Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts

Campa, Carlo C.; Weisbach, Niels R.; Santinha, António J.; Incarnato, Danny; Platt, Randall J.

Published in:
Nature Methods

DOI:
10.1038/s41592-019-0508-6

Publication date:
2019

nature methods

ARTICLES

https://doi.org/10.1038/s41592-019-0508-6

Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts

Carlo C. Campa[1,4], Niels R. Weisbach[1,4], António J. Santinha[1], Danny Incarnato[2] and Randall J. Platt[1,3]*

The ability to modify multiple genetic elements simultaneously would help to elucidate and control the gene interactions and networks underlying complex cellular functions. However, current genome engineering technologies are limited in both the number and the type of perturbations that can be performed simultaneously. Here, we demonstrate that both Cas12a and a clustered regularly interspaced short palindromic repeat (CRISPR) array can be encoded in a single transcript by adding a stabilizer tertiary RNA structure. By leveraging this system, we illustrate constitutive, conditional, inducible, orthogonal and multiplexed genome engineering of endogenous targets using up to 25 individual CRISPR RNAs delivered on a single plasmid. Our method provides a powerful platform to investigate and orchestrate the sophisticated genetic programs underlying complex cell behaviors.

CRISPR-associated (Cas) nucleases are versatile RNA-guided enzymes that facilitate a wide range of genome engineering applications[1,2]. Through the binding of short RNA molecules, known as CRISPR RNAs (crRNAs), members of the Cas nuclease family are programmed to edit specific genomic loci thereby facilitating systematic investigation of gene function[3,4]. In addition, nuclease-inactive Cas enzymes fused to transcriptional effectors enable fine control of gene expression[5–10]. Remarkably, use of either different Cas enzymes or engineered crRNAs enable distinct gene perturbations, including gene knockout, gene activation and gene repression[11–16]. Despite this versatility, either the heterologous expression of different proteins (for example, Csy4, scFV) or the presence of long RNA-based regulatory regions (for example, ribozymes, aptamers) are required to permit regulation of multiple genes using current platforms[17–19]. This inevitably limits the scalability of CRISPR-based multiplexed genome engineering approaches and consequently the possibility to investigate complex cell behaviors.

To coordinate a myriad of different processes using a limited number of cellular elements, cells evolved to maximize the efficiency of their molecular components. This principle is adopted by distinct organisms to rapidly match different environmental contexts, thereby supporting the notion that maximizing the performance of each molecular element is a general principle on which biological systems based their own survival[20]. As an example, protection against bacteriophages and other foreign genetic elements is mediated by Cas12a, a class II type V CRISPR-Cas system[21,22]. Functioning as a both RNase and DNase, Cas12a controls both processing and maturation of its own crRNA as well as DNA target cleavage[21–23]. CRISPR arrays associated with Cas12a are transcribed as a long RNA transcript, termed pre-crRNA, that contains a succession of ~30-nucleotide (nt) spacer sequences, separated by 36-nt direct repeat sequences (DRs)[21,22]. Cas12a-mediated pre-crRNA processing into mature crRNAs, as well as the recognition and cleavage of target DNA is controlled by Watson–Crick base pairing between the spacer and target sequences[21,22]. In mammalian cells, Cas12a has been used for gene editing and transcriptional gene control[24]. Optimization of this genome editing system for mammalian cells was obtained by DR length reduction and adoption of distinct promoters that differentially drive transcription of Cas12a and pre-crRNA[25–27]. Although the implementation of these strategies enables a wide range of applications, they fail to harness the full potential of Cas12a in genome engineering.

Here, we leverage the dual RNase/DNase function of *Acidaminococcus* sp. Cas12a to develop SiT-Cas12a, encoding Cas12a and dozens of crRNAs in a single transcript for multiplexed genome engineering. Stabilization of SiT-Cas12a transcripts through inclusion of a tertiary structural motif improves both pre-crRNA processing as well as Cas12a production. When coupled with transcriptional effectors, SiT-Cas12a enables multiplexed orthogonal gene transcriptional regulation and editing thereby providing a scalable way to elucidate and control the gene network's underlying cellular functions.

Results

Compact encoding of Cas12a and CRISPR arrays in a single Pol II-derived transcript. In mammalian cells, distinct promoters control transcription of different RNA molecules. While Pol II promoters are mainly used for transcription of coding genes characterized by long RNA sequences, Pol III promoters are employed for production of small non-coding RNAs including crRNAs[28]. To assess whether transcripts derived from Pol II promoters could facilitate Cas12a-based genome engineering applications[26], we expressed a crRNA targeting *DNMT1* from either a U6 (Pol III) or EF1a (Pol II) promoter along with ectopically expressed Cas12a followed by quantification of insertions and deletions (indels) 72 h after transient transfection. Expression of crRNA from both Pol II and Pol III promoters resulted in comparable gene editing efficiencies (Fig. 1a,b, conditions I and II), indicating that Cas12a processes and uses crRNAs derived from both promoter types.

To determine whether Pol II promoters facilitate simultaneous protein and crRNA expression, we cloned a CRISPR array containing a spacer targeting *DNMT1* in the 3′ untranslated region of the

[1]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [2]Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, the Netherlands. [3]Department of Chemistry, University of Basel, Basel, Switzerland. [4]These authors contributed equally: Carlo C. Campa, Niels R. Weisbach. *e-mail: rplatt@ethz.ch

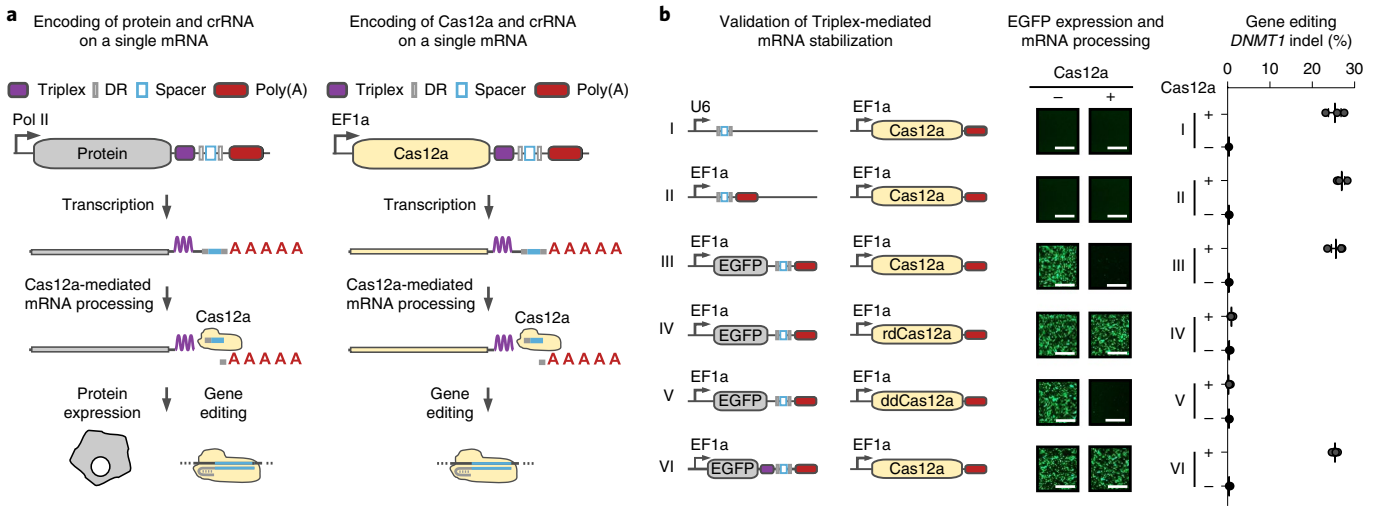

Fig. 1 | Simultaneous control of protein and crRNA expression from Pol II promoters. a, Schematic of a single transcript architecture containing both a protein-coding sequence and a CRISPR array. The transcript encodes for: Protein/Cas12a is a gray/yellow rectangle; Triplex, a tertiary RNA structural motif is a small purple rectangle; DR is a gray square; spacer is a blue square and poly(A) is a red rectangle. **b**, Validation of Triplex-mediated mRNA stabilization for concomitant gene editing and protein expression. Combinations of crRNA expression constructs and Cas12a proteins used (conditions I–VI). Representative EGFP fluorescent images after cotransfection of different plasmid combinations (conditions I–VI). Scale bar, 1.00 mm. Quantification of gene editing efficiencies (conditions I–VI). Values represent mean ± s.e.m., $n = 3$ independent experiments.

enhanced green fluorescence protein (*EGFP*) gene, and assessed both gene editing efficiency at the endogenous *DNMT1* locus and EGFP expression 72 h after transient transfection in human embryonic kidney (HEK) 293T cells (Fig. 1b, condition III). In cells harboring *EGFP* transcripts containing CRISPR arrays, we observed complete loss of EGFP expression (Fig. 1b, condition III), suggesting destabilization of *EGFP* transcripts mediated by the RNase activity of Cas12a. Consistently, expression of an RNase dead Cas12a (rdCas12a), but not DNase dead Cas12a (ddCas12a), rescued EGFP expression (Fig. 1b, conditions IV and V). Taken together, this suggests that Cas12a-mediated crRNA processing via the RNase domain results in efficient cleavage and destabilization of protein-coding messenger RNAs, which is likely a result of removal of the polyadenylation (poly(A)) tail.

To overcome mRNA destabilization and enable simultaneous expression of protein and crRNA on the same transcript, we leveraged a 110-nt structure derived from the 3′ end of the mouse non-coding RNA Metastasis-associated lung adenocarcinoma transcript 1 (*Malat1*), previously described to stabilize transcripts lacking poly(A) tails through the formation of a tertiary structure element termed 'Triplex'[29]. We cloned the Triplex sequence between the *EGFP* coding sequence and the CRISPR array (Fig. 1b, condition VI), which effectively rescued EGFP expression without affecting gene editing efficiency (Fig. 1b, condition VI). These results indicate that a Triplex sequence positioned at the 3′ of a protein-coding gene stabilizes mRNAs after Cas12a-mediated RNA processing, enabling concomitant protein expression and gene editing.

Constitutive, conditional and inducible gene editing with SiT-Cas12a. To determine whether Cas12a and CRISPR arrays could be compactly encoded on single Pol II-driven mRNA, we developed single-transcript Cas12a (SiT-Cas12a) composed of a: (1) Pol II promoter EF1a; (2) Cas12a derived from *Acidaminococcus* sp.; (3) Triplex sequence; (4) CRISPR array containing spacers targeting a set of mammalian genomic loci and (5) poly(A) signal (Fig. 1a). We evaluated the platform using a CRISPR array containing a spacer targeting *DNMT1*, and observed consistent and efficient gene editing at the *DNMT1* locus (Supplementary Fig. 1a).

We generated conditional and inducible SiT-Cas12a platforms, termed SiT-Cas12a-[Cond] (Supplementary Fig. 1b) and SiT-Cas12a-[Ind], respectively (Supplementary Fig. 1c). SiT-Cas12a-[Cond] relies on a Lox-Stop-Lox (LSL) cassette positioned downstream of an EF1a promoter and upstream of the SiT-Cas12a coding region. To demonstrate conditional genome editing, we cotransfected HEK 293T cells with SiT-Cas12a-[Cond] and either a Cre-recombinase encoding plasmid or a control plasmid. At 72 h post-transfection, we detected genome editing events exclusively in Cre recombinase expressing cells (Supplementary Fig. 1b). SiT-Cas12a-[Ind] relies on a Tetracycline responsive element positioned upstream of a minimal cytomegalovirus (CMV) promoter (minCMV) that, different from the constitutive promoter EF1a, resulted in gene editing efficiencies proportional to inducer concentration (doxycycline) both in a Tet-on and Tet-off configuration (Supplementary Fig. 1c). Taken together, SiT-Cas12a enables either constitutive or conditional and inducible gene editing through fine temporal control of Cas12a and crRNA expression.

Multiplexed gene editing with SiT-Cas12a. We evaluated the potential of the SiT-Cas12a platform for multiplexed gene editing using a CRISPR array containing five distinct spacers targeting different genomic loci (*FANCF1, EMX1, GRIN2B, VEGF, DNMT1*) and quantified both mature crRNAs and gene editing efficiency (Fig. 2a). In the SiT-Cas12a context, we observed expression and processing of each mature crRNA, albeit with varying efficiencies as observed in previous studies[25] (Fig. 2a). In addition, transcripts stabilized by the Triplex sequence (Fig. 2a, SiT-Cas12a) increased both Cas12a expression (Supplementary Fig. 2a) and processed crRNA abundance compared to transcripts without the Triplex (Fig. 2a, Cas12a). In line with these results, the gene editing efficiency was higher in cells expressing SiT-Cas12a compared to a control lacking the Triplex structure between Cas12a and the CRISPR array (Fig. 2b). The increased crRNA production facilitated by the Triplex sequence disappeared on mutation of the RNase domain of Cas12a (SiT-rdCas12a) (Fig. 2a). Consistently, gene editing efficiencies were negligible for the SiT-Cas12a RNase inactive mutant and found to be comparable to the DNase inactive mutant of Cas12a (SiT-ddCas12a) (Fig. 2b).

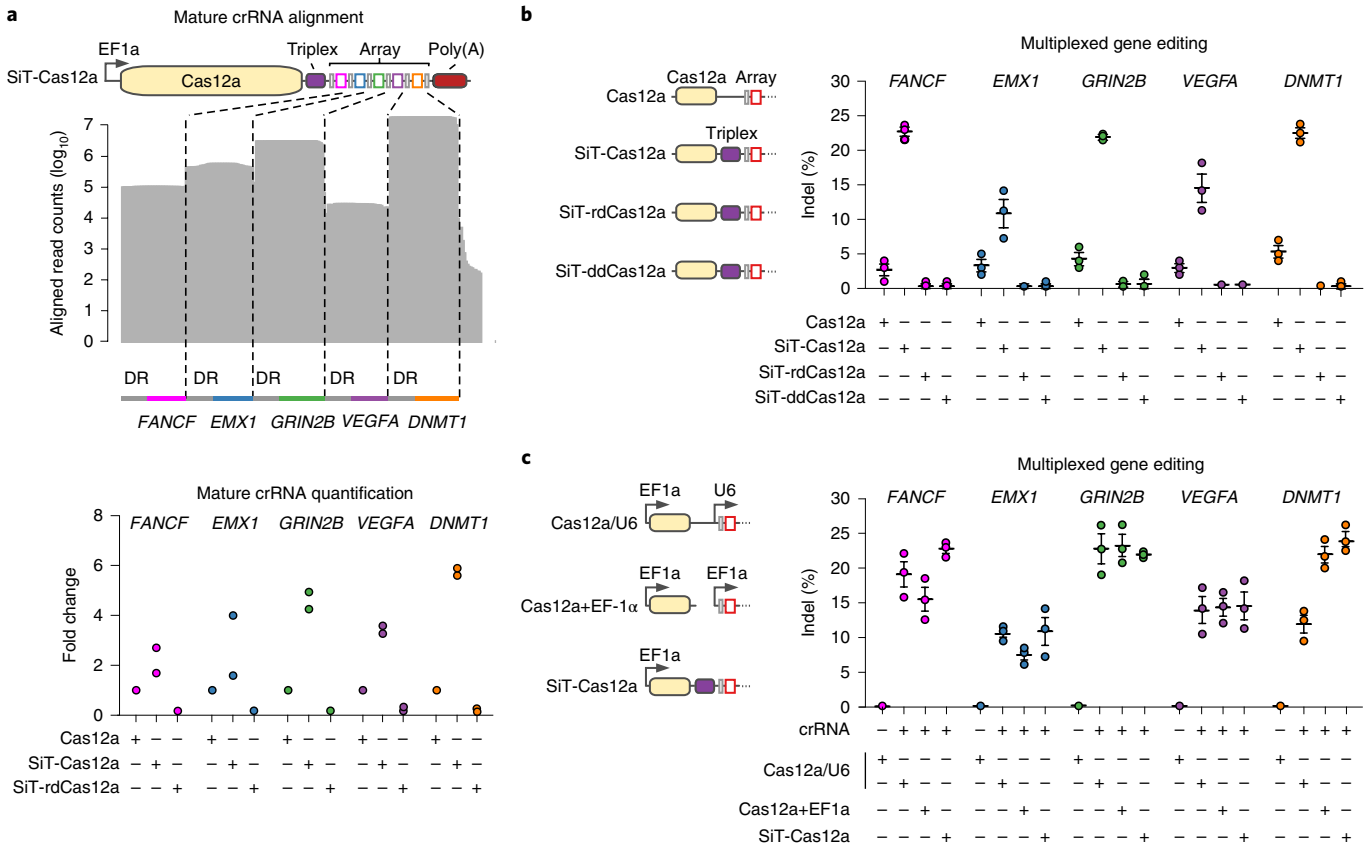

Fig. 2 | Multiplexed genome editing by SiT-Cas12a. a, Schematic of SiT-Cas12a, which includes the Pol II promoter EF1a, Cas12a (yellow rectangle), Triplex (purple rectangle), CRISPR array (colored and gray squares) and poly(A) (red rectangle). Bottom panel: quantification of mature crRNAs. $n=2$ independent experiments. **b**, Schematics of the Cas12a, SiT-Cas12a, rdSiT-Cas12a and ddSiT-Cas12a constructs. Same color scheme as in **a**. The poly(A) signal is not displayed. Quantification of multiplexed gene editing efficiencies. Values represent mean ± s.e.m., $n=3$ independent experiments. **c**, Schematics of the plasmid combinations used. Quantification of multiplexed gene editing efficiencies in cells transfected with EF1a-Cas12a and U6-crRNA on the same plasmid (Cas12a/U6), EF1a-Cas12a and EF1a-crRNA on two different plasmids (Cas12a+EF1a) or SiT-Cas12a on a single plasmid. Values represent mean ± s.e.m., $n=3$ independent experiments.

Next, we compared the gene editing efficiency of SiT-Cas12a with previously reported Cas12a platforms based on independent transcription of Cas12a and a CRISPR array from distinct promoters[25,26]. Expression of SiT-Cas12a resulted in gene editing efficiencies equal to or higher than other tested platforms (Fig. 2c). Taken together, these data demonstrate that compact encoding of Cas12a and a CRISPR array on Pol II transcripts mediates efficient multiplexed gene editing.

In contrast to Pol III promoters, Pol II promoters express transcripts of seemingly unlimited length[30]. To leverage this property, we cloned a CRISPR array harboring ten distinct spacer sequences targeting the *CD47* locus in the SiT-Cas12a context, either singularly or jointly, and performed gene editing quantification (Supplementary Fig. 2b–g). While the gene editing efficiency using single-crRNAs ranged from 2 to 17%, simultaneous expression of all crRNAs increased the gene editing efficiency up to 60% (Supplementary Fig. 2c), indicating that the targeting of multiple crRNAs in the same coding gene introduced more loss of function mutations. We obtained similar gene editing efficiency in cells infected with a SiT-Cas12a-based lentiviral vector enabling stable expression and delivery to cells that are difficult to transfect (Supplementary Fig. 3a–e). In agreement, multiple independent indels at each target site as well as large fragment deletions between target sites were detected using deep sequencing in cells expressing such a large CRISPR array (Supplementary Fig. 2d–g), confirming the targeting of all crRNAs used. Consistently, single cell analysis of CD47 protein expression showed a four-fold increase in CD47-negative cells, reaching 37% of the total population when using multiple crRNAs (Supplementary Fig. 4a).

Finally, we evaluated whether this strategy could be employed to increase the efficiency of generating multi-gene knockouts after a simple transient transfection experiment. We cloned distinct spacers targeting different coding genes (*CD47*, *CD166* and *CD97*) (Supplementary Fig. 4b) in the SiT-Cas12a context and measured the rate of single (either CD47, CD166 or CD97), double (CD47/CD166, CD47/CD97, CD166/CD97) or triple (CD47/CD166/CD97) knockout by single cell flow cytometry analysis. Targeting multiple sites per gene using SiT-Cas12a increased the rate of single, double and triple multiplexed gene knockout generation by two- to three-fold compared to single crRNA conditions (Supplementary Fig. 4b). Overall, these results demonstrate that SiT-Cas12a facilitates scalable and multiplexed genome engineering.

Multiplexed transcriptional regulation with SiT-Cas12a. To develop SiT-Cas12a-based platforms to control the expression of endogenous genes, we fused one, two or three copies of the Krüppel associated box (KRAB) domain of the transcriptional repressor ZNF10 (ref. [31]) to the C-terminus of ddCas12a, thus generating ddCas12a-$KRAB_1$, ddCas12a-$KRAB_2$ and ddCas12a-[Repr], respectively (Supplementary Fig. 5a). To assess the efficiency of the

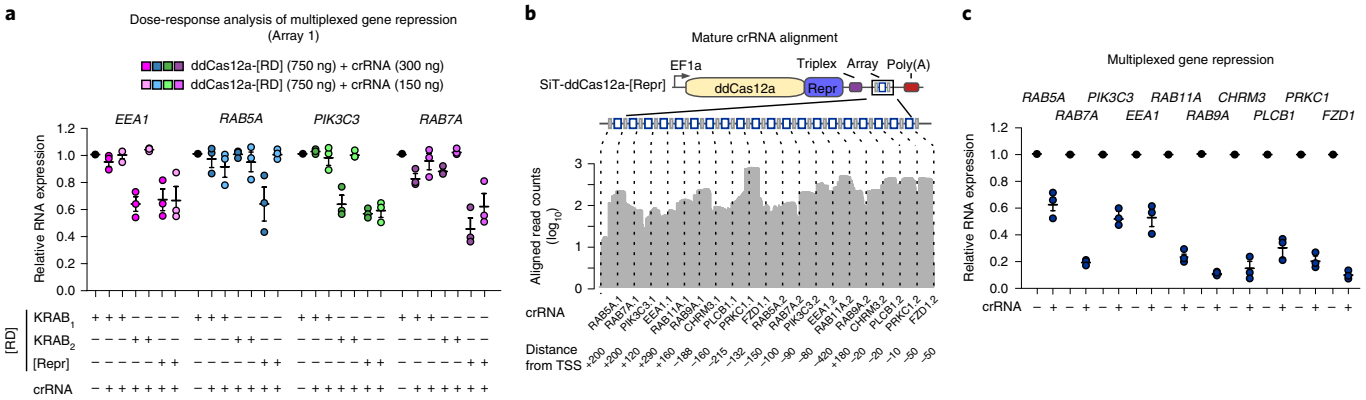

Fig. 3 | Multiplexed transcriptional repression using SiT-Cas12a-[Repr]. a, Quantification of relative mRNA expression for indicated genes in cells cotransfected with ddCas12a fused with one (KRAB), two ($KRAB_2$) or three ([Repr]) KRAB domains in combination with different crRNA array concentrations. Values represent mean ± s.e.m., $n = 3$ independent experiments. **b**, Alignments of mature crRNAs in cells transfected with SiT-ddCas12a-[Repr] containing a CRISPR array with 20 distinct spacers. **c**, Quantification of relative RNA expression for ten distinct genes after multiplexed expression of 20 distinct crRNAs from SiT-ddCas12a-[Repr]. Values represent mean ± s.e.m., $n = 3$ independent experiments.

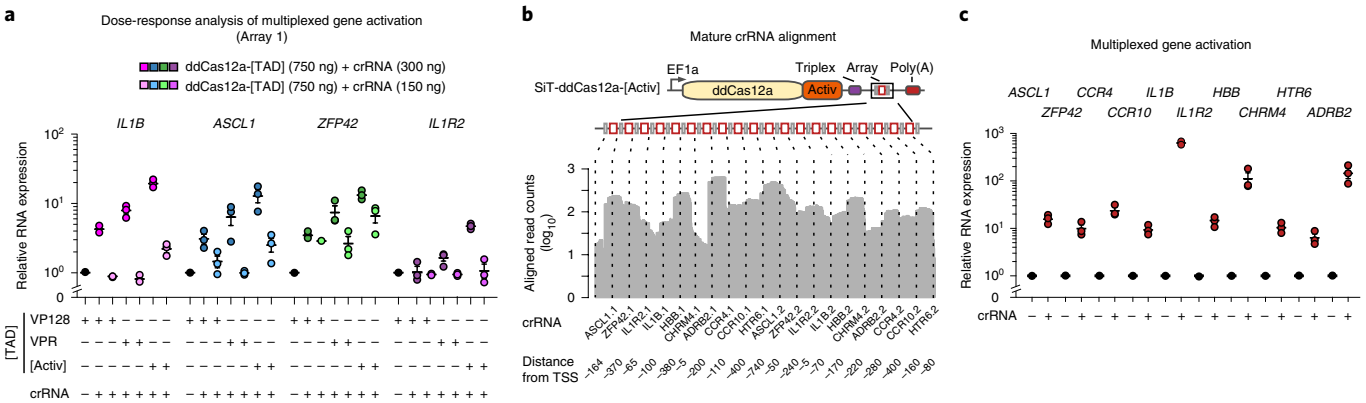

Fig. 4 | Multiplexed transcriptional activation using SiT-Cas12a. a, Quantification of relative RNA expression for indicated genes in cells cotransfected with ddCas12a fusion proteins (VP128, VPR, [Activ]) in combination with different crRNA concentrations. Values represent mean ± s.e.m., $n = 3$ independent experiments. **b**, Alignments of mature crRNAs in cells transfected with SiT-Cas12a-[Activ] containing a CRISPR array with 20 distinct spacers. **c**, Quantification of relative RNA expression for ten distinct genes after multiplexed expression of 20 distinct crRNAs from SiT-Cas12a-[Activ]. Values represent mean ± s.e.m., $n = 3$ independent experiments.

transcriptional repression conferred by multiple KRAB domains, we cotransfected the SiT-Cas12a repressor variants along with different concentrations of CRISPR arrays containing spacers targeting four distinct genomic loci (*RAB5A*, *RAB7A*, *EEA1*, *PIK3C3*). We found that three tandem KRAB domains fused at the C-terminus of ddCas12a (ddCas12a-[Repr]) conferred potent transcriptional repression (Fig. 3a and Supplementary Fig. 5b) with differential efficiencies according to the target gene and distance from the transcription start site (Supplementary Fig. 5c). Next, we cloned a CRISPR array harboring 20 different spacers targeting ten distinct genes (*RAB5A*, *RAB7A*, *PIK3C3*, *EEA1*, *RAB11A*, *RAB9A*, *CHRM3*, *PLCB1*, *PRKC1*, *FZD1*) within SiT-ddCas12a-[Repr] and quantified mature crRNAs and transcriptional repression. Small-RNA-seq analysis confirmed generation of mature crRNAs leading to efficient gene repression ranging from 40 to 80% reduction (Fig. 3b,c).

To further strengthen the finding that SiT-Cas12a enables simplified multi-gene transcriptional gene control, we replaced the transcriptional repression domain KRAB with either the VPR activator[32] or a combination of the p65 activation domain (p65) together with the heat shock factor 1 (HSF1)[3] to generate two RNA-guided transcriptional activators: ddCas12a-VPR and ddCas12a-p65-HSF1 (Supplementary Fig. 6a). Both of these Cas12a-based chimeric proteins induced gene activation at different efficiencies (Supplementary Fig. 6b). Comparative analyses of these gene activators (Fig. 4a,b) showed a ten-fold increase in activation efficiency of ddCas12a-[Activ] compared to ddCas12a-VPR (Supplementary Fig. 6b) also in the presence of limiting spacer concentrations (Fig. 4a and Supplementary Fig. 6c). We also observed that transcriptional activation efficiency varies according to the target gene and distance from the transcription start site (Supplementary Fig. 6d). Next, to explore the potential of multiplexed transcriptional activation in the SiT-Cas12a context, we combined SiT-ddCas12a-[Activ] with a CRISPR array harboring 20 spacers targeting ten distinct genes (*ASCL1*, *ZFP42*, *CCR4*, *CCR10*, *IL1B*, *IL1R2*, *HBB*, *CHRM4*, *HTR6*, *ADRB2*) and quantified mature crRNAs and transcriptional activation. Small RNA-seq analysis confirmed generation of all 20 mature crRNAs (Fig. 4b). In addition, we measured robust gene activation (10–1000-fold) for all target genes (Fig. 4c and Supplementary Fig. 6e). Overall, these data indicate that both repressor and activator transcriptional domains when combined with SiT-Cas12a enable multi-gene transcriptional control.

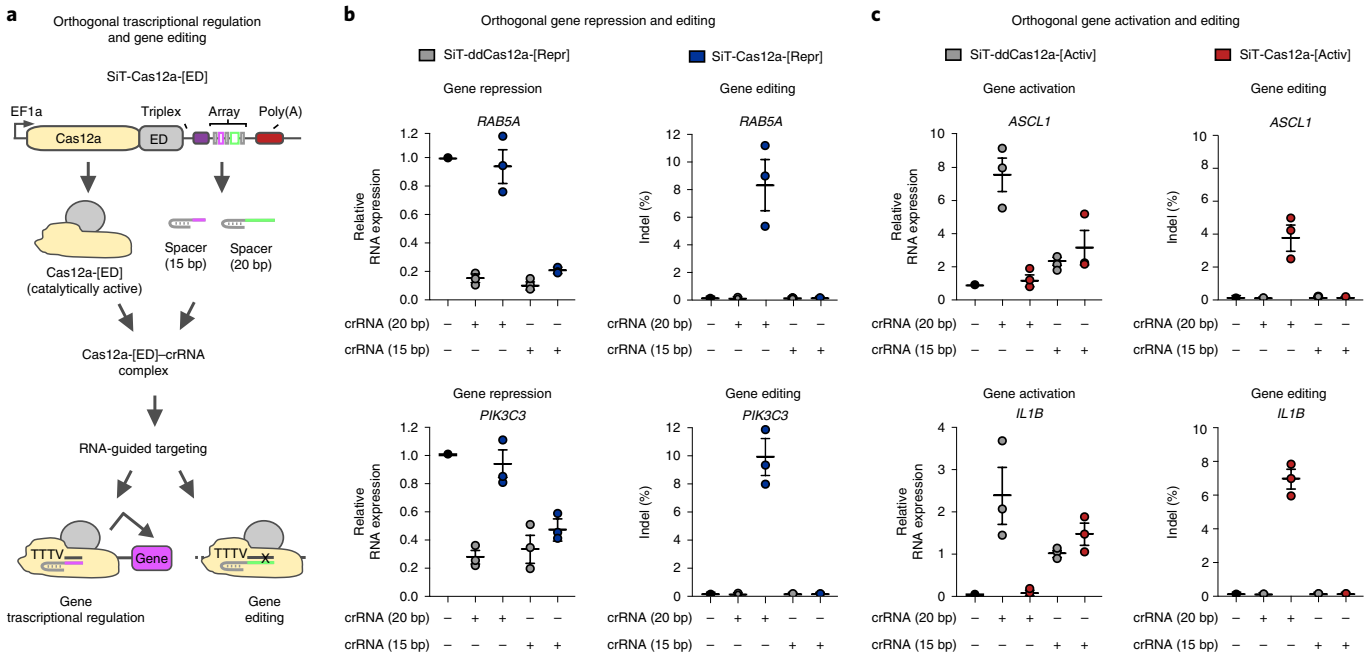

Fig. 5 | Orthogonal transcriptional gene regulation and editing. a, Schematic of the SiT-Cas12a platform for orthogonal gene editing and transcriptional regulation with the Pol II promoter EF1a, Cas12a (yellow rectangle), effector domain (ED, gray rectangle), Triplex (purple rectangle), CRISPR array (colored and gray squares) and poly(A) (red rectangle) indicated. Delivery of SiT-Cas12a-[ED] coupled to a CRISPR array consisting of either long (20 bp) or short (15 bp) spacer sequences enables orthogonal gene editing and transcriptional gene regulation. **b**, Quantification of relative *RAB5A* or *PIK3C3* RNA expression and gene editing efficiencies in cells expressing either long (20 bp) or short (15 bp) spacers in combination with either SiT-Cas12a-[Repr] or ddCas12a-[Repr]. Values represent mean ± s.e.m., $n = 3$ independent experiments. **c**, Quantification of relative RNA expression and gene editing efficiencies for two distinct genes (*ASCL1* and *IL1B*) in cells expressing long (20 bp) or short (15 bp) spacers in combination with either SiT-Cas12a-[Activ] or ddCas12a-[Activ]. Values represent mean ± s.e.m., $n = 3$ independent experiments.

Orthogonal gene editing and transcriptional control with SiT-Cas12a. The Cas12a endonuclease, similar to Cas9, is characterized by unique DNA binding kinetics, which enables binding while avoiding cleavage in the presence of truncated crRNAs[13–15,33]. To fully harness the potential of Cas12a for genome engineering, we set out to explore these unique properties for generating a SiT-Cas12a-based platform that could facilitate orthogonal gene editing and transcriptional gene control.

Toward developing such a platform, we assessed *Acidaminococcus* sp. Cas12a (*As*Cas12a) processing efficiency using CRISPR arrays containing both long (20 base pairs (bp)) and short (15 bp) spacers (Supplementary Fig. 7a). We measured three- to five-fold higher amounts of mature crRNAs in cells expressing arrays containing short spacers compared to those with long spacers (Supplementary Fig. 7b). Furthermore, both short and long spacers generated comparable transcriptional gene control when combining SiT-Cas12a-[Repr], SiT-Cas12a-p65-HSF1 and SiT-Cas12a-[Activ] with two distinct CRISPR arrays (Supplementary Fig. 7c,d). This strategy does not extend with a similar efficiency to *Lachnospiraceae bacterium* Cas12a (*Lb*Cas12a), whereby crRNAs containing short spacers did not induce substantial gene activation (Supplementary Fig. 7e). Next, we evaluated the SiT-Cas12a platform in an orthogonal transcriptional control and gene editing context (Fig. 5a). We combined both active and inactive DNase versions of SiT-Cas12a-based transcriptional repressor and activator (SiT-ddCas12a-[Repr], SiT-Cas12a-[Repr], SiT-ddCas12a-[Activ] and SiT-Cas12a-[Activ]) with two sets of CRISPR arrays harboring spacers targeting two distinct promoters using either short or long spacers (Fig. 5b,c). Subsequently, we quantified gene expression and genome editing and determined that only DNase active SiT-Cas12a effectors combined with 20 bp spacers facilitated gene editing. In contrast, SiT-Cas12a effectors combined with 15 bp spacers induced either gene repression (Fig. 5b) or gene activation (Fig. 5c) with comparable efficiencies and without any detectable gene editing events. Last, large CRISPR arrays containing both short and long spacers enabled coordinated and highly multiplexed regulation of ten distinct genes simultaneously with gene editing of another five distinct genes (Fig. 6a,b). Taken together, SiT-Cas12a effectors, based on *As*Cas12a, facilitated orthogonal transcriptional control and gene editing simply by altering spacer length.

Discussion

In contrast to most CRISPR-Cas gene editing expression strategies, where Cas enzymes and guide RNAs are expressed from distinct promoters, in our platform a single Pol II promoter expresses a single transcript harboring both Cas12a and a CRISPR array. Consequently, the ratio of Cas12a and CRISPR array is fixed, which introduces distinct disadvantages and advantages. A potential disadvantage is that the fixed ratio of the two components may be suboptimal, especially in expression-limited conditions. Such a theoretical disadvantage could be overcome by either increasing SiT-Cas12a transcript abundance or encoding multiple crRNAs per gene. On the other hand, the advantage of the fixed ratio is that it enables tight control of both Cas12a and crRNA expression, facilitating conditional and inducible genome engineering applications. Last, as the extent of genome editing reflects the underlying expression of the components, derived from either the intrinsic variability among different cell types or the cellular environment (for example, Cre recombinase, rtTA, tTA in this work), the SiT-Cas12a platform empowers further applications in DNA writing, molecular recording and synthetic biology[34–36].

Considering the mean natural length of protein-coding transcripts found within mammalian cells (13.5 kb), the potential for

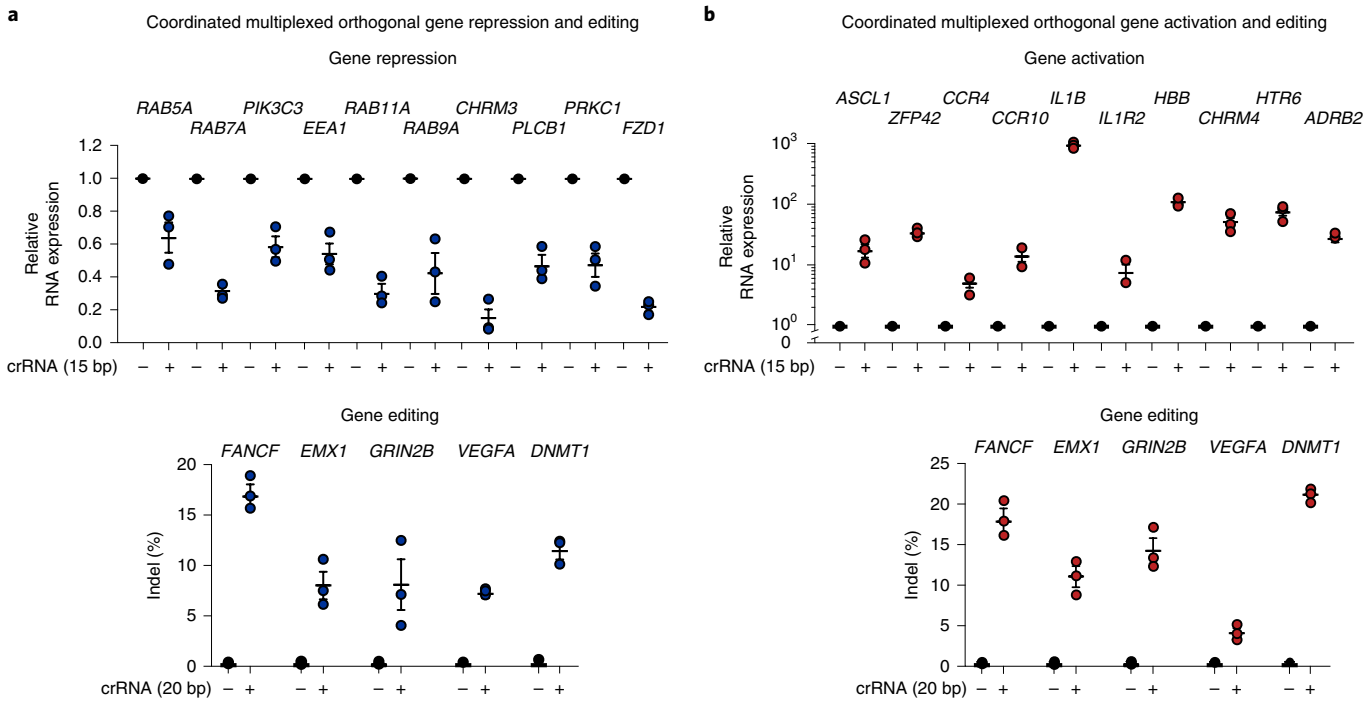

Fig. 6 | Multiplexed orthogonal gene editing and transcriptional activation. a, Quantification of relative RNA expression and gene editing efficiencies in cells expressing a large CRISPR array (25 crRNAs) harboring both short (15 bp) and long (20 bp) spacers using SiT-Cas12a-[Repr]. Values represent mean ± s.e.m., $n = 3$ independent experiments. **b**, Quantification of relative RNA expression and gene editing efficiencies in cells expressing a large CRISPR array (25 crRNAs) harboring both short (15 bp) and long (20 bp) using SiT-Cas12a-[Activ]. Values represent mean ± s.e.m., $n = 3$ independent experiments.

expressing multiple crRNAs in the SiT-Cas12a context is profound. In the future this could theoretically be used to enable massively multiplexed expression of hundreds to thousands of independent crRNAs, opening up avenues for large-scale genome engineering efforts[30,37]. While SiT-Cas12a offers a seemingly unlimited potential for crRNA expression, longer pre-crRNA transcripts will inevitably be challenging to synthesize and clone. Furthermore, DRs and spacers containing complementary sequences that could generate complex secondary RNA structures affecting the maturation of crRNAs in cells[38,39]. Consequently, complementary regions in pre-crRNA must be considered to improve crRNA maturation. Future work overcoming these limitations will open up numerous applications for highly multiplexed genome engineering.

In the control of cell behavior, genetic elements act together. Deciphering such complexity requires fine modulation of multiple genetic elements. Simultaneous encoding of dozens of crRNAs in a single plasmid simplifies both guide RNA testing and validation of gene function. Inspired by design principles that embody biological efficiency, our genome engineering platform harnesses the full potential of the Cas12a enzyme, providing an easy and customizable way to allow highly multiplexed gene editing and transcriptional control making it possible, in the future, to systematically interrogate complex genetic interactions and cellular behaviors.

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/s41592-019-0508-6.

Received: 23 January 2019; Accepted: 7 June 2019;
Published online: 12 August 2019

References

1. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
2. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
3. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
4. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
5. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
6. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
7. Maeder, M. L. et al. CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
8. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
9. Cheng, A. W. et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).
10. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
11. Zalatan, J. G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
12. Boettcher, M. et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170–178 (2018).
13. Dahlman, J. E. et al. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159–1161 (2015).
14. Kiani, S. et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Methods* **12**, 1051–1054 (2015).
15. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
16. Leenay, R. T. et al. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* **62**, 137–147 (2016).
17. Rogers, J. K. & Church, G. M. Multiplexed engineering in biology. *Trends Biotechnol.* **34**, 198–206 (2016).

18. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl Acad. Sci. USA* **112**, 3570–3575 (2015).
19. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**, 698–710 (2014).
20. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Chemical roots of biological evolution: the origins of life as a process of development of autonomous functional systems. *Open Biol.* **7**, 170050 (2017).
21. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
22. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
23. Swarts, D. C., van der Oost, J. & Jinek, M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol. Cell* **66**, 221–233.e224 (2017).
24. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651 (2018).
25. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
26. Zhong, G., Wang, H., Li, Y., Tran, M. H. & Farzan, M. Cpf1 proteins excise CRISPR RNAs from mRNA transcripts in mammalian cells. *Nat. Chem. Biol.* **13**, 839–841 (2017).
27. Tak, Y. E. et al. Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* **14**, 1163–1166 (2017).
28. Arimbasseri, A. G., Rijal, K. & Maraia, R. J. Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. *Transcription* **5**, e27639 (2014).
29. Wilusz, J. E. et al. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392–2407 (2012).
30. Meyers, R. A. *Encyclopedia of Molecular Cell Biology and Molecular Medicine* 2nd edn (Wiley-VCH, 2004).
31. Huntley, S. et al. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* **16**, 669–677 (2006).
32. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
33. Singh, D. et al. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl Acad. Sci. USA* **115**, 5444–5449 (2018).
34. Schmidt, F. & Platt, R. J. Applications of CRISPR-Cas for synthetic biology and genetic recording. *Curr. Opin. Syst. Biol.* **5**, 9–15 (2017).
35. Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).
36. Farzadfard, F. & Lu, T. K. Emerging applications for DNA writers and molecular recorders. *Sci.* **361**, 870–875 (2018).
37. Tennyson, C. N., Klamut, H. J. & Worton, R. G. The human dystrophin gene requires 16 h to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **9**, 184–190 (1995).
38. Zoephel, J. & Randau, L. RNA-Seq analyses reveal CRISPR RNA processing and regulation patterns. *Biochem. Soc. Trans.* **41**, 1459–1463 (2013).
39. Liao, C., Slotkowski, R. A., Achmedov, T. & Beisel, C. L. The Francisella novicida Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. *RNA Biol.* **16**, 1–9 (2018).

Acknowledgements

We thank the entire Platt Laboratory for productive discussions; C. Beisel, E. Burcklen, K. Eschbach, I. Nissen and M. Kohler from the Genomics Facility Basel for assistance in Illumina sequencing. This project was supported by the Swiss National Science Foundation, ETH domain Personalized Health and Related Technologies, Brain and Behavior Research Foundation, the National Centres of Competence—Molecular Systems Engineering and ETH Zurich. C.C.C. was supported by the ETH Postdoctoral Fellowship.

Author contributions

C.C.C., N.R.W. and R.J.P. conceived and designed the experiments. C.C.C. and N.R.W. performed the experiments and analyzed the data. A.J.S analyzed the deep sequencing data. D.I. analyzed the RNA-seq data. C.C.C., N.R.W. and R.J.P. wrote the manuscript. All authors reviewed the paper and provided comments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0508-6.

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

Correspondence and requests for materials should be addressed to R.J.P.

Peer review information: Nicole Rusk was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mammalian cell culture. A HEK 293T cell line (SIGMA-Aldrich) was maintained in Dulbecco's modified Eagle's Medium (DMEM) (SIGMA-Aldrich) supplemented with 10% FBS (HyClone) at 37 °C with 5% CO_2.

Transient transfection. HEK 293T cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transient transfections were performed using either 0.6 or 1 µg of plasmid DNA per well (24-well plate).

Plasmids. All plasmids were generated using restriction enzyme-based cloning, Golden Gate Cloning or Gibson Assembly. SiT-Cas12a constructs for multiplexed genome editing were generated by replacing the U6-single-guide RNA from pY026 (Addgene, 84741) with the Triplex sequence and two DR sequences, separated by two BsmBI restriction sites. SiT-Cas12a constructs for transcriptional control were generated by replacing the U6-gRNA and Cas9-Puromycin resistance from lentiCRISPR v.2 (Addgene, 52961) with AsCas12a, Triplex sequence and two DR sequences, separated by two SapI restriction sites. Constructs for production of lentivirus were generated by replacing the U6-gRNA, Cas9-Puromycin and WPRE element from lentiCRISPR v.2 (Addgene, 52961) with a CMV promoter together with *EGFP* or AsCas12a, Puromycin, Triplex sequence, two DR sequences, separated by two BsmBI restriction sites and/or a WPRE element and/or a polyadenylation sequence in inverted orientation. Depending on the experiment setup, either a CMV or an EF1a promoter was used. The ddCas12a and rdCas12a mutants were generated by site-directed mutagenesis through the substitution of either E993A or H800A, respectively. Transcriptional control elements were amplified from gene fragments (IDT) and integrated 3′ of AsCas12a using a BamHI restriction site. Assembly of CRISPR arrays was performed by Golden Gate Cloning via BsmBI or SapI using either annealed oligonucleotides (IDT) for single spacers or small arrays (2–4 spacers), or gene fragments (IDT) with flanking Type IIS restriction sites for medium (10 spacers) and large (20–25 spacers) arrays. The tetracycline/doxycycline inducible plasmid was constructed by exchanging the constitutive promoter of SiT-Cas12a with tetracycline response element consisting of five repeats of bacterial TetO sequence upstream of a minCMV promoter by Gibson assembly. SiT-Cas12a-[Cond] was generated by adding a LSL cassette between the promoter and Cas12a by Gibson assembly. DNA and spacer target sequences are listed in the Supplementary Information. Relevant plasmids used in this study will be deposited to AddGene.

Large CRISPR array assembly. Large CRISPR arrays containing 10, 20 or 25 spacer sequences were assembled based on previous procedures[40]. Double-stranded DNA fragments encoding for 4–6 crRNAs were purchased from IDT Technologies (gBlock or custom gene synthesis services) and amplified by primers containing type IIS restriction site (BsmBI or SapI). Type IIS recognition sites were designed to allow cleavage, at different positions, on the DR sequence. This procedure enables both generation of non-identical 5′ or 3′ ends and reduction of repetitive elements (in our case, DR) that prohibit the chemical synthesis of double-stranded crRNAs fragments by IDT Technologies. Overhangs generated by type IIS restriction enzyme digestion were intended to be complementary, enabling assembly of a large CRISPR array with defined directionality. DNA fragments encoding for Cas12a crRNAs and the vector backbone were assembled by standard Golden Gate Cloning. In brief, simultaneous type II enzymatic digestion using either BsmBI (Thermo Scientific) or SapI (Thermo Scientific) and ligation using T7 DNA ligase (NEB) in 1× T4 Ligase buffer (NEB) reaction was performed. The reaction was incubated for 30 cycles (37 °C for 5 min; 16 °C for 5 min) followed by 55 °C for 10 min. Ligation reaction was transformed into chemical competent *E. coli* Stbl3 bacteria strain to avoid potential plasmid recombination. Bacteria cells were plated on agar plates supplemented with ampicillin (100 mg l^{-1}). Single colonies were cultivated overnight in liquid LB supplemented with ampicillin (100 mg l^{-1}) and DNA were isolated. Correct assembly of CRISPR array was verified by SANGER sequencing.

Inducible gene editing. HEK 293T cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transient transfections were performed using a total of 0.6 µg (0.3 µg of SiT-Cas12a-[Ind] and 0.3 µg of either rtTA or tTA-Advanced (Clontech) plasmid DNA per well (24-well plate) and Doxycycline (Sigma) was added (final concentrations, 0, 0.01, 0.10 and 1.00 µg ml^{-1}). Quantification of gene editing was performed 72 h post-transfection.

Quantification of gene editing. Roughly 5×10^4 HEK 293T cells per well were seeded in 24-well-plates and transfected with 1 µg of DNA plasmid. Then, 72 h post-transfection, cells were gathered and lysed in QE buffer (1 mM $CaCl_2$, 3 mM $MgCl_2$, 1 mM EDTA, 1% Triton X-100, 10 mM Tris pH 7.5, 0.2 mg ml^{-1} Proteinase K) using the following temperature conditions: 65 °C (15 min), 68 °C (15 min) and 98 °C (10 min). Genomic DNA was used as a template for PCR-based amplification of targeted genomic regions using Phusion flash polymerase (Thermo Scientific) and primers specific for each target site (Supplementary Information (Indel analyses primer list)). PCR amplicons were purified using the Zymo PCR purification Kit (Zymo Research) and quantified using Nanodrop 3000 (Thermo Scientific).

To generate heteroduplex DNA fragments, 250 ng purified PCR amplicons were mixed with 10× annealing buffer (500 mM $NaCl_2$, 100 mM $MgCl_2$, 100 mM Tris-HCl) and incubated for 10 min at 95 °C followed by ramping 95 °C to 85 °C with 1.34 °C s^{-1}, 85 °C to 75 °C with 0.2 °C s^{-1}, 75 °C to 65 °C with 0.2 °C s^{-1}, 65 °C to 55 °C with 0.2 °C s^{-1}, 55 °C to 45 °C with 0.2 °C s^{-1}, 45 °C to 35 °C with 0.2 °C s^{-1} and 35 °C to 25 °C with 0.2 °C s^{-1}. Heteroduplex DNA were treated with Surveyor enhancer and Surveyor nuclease from Surveyor Mutation Detection Kit (IDT) according to the manufacturer's protocol and separated on 2% E-Gel (Thermo Scientific). Separated cleavage products were imaged using Gel DOC EZ imager (Biorad) and quantified using Image Lab software (Biorad). The percentage of heteroduplex DNA formation was quantified using the formula: $(1-(1-(b/(a+b))\times 0.5))\times 100$ with a being equivalent to the integrated intensity of uncut DNA fragments and b being equivalent to the sum of the integrated intensity of all cleavage products.

FACS analyses. Roughly 1.2×10^5 HEK 293T cells were plated in 24-well-plate and transfected with 600 ng plasmid. Then, 48 h after transfection, cells were split and 120 h post-transfection, cells were gathered (using PBS supplemented with 0.5 mM EDTA) and fixed in 1.8% PFA (Electron Microscopy Sciences). Fixed cells were stained using conjugated antibodies CD166 conjugated PE (Miltenyi Biotec 130-106-575, flow cytometry dilution (FC) 1:11); CD47 conjugated FITC (Miltenyi Biotec 130-101-344, FC 1:11); CD97 conjugated APC-Vio770 (Miltenyi Biotec 130-105-526, FC 1:11). Cytometric analysis was conducted using five-color Fortessa (BD). Data from 10,000 events was collected and analyzed by FlowJo (BD). Cell debris was removed by SSC-A/FSC-A gating and fluorescent intensity (FITC, PE or APC-Vio770) was measured. Cells transfected with empty plasmids and stained were used as positive controls. Non-transfected cells were used as a negative control. Percentage of negative cells was calculated by gating on the positive control cell population. Quantification of EGFP-positive cells and EGFP expression of infected cells (72 h post infection) were performed by gathered HEK 293T cells (using PBS supplemented with 0.5 mM EDTA) and fixed in 1.8% PFA (Electron Microscopy Sciences). Data from 10,000 events was collected and analyzed by FlowJo (BD). Cell debris was removed by SSC-A/FSC-A gating and fluorescent intensity (EGFP) was measured. Non-infected cells were used as a negative control. Both percentage of positive cells and EGFP expression (median fluorescence intensity) were calculated.

Quantification of mRNA expression. Gene expression analyses were conducted 48 h after transfection according to a previously established protocol[3]. In brief, RNA was isolated using Quick RNA Miniprep Plus kit (Zymo Research), followed by reverse transcription of 500 ng RNA using Qscript cDNA supermix (Quantabio). A quantitative PCR (qPCR) reaction was performed using Fast Plus EvaGreen qPCR master mix (Biotium) according to the manufacturer's protocol (primers used are indicated in the Supplementary Information (qPCR primer list)). Quantification of RNA expression was normalized based on expression of glyceraldehyde 3-phosphate dehydrogenase and calculated using $\Delta\Delta Ct$[41]. For samples with a nearly undetectable amount of mRNA and Ct values exceeding 45, an arbitrary cycle number of 45 was assigned.

Small RNA-seq library preparation. Roughty 1.2×10^7 HEK 293T cells were plated in a 150 mm dish (Thermo Scientific) and transfected with 30 µg of plasmid. After 48 h, purification of small RNAs was conducted using Quick RNA Miniprep Plus kit (Zymo Research) according to the manufacturer's instructions. Small RNA library preparation was performed in line with a previous published protocol[42]. In brief, residual genomic DNA was removed by DNase I digestion. Then, 20 µg of small RNAs were dissolved in 39.5 µl water and denatured for 5 min at 65 °C. After cooling on ice for 5 min, 5 µl 10× DNase I buffer including $MgCl_2$ (NEB), 10 units of Superase-In RNase Inhibitor (Thermo Scientific) and 5 units of DNase I (NEB) were added and incubated for 45 min at 37 °C. Purification of this pure small RNA fraction was conducted using Quick RNA Miniprep Plus kit (Zymo Research) and quantified using Nanodrop 3000 (Thermo Scientific). To capture 5′ phosphorylated crRNA, an additional 5′ phosphorylation step was performed. RNA samples, treated with 10 units of DNase I (NEB), were denatured for 2 min at 90 °C and stored on ice for 5 min. Subsequently, 20 units of T4 PNK (Thermo Scientific) together with 10 units of Superase-In RNase Inhibitor (Thermo Scientific) were added in the presence of T4 PNK buffer (Thermo Scientific) to a final volume of 50 µl. After 6 h of incubation at 37 °C, an additional 10 units of T4 PNK (Thermo Scientific) and 10 mM ATP (2 mM final concentration) (Thermo Scientific) were added to the samples and incubated for 1 h at 37 °C. Purification of 5′ phosphorylated RNA fractions were performed using Quick RNA Miniprep Plus kit (Zymo Research) and quantified. As T4 PNK treatment enriched transcript with 5′ pyrophosphates, thereby interfering with the following step of small RNA library preparations, 5′ pyrophosphates were removed by Tobacco Acid Pyrophosphatase (TAP) treatment (Epicentre). In a total volume of 20 µl, 10 units of TAP, 2 µl TAP buffer, 10 units of Superase-In RNase Inhibitor (Thermo Scientific) and 10 µg of PNK-treated small RNA were mixed and incubated for 1 h at 37 °C and purified using Quick RNA Miniprep Plus kit (Zymo Research). These enriched pre-crRNA transcripts were used as template for the preparation of the small RNA library. Library preparation was performed using NEBNext multiplex small RNA library prep set (set 1) (NEB), separated on a 6% polyacrylamide gel and purified using

SpinX columns (Costar). This step ensured enrichment of mature crRNAs (<60 nt). Five million reads were sequenced for each sample using Illumina NextSeq 500 (Illumina) and analyzed as described below.

Mature crRNA quantification. Adapter sequences (described in NEBNext multiplex small RNA library prep set (set 1)) were clipped from sequencing reads using CutAdapt v.1.10 (ref. [43]). Only reads ≥10 nt were retained. Reads were mapped to the reference sequence of the entire crRNA array, using Bowtie v.1.2.2 (ref. [44]), with the parameters '—norc -l 28 -n 2 -m 1 —best —strata'. For the quantification of mature crRNAs, a BED file containing the coordinates of each crRNA within the array was created and used to extract the number of reads mapping onto each crRNA using the intersectBed utility from BEDTools v.2.27.0 (ref. [45]). Raw counts were then normalized to the number of non-crRNA reads mapping on human small nucleolar RNAs through the following formula: $N_i = (c_i/S) \times 10^6$; where N_i and c_i are, respectively, the normalized count and the raw read count for crRNA i, and S is the count of reads mapping on human snoRNAs, a small RNA population largely recognized for performing housekeeping function and used to normalize human microRNA expression[46,47]. Finally, a fold change over control condition was calculated to quantify changes across distinct experiments.

Deep sequencing-based *CD47* editing analyses. Analyses of deep sequencing data derived from HEK 293T cells stably expressing SiT-Cas12a harboring ten distinct crRNAs were performed using the CRISPResso2 tool[48], the *CD47* edited region was amplified, size was selected between 200 and 500 bp and then sequenced. Analysis was carried out using the CRISPResso2 tool[48] with the following parameters: -w 10 –cleavage offset 1 –S 20; -w 10 –cleavage offset 1 –g.

Statistics. Unless otherwise noted, experiments in this study were performed using three independent biological replicates. Tests for determination of statistical significance were not implemented.

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

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The datasets generated during the current study are available in the NCBI Sequence Read Archive (BioProject ID PRJNA530879). The data sets generated and/or analysed during the current study are available from the corresponding author upon reasonable request. The custom scripts used for the described data analysis areavailable on the Platt Laboratory website (http://www.platt.ethz.ch).

References

40. Liao, C. et al. One-step assembly of large CRISPR arrays enables multi-functional targeting and reveals constraints on array design. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/312421v1 (2018).
41. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).
42. Heidrich, N., Dugar, G., Vogel, J., Sharma, C. M. & Investigating CRISPR, R. N. A. Biogenesis and function using RNA-seq. *Methods Mol. Biol.* **1311**, 1–21 (2015).
43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10 (2011).
44. 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).
45. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12.11–11.12.34 (2014).
46. Roa, W. et al. Identification of a new microRNA expression profile as a potential cancer screening tool. *Clin. Invest. Med.* **33**, E124 (2010).
47. Hui, A. B. et al. Comprehensive microRNA profiling for head and neck squamous cell carcinomas. *Clin. Cancer Res.* **16**, 1129–1139 (2010).
48. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

nature research

Corresponding author(s): Randall J. Platt

Last updated by author(s): May 10, 2019

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	qPCR data were collected using Light Cycler 96 Application Software Version 1.1
Data analysis	FlowJo version 10.3, Image Lab 5.2.1, CRISPResso2, CutAdapt v1.10, intersectBed utility from BEDTools v2.27.0, Bowtie v1.2.2, Genious v10.1.2

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data were submitted in public repositories (NCBI sequence read archive BioProject ID PRJNA530879). Data will be provided upon reasonable request

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

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

Sample size	no sample size calculation was performed. Based on previous publications, experiments in cell lines were performed in triplicates n = 3, unless otherwise noted.
Data exclusions	no data was excluded
Replication	Unless otherwise noted, experiments were performed using three independent biological replicates (indicated in Figure legends).
Randomization	due to the small sample, randomization was not relevant for this study
Blinding	blinding was not relevant for this study due to the limitation on in vitro experiments

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CD166 conjugated PE (Miltenyi Biotec 130-106-575, flow cytometry dilution (FC) 1:11); CD47 conjugated FITC (Miltenyi Biotec 130-101-344, FC 1:11); CD97 conjugated APC-Vio770 (Miltenyi Biotec 130-105-526, FC 1:11).
Validation	All antibodies are commercially available. CD166 conjugated PE was validated against CD166 by the manufacturer and used in several publications (e.g. Chitteti, B.R. (2014)). CD47 conjugated FITC was validated against CD47 by the manufacturer and used in several publications (e.g. Demeure,C.E. (2000)). CD97 conjugated APC-Vio770 was validated against CD97 by the manufacturer.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (ATCC)
Authentication	none of the cell line were authentichated
Mycoplasma contamination	cell line was negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	Exclusively HEK293T cells were used in this study.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were detached using TRIS/EDTA solution and fixed cells were stained with the following antibodies: CD166 conjugated PE (Miltenyi Biotec 130-106-575, flow cytometry dilution (FC) 1:11); CD47 conjugated FITC (Miltenyi Biotec 130-101-344, FC 1:11); CD97 conjugated APC-Vio770 (Miltenyi Biotec 130-105-526, FC 1:11).
Instrument	5 color fortessa BD
Software	FlowJo version 10.3
Cell population abundance	100%
Gating strategy	cells were gated by FSC/SSC followed by PE, FITC, APC, intensities

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.