author manuscript

*Nat Biotechnol.* Author manuscript; available in PMC 2020 March 23.

Published in final edited form as:

*Nat Biotechnol.* 2019 December ; 37(12): 1493–1501. doi:10.1038/s41587-019-0235-7.

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

**Adrian Pickar-Oliver<sup>1,2</sup>, Joshua B. Black<sup>1,2</sup>, Mae M. Lewis<sup>1,2</sup>, Kevin J. Mutchnick<sup>1,2</sup>, Tyler S. Klann<sup>1,2</sup>, Kylie A. Gilcrest<sup>1,2</sup>, Madeleine J. Sitton<sup>1,2</sup>, Christopher E. Nelson<sup>1,2</sup>, Alejandro Barrera<sup>3</sup>, Luke C. Bartelt<sup>2,3</sup>, Timothy E. Reddy<sup>2,3,4</sup>, Chase L. Beisel<sup>5,6,7</sup>, Rodolphe Barrangou<sup>8</sup>, Charles A. Gersbach<sup>1,2,9,10,\*</sup>**

<sup>1</sup>Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA.

<sup>2</sup>Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA.

<sup>3</sup>Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA.

<sup>4</sup>University Program in Genetics and Genomics, Duke University, Durham, NC 27708, USA.

<sup>5</sup>Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695, USA.

<sup>6</sup>Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Centre for Infection Research (HZI), 97080 Würzburg, Germany.

<sup>7</sup>Medical Faculty, University of Würzburg, 97080 Würzburg, Germany.

<sup>8</sup>Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC 27695, USA.

<sup>9</sup>Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC 27710, USA.

<sup>10</sup>Lead Contact

### Abstract

Class 2 CRISPR-Cas systems, such as Cas9 and Cas12, have been widely applied for targeting DNA sequences in eukaryotic genomes. However, class I 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

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Correspondence to: Charles A. Gersbach, Ph.D., Department of Biomedical Engineering, Room 1427 FCIEMAS, 101 Science Drive, Box 90281, Duke University, Durham, NC 27708-0281, Phone: 919-613-2147, Fax: 919-668-0795, [charles.gersbach@duke.edu](mailto:charles.gersbach@duke.edu).

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 Financial Interests Statement

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. C.A.G. is a co-founder and advisor to Locus Biosciences and Element Genomics, and an advisor to Sarepta Therapeutics. R.B. is a co-founder and SAB member of Locus Biosciences and Intellia Therapeutics. C.L.B. is a co-founder and SAB member of Locus Biosciences.

cells and used for DNA-targeting and transcriptional control. We repurpose type I variants of class 1 CRISPR-Cas systems from *E. coli* and *L. 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 I CRISPR-Cas systems for further exploration.

## Introduction

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 tremendous advances in science, biotechnology, and medicine<sup>1</sup>. 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 with a vast range of applications<sup>2</sup>. In fact, bioinformatics analyses have further revealed a diversity of CRISPR-Cas systems, and the most recent classification encompasses six major types (I through VI)<sup>3</sup>. However, single-effector class 2 systems (types II, V and VI) have been primarily 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 archaea<sup>4</sup>. The 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 V<sup>5</sup> and Cas13-based type VI<sup>6</sup> CRISPR-Cas systems of class 2 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) based on related, but subtype-specific, accessory *cas* genes<sup>3,4,7</sup>. The well-studied prototypical type I-E system of *Escherichia coli* K12 consists of eight *cas* genes and a downstream CRISPR array<sup>8-10</sup>. Following transcription of the CRISPR array, the 29-bp repeat sequences flanking the variable spacer sequences are cleaved during crRNA biogenesis by the Cas6 endoribonuclease<sup>10</sup>. Together, five protein subunits (Cas8e, Cse2, Cas7, Cas5, and Cas6, previously referred to as CasA, CasB, CasC, CasD, and CasE, respectively)<sup>4</sup> and the mature 61-nt crRNA form the 405-kDa CRISPR-associated complex for antiviral defense (Cascade)<sup>11-13</sup> (Figure 1A). Unlike type II CRISPR systems, there is no tracrRNA required for effector complex formation. For type I CRISPR systems, CasE 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 (Figure 1B). We sought to harness this prokaryotic immune system for genome targeting in eukaryotes.

## Results (with Subheadings)

### Expressing the Cascade complex in human cells

To repurpose Cascade for use in mammalian cells, we used a CMV promoter to express each Cascade subunit of the *E. coli* K12 system (EcoCascade). Based on available structural information<sup>13–16</sup>, 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 CasE (Figure 1C)<sup>17</sup>. Heterologous expression of all EcoCascade constructs was confirmed by Western blot following transfection of each plasmid individually into human HEK293T cells (Figure 1D). The *cas* genes were originally optimized based on human codon usage, but variable expression of these *cas* constructs indicated a need for additional codon-optimization of *Cas5*, which was performed by ATUM/DNA2.0 (DNA sequences provided in Supplementary Information). To determine if EcoCascade complex formation occurred in human cells, the six plasmids encoding each of the five Cas subunits and the crRNA cassette were co-transfected into HEK293T cells. Co-immunoprecipitation by pull-down of the V5 epitope on Cas7 and blotting for the Flag epitope on the other subunits confirmed proper complex assembly (Figure 1E). Interestingly, EcoCascade complexes can be purified from bacteria in the absence of a crRNA<sup>18</sup>, however we observed EcoCascade formation in human cells only in the presence of a crRNA (Figure 1E). Additionally, nuclear localization of each subunit with the single NLS was confirmed by immunofluorescence staining of transfected HEK293T cells (Figure 1F).

### Programmable transcriptional activation by EcoCascade-p300

Next, we sought to repurpose EcoCascade for CRISPR-based programmable transcriptional activation (CRISPRa) 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 DNA<sup>19</sup>. For example, the nuclease-deficient Cas9 (dCas9) can be engineered to function as a synthetic transcriptional activator by genetically fusing it to a transactivation or epigenome-modifying domain<sup>20–24</sup>. 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* was utilized for a CRISPR interference (CRISPRi) strategy in bacteria by permitting Cascade to bind target DNA and block transcription without DNA degradation<sup>25</sup>. Therefore, we hypothesized 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 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<sup>23</sup>. The EcoCascade system contains five Cas

subunits available at various stoichiometry (Figure 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 (Figure 2A). The ability of each of these subunits to accommodate the fusion of the large p300 core catalytic domain (72kDa) 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 *ILIRN* promoter at protospacer targets downstream of known PAMs (5'-AAG, AGG, ATG, GAG, TAG-3')<sup>10,26-28</sup> (Figure 2B, Table S1). Co-transfection of HEK293T cells with plasmids encoding EcoCascade with Cas6-p300 and individual crRNAs revealed robust *ILIRN* activation with many crRNAs, including >3,000-fold *ILIRN* activation with cr26 (\*\* $P < 0.001$ , Figure 2C). Importantly, cr26 with EcoCascade lacking a p300 domain, or cr26 alone did not activate *ILIRN* (Figure S1), suggesting target-specific activation by EcoCascade-p300. Based on EcoCascade stoichiometry (Figure 1A) and relative Cas construct expression (Figure 1D), we optimized the relative masses of transfected plasmids to maximize gene activation (Figure S2). 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 displayed significant transactivation of *ILIRN* (\* $P < 0.05$ , Figure 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 (Figure 2B). We observed significant enrichment of the target regions in EcoCascade-p300 samples co-transfected with cr25 or cr26 compared to a control crRNA (\*\* $P < 0.001$ , Figure 2E). These results further confirm EcoCascade as a programmable DNA-binding platform for efficient targeting of specific loci in the human genome. Intriguingly, we observed enhanced enrichment of *ILIRN* signal in samples treated with Flag-tagged EcoCascade-p300 and an *ILIRN*-targeted crRNA compared to samples treated with Flag-tagged dCas9-p300 and an *ILIRN*-targeted single-guide RNA (sgRNA) (Figure S3). These results may indicate increased occupancy by EcoCascade relative to dCas9 but could also be the result of differences in epitope avidity or presentation. Targeted endogenous *ILIRN* activation was also achieved by tethering Cas6 to the tripartite activator, VP64-p65-Rta (VPR)<sup>24</sup>, although both p300 and VPR tethered to Cas6 led to reduced activation levels compared to fusion to dCas9 (Figure S4). To assess activation of other endogenous targets in the human genome, we targeted the *HBG* promoter with EcoCascade-p300 (Figure 2F) and observed robust transactivation (Figure 2G). The natural function of Cascade to process crRNAs suggests the possibility of using arrayed spacers for multiplexed genome engineering. By generating a CRISPR array containing multiple crRNA spacers that target both *ILIRN* and *HBG*, we demonstrated multiplexed activation of endogenous genes (Figure S5). 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

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<sup>29–31</sup>, which has led to the development of a variety of strategies to improve specificity<sup>32</sup>. To quantify the genome-wide binding specificity of EcoCascade-p300, we performed 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 (Figure 3A). Interestingly, the strength of binding signal was comparable between cr25 and cr26, even though they differed substantially in their induction of *IL1RN* transcription (Figure 2C). With a genome-wide false discovery rate (FDR) < 0.001, there were a few off-target differential binding sites observed when comparing cr25 and cr26 to Ctrl crRNA (Figure 3B–C). These were all substantially weaker sites compared to the signal at the *IL1RN* locus, with the exception of one genomic window located on chromosome 6 about 5 kilobases adjacent to *TAAR4P* that was significantly enriched only with cr26 (Figure 3C and Figure S6A). 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, indicating a possible off-target site of this crRNA (Figure 3B–C).

To evaluate the specificity of crRNA-mediated endogenous gene activation with EcoCascade-p300, we performed RNA-seq to quantify transcriptome-wide changes when targeting *IL1RN* with cr26 or with a *GFP*-targeting control crRNA (Figure 3D). Targeted activation was highly specific to the target gene when EcoCascade-p300 was co-expressed with cr26, with detection of a modest change to only six other genes (Figure 3D). However, we did observe that the addition of the p300 domain to EcoCascade resulted in significant off-target transcriptional changes when compared to EcoCascade alone (Figure S6B). Given the highly specific DNA-targeting by EcoCascade (Figure 3B–C) and cr26-dependent activation of *IL1RN* (Figure 3D), these results may 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 repurposing the type I-B CRISPR-Cas system of *L. monocytogenes* Finland\_1998 (LmoCascade) (Figure 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 (Figure 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 the predicted spacer length of 36 nucleotides, was generated tiling the endogenous *ILIRN* promoter at protospacer targets downstream of the known PAM (5'-CCA-3')<sup>33</sup> (Figure 4C). Co-transfection of HEK293T cells with plasmids encoding LmoCascade with Cas6-p300 and individual crRNAs revealed robust *ILIRN* activation with most crRNAs (\*\* $P < 0.001$ , Figure 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 displayed significant transactivation of *ILIRN* (\*\* $P < 0.001$ , Figure 4E). In contrast to the panel of *ILIRN*-targeting crRNAs for EcoCascade (Figure 2C), almost all of the LmoCascade crRNAs, as well as three of the four Cas-p300 effector fusions, achieved significant *ILIRN* activation (Figure 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 (Figure S7). 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 Cascade to target DNA<sup>34</sup>, and tethering of transcriptional repressor domains such as KRAB<sup>20,31,35</sup>, to repurpose LmoCascade for targeted transcriptional repression in eukaryotic cells. To achieve stable expression of LmoCascade-KRAB in mammalian cells, we generated lentivirus expression constructs for each Cascade subunit, including the addition of a T2A-Blasticidin<sup>R</sup> sequence downstream of Cas6-KRAB (Figure S8A). Following co-transduction of a K562-HBE1-mCherry endogenous gene reporter cell line<sup>36</sup>, LmoCascade-KRAB expressing cells were selected with blasticidin S, followed by clonal isolation and expansion. Clone 2 was selected following confirmation of LmoCascade-KRAB expression by Western blot analysis (Figure S8B).

To test programmable endogenous gene repression in human cells, a panel of crRNAs was generated tiling the endogenous 5'-untranslated region of *HBE1* (Figure 5A). Lentiviral expression constructs were generated for stable, independent expression of a crRNA and an eGFP-2A-Puro<sup>R</sup> selection cassette (Figure 5B). Transduction, selection, and expansion of K562-HBE1-mCherry cells expressing LmoCascade-KRAB (Figure 5C) revealed *HBE1* transcriptional repression with all crRNAs (Figure 5D). To further assess the repressive capacity of LmoCascade-KRAB, protein expression was evaluated by flow cytometry analysis of the HBE1-mCherry reporter (Figure 5E–F). Significant reduction in mCherry fluorescence was observed for all crRNA targets ( $P < 0.05$ , Figure 5E), including robust repression in 5 of 6 crRNAs (\*\* $P < 0.001$ , Figure 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<sup>16,37</sup>, enables a larger set of available CRISPR target sequences.

By tiling crRNAs along endogenous promoters, our studies revealed PAM-independent (Table S1) variation in transactivation potential of crRNAs (Figures 2B–C, 2F–G, 4C–D). The reason for these differences remains to be elucidated but is similar to differences observed among gRNAs and crRNAs for class 2 systems. Ineffective class 2 effectors can gain activity when co-targeting dCas9 to proximal sites<sup>38</sup>, suggesting dCas9 increases accessibility of these sites. Similarly, we observed 16-fold enhanced transactivation with EcoCascade-p300 and cr22 when dCas9 was co-targeted to *ILIRN* (Figure S9). These results suggest a narrower ability for Cascade to recognize genomic DNA targets in eukaryotes, however this may also serve as a mechanism to increase targeting specificity.

Additionally, the preservation of complex formation observed after effector tethering suggests opportunities to utilize the stoichiometry of the Cascade complex for exploring synergistic activities of multiple effector domains. For dCas9, combinatorial targeting by tethering KRAB and DNA methyltransferases has been used to achieve heritable transcriptional silencing<sup>39</sup>. Additionally, the stoichiometry and architecture of Cascade have been tuned in bacteria by altering crRNA protospacer length<sup>25,40</sup>. 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<sup>41–43</sup> 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 (Figure S10). Alternatively, Cascade can be co-expressed with the Cas3 helicase-nuclease to generate a spectrum of long-range chromosomal deletions<sup>44</sup>.

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 Methods

### Mammalian expression plasmid construction

*E. coli* K-12 Cascade sequences were originally codon-optimized by human codon usage tables using Integrated DNA technology (IDT), synthesized as gene blocks and integrated into expression plasmids containing a CMV-driven cassette by Gibson cloning strategies. ATUM/DNA2.0 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 PNK and ligated into digested pAPcrRNA\_Eco. See Supplementary Figure S11 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 S3 for Cas9 gRNA target sequences<sup>23</sup>.

*L. monocytogenes* Finland\_1998 Cascade sequences were synthesized by ATUM/DNA2.0 as human codon-optimized constructs using proprietary methods. See Supplementary Text for 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 PNK and ligated into digested pAPcrRNA\_Lmo. See Supplementary Figure S12 for an illustration of the cloning scheme. The control crRNA used throughout this study targets Cascade to an irrelevant control locus at *HBE1*. Plasmids used throughout this study are available through Addgene (Plasmid IDs: 106270–106276, 126481–126494, 126501).

### Cell culture and transfections

HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (Invitrogen) with 10% Fetal Bovine Serum (Sigma) and 1% penicillin-streptomycin (Gibco). K562 cells were maintained in RPMI-1640 Medium (Invitrogen) with 10% Fetal Bovine Serum (Sigma) and 1% penicillin-streptomycin (Gibco). Cells were incubated at 37°C with 5% CO<sub>2</sub>. Lentivirus was produced in HEK293Ts 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 three days post-transfection, cells were washed with PBS and fixed with 4% paraformaldehyde (Sigma). 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, 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 set by fluorescence of 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 three days, cells were lysed in RIPA buffer (Sigma) with a proteinase inhibitor cocktail (Roche). Samples were centrifuged at 12,000rpm for 5min and the supernatant was isolated and quantified using a bicinchronic acid assay (BCA) protein standard curve (Thermo Scientific) on the 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 200V on ice. Protein was transferred to nitrocellulose membranes for 1 hour in 1X tris-glycine transfer buffer containing 10% methanol and 0.01% SDS at 4°C at 400 mA. The blot was blocked at room temperature 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, 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 (1:2,500 dilution, Sigma) 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 Scientific), blocked, and re-blotted with rabbit anti-GAPDH (1:1,000 dilution, Cell Signaling, 14C10) or anti-Actin (1:1,000 dilution, Sigma, A2066) and goat anti-rabbit-conjugated horseradish peroxidase (1:2,500 dilution, Sigma). 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 three days post transfection, cells were lysed with IP lysis buffer (Thermo Scientific) with a proteinase inhibitor cocktail (Roche). Samples were centrifuged at 12,000 rpm 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 IP 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, M2 clone) and mouse anti-V5 (1:40,000 dilution, Abcam, SV5-Pk1 clone) then with goat anti-mouse-conjugated horseradish peroxidase (1:2,500 dilution, Sigma). Blots were visualized as described above.

## RNA analysis

For quantitative PCR (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 three days, 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 ran using the CFX96 Real-Time PCR Detection System

(Bio-Rad). All sequences for qPCR primers can be found in Supplementary Table S3. 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 ug total plasmid in 6-well plates. After three days, cells were washed twice with PBS and 350 uL of 1:10 mixture of  $\beta$ -mercaptoethanol and Buffer RLT (Qiagen) was added to each well. While on ice, cells were lysed NA was quantified using a Nanodrop instrument, and RNA quality was assessed using an Agilent TapeStation 2200 with RNA ScreenTape (Agilent). Using 1 ug of total RNA input, stranded mRNA sample preparation was performed with the Illumina TruSeq Stranded mRNA Library Prep Kit (Illumina) following manufacturer's protocol except the enzymatic fragmentation time was reduced from 8 min to 1 min. Libraries were sequenced at the Duke GCB Sequencing Core as 51 cycles paired-end runs (51PE), in an Illumina HiSeq 4000 platform. Reads were aligned against the human reference genome GRCh38 using the aligner STAR v2.4.1a<sup>45</sup> following the proposed 2-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-p6<sup>46</sup>, using gene annotations from Refseq<sup>47</sup> and allowing for multimapping reads (-M and --fraction arguments). Differential expression analyses were performed using the DESeq2 package<sup>48</sup> 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  $\mu$ g total plasmid in 15cm dishes. After three days, cells were fixed in 1% formaldehyde for 10 min at room temperature. The reaction was quenched with 0.125M glycine and the cells were lysed using Farnham lysis buffer with a protease inhibitor cocktail (Roche). Nuclei were collected by centrifugation at 2,000rpm for 5min 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 anti-Flag (Sigma, M2). 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 ChIP'd 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  $\beta$ -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 S4.

### Chromatin immunoprecipitation sequencing

Reverse crosslinked ChIP'd DNA was cleaned using PCR purification columns (Qiagen). DNA concentration was determined using Qubit dsDNA High Sensitivity and Broad Range assay kit (Invitrogen). To prepare sequencing libraries for Illumina sequencing, 7 ng input of ChIP'd 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 ~20 million, 50 bp single-end reads per sample.

Sequences for TruSeq Illumina adapters were removed from the raw reads using Trimmomatic v0.32<sup>49</sup>. Then, reads were aligned using Bowtie v1.0.0<sup>50</sup>, reporting the best alignment with up to 2 mismatches (parameters --best --strata -v 2). Duplicates were removed using Picard MarkDuplicates v1.130, while low mappability or blacklisted regions identified by the ENCODE project<sup>51</sup> were filtered out from the final BAM files. Using the sequenced input controls, binding regions were identified using the callpeak function in MACS2 v2.1.0.20151222<sup>52</sup>. For the differential binding analysis, first, a union peakset was computed merging individual peak calls using bedtools2 v2.25.0<sup>53</sup>. 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 Ctrl 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 package<sup>54</sup> which implements the Smith-Waterman algorithm for local alignment, we were only able to map 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 *ILIRN* gene, a nearly perfect continuous match (30/32 bp, 94% sequence identity) starting at the 5' end occurs 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 cells<sup>36</sup> 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 then clonally isolated and expanded. For protein analysis, cells were lysed in RIPA buffer (Sigma) with a proteinase inhibitor cocktail (Roch) and a western blot was completed using 25 µg protein and mouse anti-Flag (1:1,000 dilution, Sigma, M2 clone) using the methods described above.

LmoCascade-KRAB clone 2 cells were transduced with lentivirus expressing individual crRNAs and selected with puromycin (Sigma) at a concentration of 1 µg/ml. After seven days, cells were harvested, washed twice with 2 mM EDTA (ThermoFisher) and 0.5% bovine serum albumin (Sigma) in PBS. Flow cytometry was done with a SH800 (Sony Biotechnology). Total RNA was isolated and used for qPCR analysis using the methods described above. The qPCR data is 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 *ILIRN*.

## Deep sequencing and indel analysis

HEK293T cells were co-transfected with 600 ng total plasmid in 24-well plates. After four days, 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 CRISPResso<sup>55</sup> 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 S5.

## Statistical analysis

All data analyzed with two to three biological replicates and presented as mean  $\pm$  SEM. Logarithmic transformations were completed prior to statistical analysis where indicated. All p values calculated by global one-way 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<sup>56,57</sup> and are accessible through GEO Series Accession number GSE114859. All sequencing data used in this study has been deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) database (SUB5479090). 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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 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, a NIH Director's New Innovator Award (DP2OD008586), and US National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award (CBET-1151035). 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).

## References (for main text only)

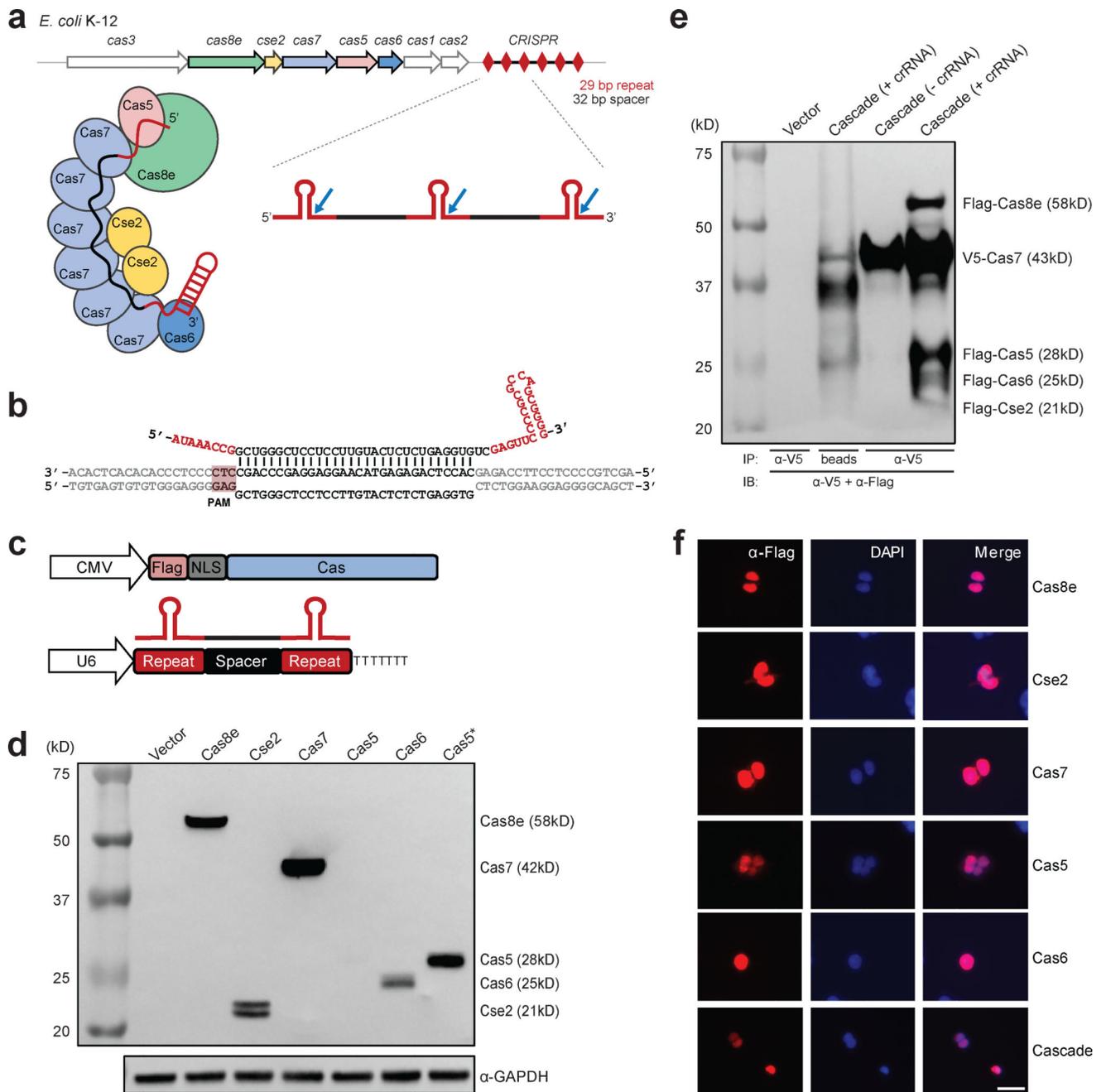
1. Gaj T, Gersbach CA & Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31, 397–405, doi:10.1016/j.tibtech.2013.04.004 (2013). [PubMed: 23664777]
2. Barrangou R & Doudna JA Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* 34, 933–941, doi:10.1038/nbt.3659 (2016). [PubMed: 27606440]
3. Koonin EV, Makarova KS & Zhang F Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 37, 67–78, doi:10.1016/j.mib.2017.05.008 (2017). [PubMed: 28605718]
4. Makarova KS et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13, 722–736, doi:10.1038/nrmicro3569 (2015). [PubMed: 26411297]
5. Zetsche B et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771, doi:10.1016/j.cell.2015.09.038 (2015). [PubMed: 26422227]
6. Cox DBT et al. RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027, doi:10.1126/science.aag0180 (2017). [PubMed: 29070703]
7. Majumdar S et al. Three CRISPR-Cas immune effector complexes coexist in *Pyrococcus furiosus*. *Rna* 21, 1147–1158, doi:10.1261/rna.049130.114 (2015). [PubMed: 25904135]
8. Bolotin A, Quinquis B, Sorokin A & Ehrlich SD Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561, doi:10.1099/mic.0.28048-0 (2005). [PubMed: 16079334]
9. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J & Almendros C Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740, doi:10.1099/mic.0.023960-0 (2009). [PubMed: 19246744]
10. 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, doi:10.1074/jbc.M113.472233 (2013). [PubMed: 23760266]
11. Brouns SJ et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964, doi:10.1126/science.1159689 (2008). [PubMed: 18703739]
12. Jore MM et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18, 529–536, doi:10.1038/nsmb.2019 (2011). [PubMed: 21460843]
13. Wiedenheft B et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477, 486–489, doi:10.1038/nature10402 (2011). [PubMed: 21938068]
14. Jackson RN et al. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* 345, 1473–1479, doi:10.1126/science.1256328 (2014). [PubMed: 25103409]
15. Zhao H et al. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* 515, 147–150, doi:10.1038/nature13733 (2014). [PubMed: 25118175]
16. Hayes RP et al. Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* 530, 499–503, doi:10.1038/nature16995 (2016). [PubMed: 26863189]
17. Gesner EM, Schellenberg MJ, Garside EL, George MM & Macmillan AM Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* 18, 688–692, doi:10.1038/nsmb.2042 (2011). [PubMed: 21572444]
18. Beloglazova N et al. CRISPR RNA binding and DNA target recognition by purified Cascade complexes from *Escherichia coli*. *Nucleic Acids Res* 43, 530–543, doi:10.1093/nar/gku1285 (2015). [PubMed: 25488810]
19. Jinek M et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821, doi:10.1126/science.1225829 (2012). [PubMed: 22745249]
20. Gilbert LA et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451, doi:10.1016/j.cell.2013.06.044 (2013). [PubMed: 23849981]
21. Perez-Pinera P et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10, 973–976, doi:10.1038/nmeth.2600 (2013). [PubMed: 23892895]

22. Maeder ML et al. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10, 977–979, doi:10.1038/nmeth.2598 (2013). [PubMed: 23892898]
23. Hilton IB et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33, 510–517, doi:10.1038/nbt.3199 (2015). [PubMed: 25849900]
24. Chavez A et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12, 326–328, doi:10.1038/nmeth.3312 (2015). [PubMed: 25730490]
25. Luo ML, Mullis AS, Leenay RT & Beisel CL Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res* 43, 674–681, doi:10.1093/nar/gku971 (2015). [PubMed: 25326321]
26. Westra ER 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, doi:10.1016/j.molcel.2012.03.018 (2012). [PubMed: 22521689]
27. Hochstrasser ML, Taylor DW, Kornfeld JE, Nogales E & Doudna JA DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Mol Cell* 63, 840–851, doi:10.1016/j.molcel.2016.07.027 (2016). [PubMed: 27588603]
28. Leenay RT et al. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* 62, 137–147, doi:10.1016/j.molcel.2016.02.031 (2016). [PubMed: 27041224]
29. Wu X et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 32, 670–676, doi:10.1038/nbt.2889 (2014). [PubMed: 24752079]
30. 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, doi:10.1038/nbt.2916 (2014). [PubMed: 24837660]
31. Thakore PI et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* 12, 1143–1149, doi:10.1038/nmeth.3630 (2015). [PubMed: 26501517]
32. Bolukbasi MF, Gupta A & Wolfe SA Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Methods* 13, 41–50, doi:10.1038/nmeth.3684 (2016). [PubMed: 26716561]
33. Di H et al. Comparative analysis of CRISPR loci in different *Listeria monocytogenes* lineages. *Biochem Biophys Res Commun* 454, 399–403, doi:10.1016/j.bbrc.2014.10.018 (2014). [PubMed: 25445602]
34. Rath D, Amlinger L, Hoekzema M, Devulapally PR & Lundgren M Efficient programmable gene silencing by Cascade. *Nucleic Acids Res* 43, 237–246, doi:10.1093/nar/gku1257 (2015). [PubMed: 25435544]
35. Margolin JF et al. Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci U S A* 91, 4509–4513 (1994). [PubMed: 8183939]
36. Klann TS et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* 35, 561–568, doi:10.1038/nbt.3853 (2017). [PubMed: 28369033]
37. Sinkunas T et al. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J* 32, 385–394, doi:10.1038/emboj.2012.352 (2013). [PubMed: 23334296]
38. Chen F et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun* 8, 14958, doi:10.1038/ncomms14958 (2017). [PubMed: 28387220]
39. Amabile A et al. Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* 167, 219–232.e214, doi:10.1016/j.cell.2016.09.006 (2016). [PubMed: 27662090]
40. Gleditsch D et al. Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res* 44, 5872–5882, doi:10.1093/nar/gkw469 (2016). [PubMed: 27216815]
41. Kleinstiver BP et al. Monomeric site-specific nucleases for genome editing. *Proc Natl Acad Sci U S A* 109, 8061–8066, doi:10.1073/pnas.1117984109 (2012). [PubMed: 22566637]
42. Beurdeley M et al. Compact designer TALENs for efficient genome engineering. *Nat Commun* 4, 1762, doi:10.1038/ncomms2782 (2013). [PubMed: 23612303]
43. Kleinstiver BP et al. The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3 (Bethesda, Md.)* 4, 1155–1165, doi:10.1534/g3.114.011445 (2014).

44. Dolan AE et al. Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Mol Cell In Press* (2019).

## References (Methods only)

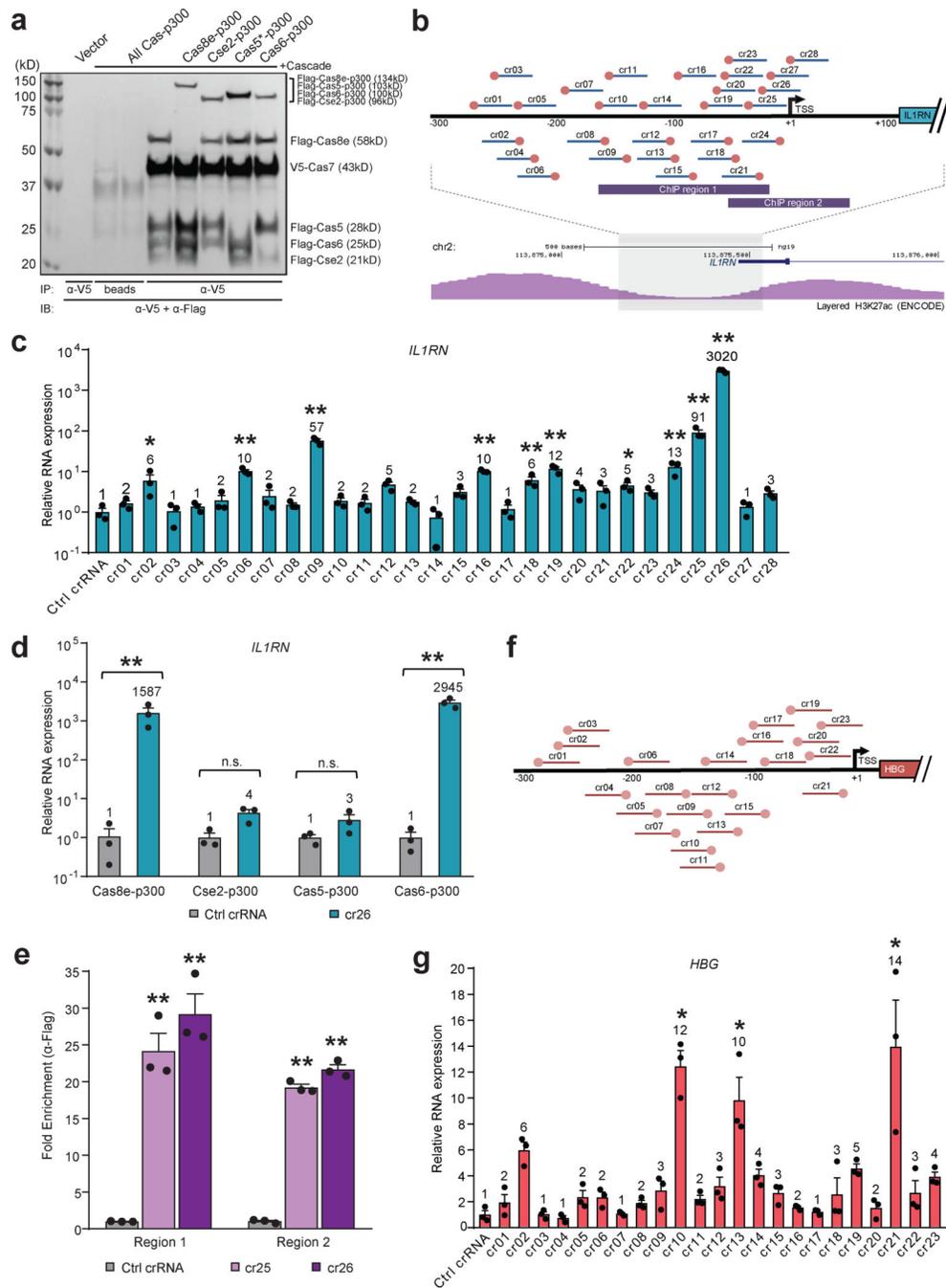
45. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* 29, 15–21, doi:10.1093/bioinformatics/bts635 (2013).
46. Liao Y, Smyth GK & Shi W The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41, e108, doi:10.1093/nar/gkt214 (2013). [PubMed: 23558742]
47. O’Leary NA et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44, D733–745, doi:10.1093/nar/gkv1189 (2016). [PubMed: 26553804]
48. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550, doi:10.1186/s13059-014-0550-8 (2014). [PubMed: 25516281]
49. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30, 2114–2120, doi:10.1093/bioinformatics/btu170 (2014).
50. Langmead B, Trapnell C, Pop M & Salzberg SL Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25, doi:10.1186/gb-2009-10-3-r25 (2009). [PubMed: 19261174]
51. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74, doi:10.1038/nature11247 (2012). [PubMed: 22955616]
52. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137, doi:10.1186/gb-2008-9-9-r137 (2008). [PubMed: 18798982]
53. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* 26, 841–842, doi:10.1093/bioinformatics/btq033 (2010).
54. Rice P, Longden I & Bleasby A EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG* 16, 276–277 (2000). [PubMed: 10827456]
55. Pinello L et al. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* 34, 695–697, doi:10.1038/nbt.3583 (2016). [PubMed: 27404874]
56. Edgar R, Domrachev M & Lash AE Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30, 207–210 (2002). [PubMed: 11752295]
57. Barrett T et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41, D991–995, doi:10.1093/nar/gks1193 (2013). [PubMed: 23193258]



**Figure 1. EcoCascade expression and complex formation in human cells.**

(a) Schematic of type I-E CRISPR-Cas system in *E. coli* K-12 showing EcoCascade stoichiometry and crRNA processing. Genes encoding proteins comprising the EcoCascade complex are presented in different colors. CRISPR repeats are indicated with a red diamond. Cas6 cleaves the primary CRISPR RNA (crRNA) transcript at the regions indicated with blue arrows yielding mature crRNAs. (b) Schematic representation of processed crRNA with 5' PAM recognition and base pairing at the DNA target site. (c) Cas subunits driven by human cytomegalovirus (CMV) promoter and pre-processed crRNA driven by U6 promoter for expression and processing in mammalian cells. (d) Western blot showing expression of

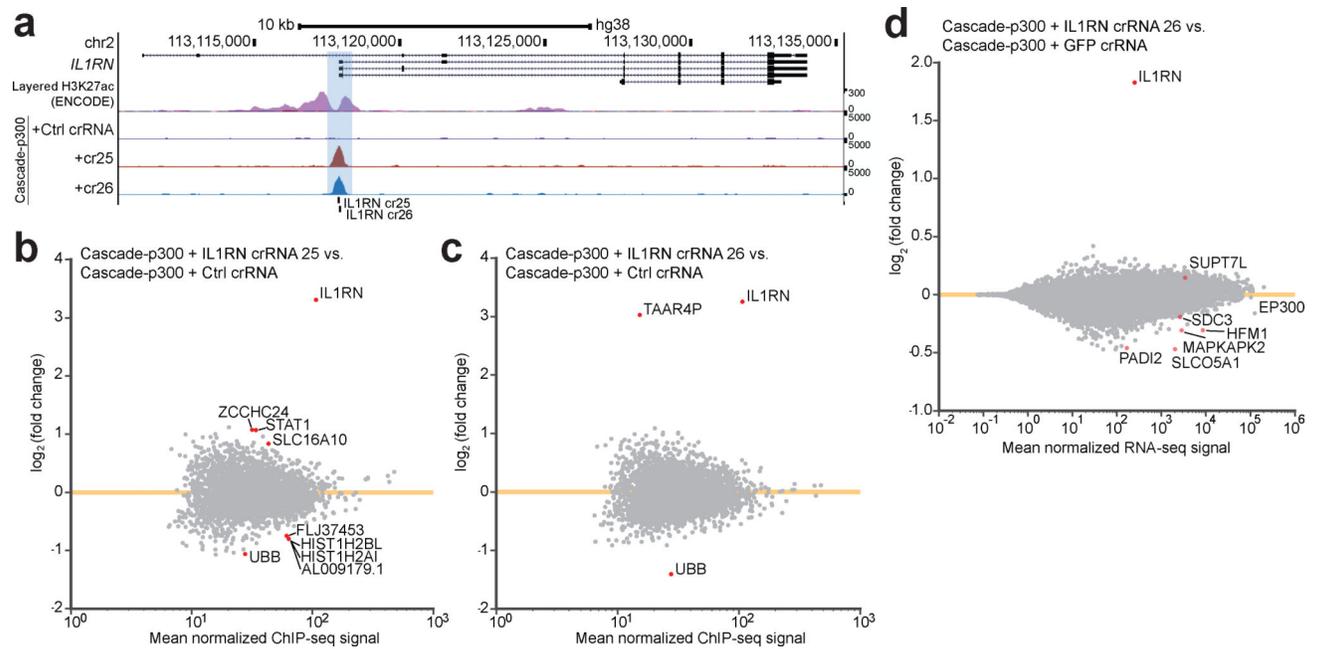
human codon-optimized individual Cas proteins and GAPDH loading control in human HEK293T cells. (\* indicates second round of codon optimization). **(e)** Co-IP and western blot showing crRNA-dependent Cascade formation following co-transfection with V5-tagged Cas7 and Flag-tagged Cas8e, Cse2, Cas5, and Cas6 (immunoprecipitation with  $\alpha$ -V5, and detection with  $\alpha$ -V5 and  $\alpha$ -Flag). **(f)** Immunofluorescence imaging showing engineered Cas subunits with NLSs enables import into the human nucleus. Red indicates Cas subunit; scale bar, 25 $\mu$ m. For **d-f**, two independent experiments were conducted with similar results. All samples processed at 3 days post-transfection. CMV, human cytomegalovirus; CoIP, Co-immunoprecipitation; NLS, nuclear localization signal.



**Figure 2. EcoCascade activates transcription of endogenous genes in human cells.**

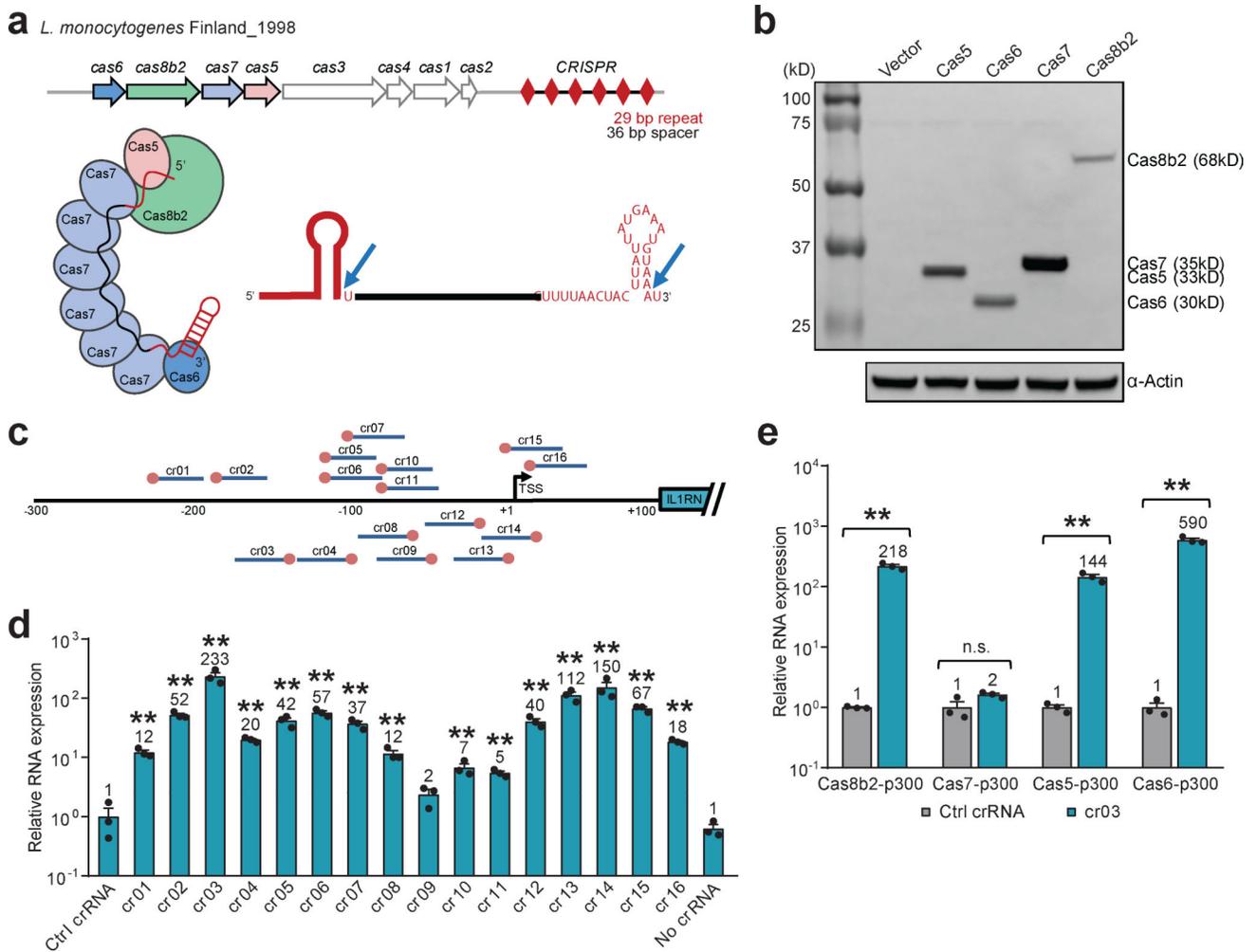
(a) Co-IP showing EcoCascade formation following co-transfection with plasmids encoding the crRNA, Flag-tagged EcoCascade subunits with V5-tagged Cas7, and various Cas-p300 fusions. IP performed with α-V5 and IB performed with α-V5 and α-Flag. Two independent experiments were conducted with similar results. (b) Schematic of the *IL1RN* locus along with tiled *IL1RN* crRNA target sites. H3K27ac enrichment from the ENCODE Consortium is indicated 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 individual crRNAs (100 ng) and EcoCascade (50 ng Cas8e, 100 ng Cse2, 50 ng Cas7, 250 ng Cas5\*) with Cas6-p300 (50 ng). (n=3 biological independent samples; mean  $\pm$  SEM). **(d)** Relative *ILIRN* expression following co-transfection of Ctrl crRNA or cr26 and EcoCascade with various Cas-p300 effectors. (n=3 biological independent samples; mean  $\pm$  SEM). **(e)** ChIP-qPCR enrichment following co-transfection of individual crRNAs with EcoCascade and Cas6-p300. IP performed with  $\alpha$ -Flag and qPCR performed with primers for amplicon regions designated in **b**. (n=3 biological independent samples; mean  $\pm$  SEM; 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. (n=3 biological independent samples; mean  $\pm$  SEM). All samples processed at 3 days post transfection. (Tukey-test following log transformation, \*\* $P$ <0.001 and \* $P$ <0.05 compared to Ctrl crRNA). Numbers above bars indicate mean relative expression. TSS, Transcription start site; Ctrl crRNA, Control crRNA; IP, immunoprecipitation; IB, immunoblot; ChIP, chromatin immunoprecipitation; qPCR, quantitative PCR.



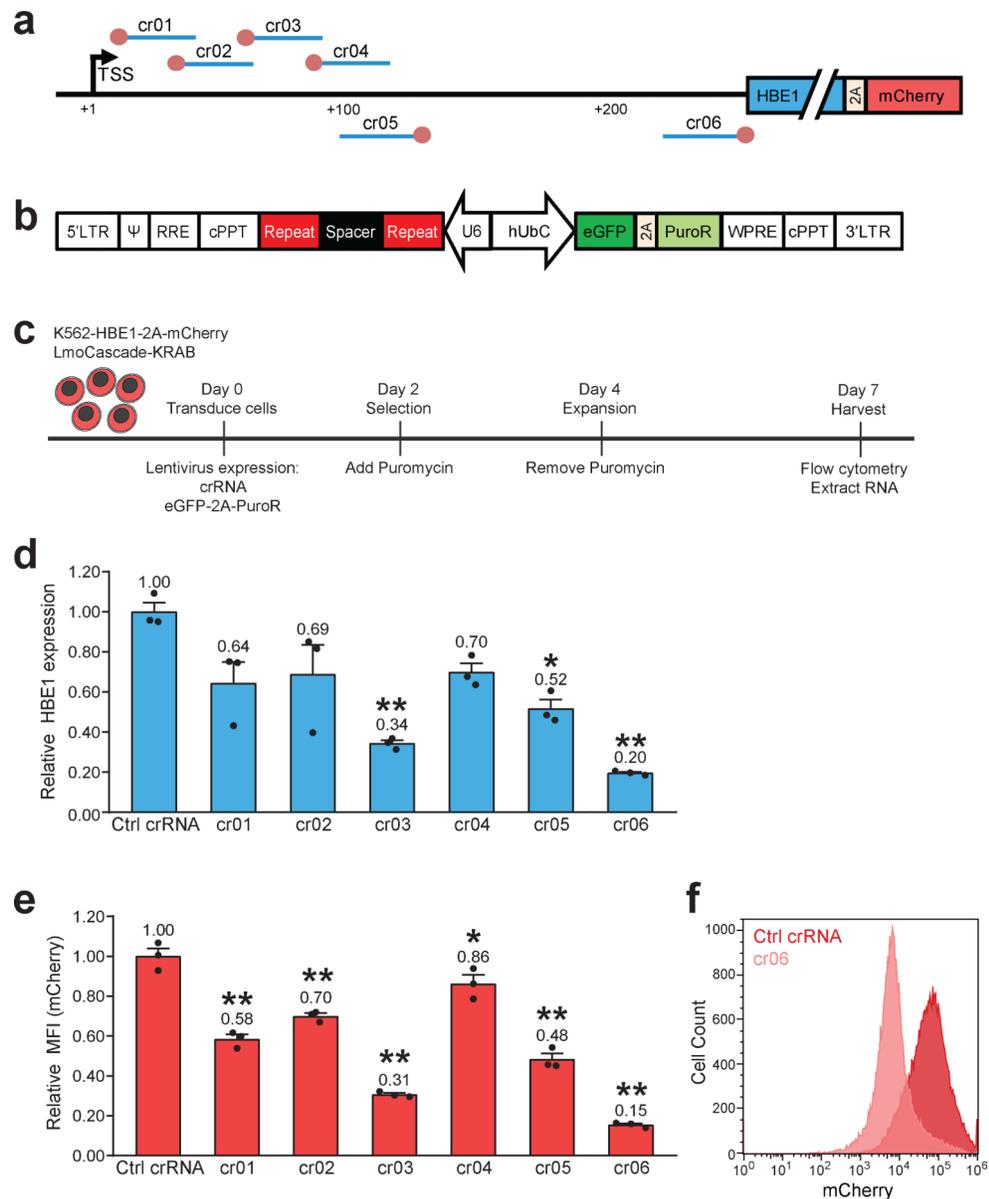
**Figure 3. Genome-wide specificity of EcoCascade-p300 targeted to *IL1RN*.**

(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 Ctrl 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 (b) cr25 and (c) cr26 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. (n=3 biological independent samples). Ctrl crRNA, Control crRNA.



**Figure 4. LmoCascade activates transcription of *ILIRN* gene in human cells.**

(a) Schematic of type I-B CRISPR-Cas system in *L. monocytogenes* Finland\_1998 showing predicted LmoCascade stoichiometry and crRNA processing. Genes encoding proteins comprising the LmoCascade complex are presented in different colors. CRISPR repeats are indicated with a red diamond. 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 *ILIRN* locus along with tiled *ILIRN* crRNA target sites. (d) Relative *ILIRN* expression following co-transfection of individual crRNAs and LmoCascade with Cas6-p300. (e) Relative *ILIRN* expression following co-transfection of Ctrl crRNA or cr03 and LmoCascade with various Cas-p300 effectors. All samples were processed at 3 days post transfection. (Tukey-test following log transformation, \*\* $P < 0.001$  compared to Ctrl crRNA,  $n = 3$  biological independent samples; error bars, SEM). Numbers above bars indicate mean relative expression. TSS, Transcription start site; Ctrl crRNA, Control crRNA; qPCR, quantitative PCR.



**Figure 5. LmoCascade represses transcription of *HBE1* gene in human cells.**

(a) Schematic of the *HBE1* locus along with tiled *HBE1* crRNA target sites. (b) Schematic of lentiviral expression constructs with dual promoters for co-expression of crRNAs with eGFP and Puromycin<sup>R</sup>. (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 two days with puromycin and harvested on day seven. (d) Relative *HBE1* expression measured by qPCR. (e) Relative 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 Ctrl crRNA,  $n = 3$  biological independent samples; error bars, SEM). Numbers above bars indicate

mean relative expression. TSS, Transcription start site; Ctrl crRNA, Control crRNA; qPCR, quantitative PCR. MFI, mean fluorescence intensity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript