

ARTICLE OPEN



gDesigner: computational design of synthetic gRNAs for Cas12a-based transcriptional repression in mammalian cells

Michael A. Crone^{1,2,3}, James T. MacDonald¹, Paul S. Freemont^{1,2,3} and Velia Siciliano^{1,4}

Synthetic networks require complex intertwined genetic regulation often relying on transcriptional activation or repression of target genes. CRISPRi-based transcription factors facilitate the programmable modulation of endogenous or synthetic promoter activity and the process can be optimised by using software to select appropriate gRNAs and limit non-specific gene modulation. Here, we develop a computational software pipeline, gDesigner, that enables the automated selection of orthogonal gRNAs with minimized off-target effects and promoter crosstalk. We next engineered a *Lachnospiraceae* bacterium Cas12a (dLbCas12a)-based repression system that downregulates target gene expression by means of steric hindrance of the cognate promoter. Finally, we generated a library of orthogonal synthetic dCas12a-repressed promoters and experimentally demonstrated it in HEK293FT, U2OS and H1299 cells lines. Our system expands the toolkit of mammalian synthetic promoters with a new complementary and orthogonal CRISPRi-based system, ultimately enabling the design of synthetic promoter libraries for multiplex gene perturbation that facilitate the understanding of complex cellular phenotypes.

npj Systems Biology and Applications (2022)8:34; <https://doi.org/10.1038/s41540-022-00241-w>

INTRODUCTION

Mammalian synthetic biology leverages characterized genetic regulatory parts to perform input-output processing algorithms, enabling novel biological functions in eukaryotic cells^{1–3} and tackling fundamental questions underlying biological processes^{4–7}. To achieve a comprehensive understanding of cellular phenotypes and functions, new genetic modules (transcription factors, post-transcriptional and translational modulators) are required tools for multi-layered regulation of synthetic networks^{8–12}.

Orthogonal transcriptional repressors provide the means to design universal logic circuits, enabling the assembly of tunable and composable devices of increasing complexity¹³. Synthetic transcriptional repressors such as TALE¹⁴ or zinc-finger proteins^{15,16} fused to effector domains (e.g., KRAB) efficiently down-modulate gene expression, but are limited by slow temporal gene reactivation¹⁷. Alternatively, transcriptional repressors that directly bind to operator sequences in proximity to the transcription start site and inhibit transcription by steric hindrance have been successfully demonstrated^{10,11}.

Catalytically dead Cas9 from *Streptococcus pyogenes* (dSpCas9) as a transcriptional tool to activate¹⁸ or repress¹⁹ endogenous and exogenous gene expression, allows transcriptional modulation in a multiplexed and orthogonal fashion, with ground-breaking potential in the biomedical and industrial fields. By means of single guide RNAs (gRNAs) to potentially target nearly any DNA sequence, CRISPR-Cas9 based transcriptional regulation also overcomes the issue of metabolic burden imposed on cells by circuits that simultaneously express a large number of transcription factors to achieve sophisticated regulation^{20–22}. However, the common gRNA scaffold sequence prevents the simultaneous use of dCas9 as activator and repressor, limiting possible genetic network designs, and thus hampering their applicability in sophisticated circuits development.

It has been shown that CRISPR-Cas12a nucleases offer several additional features and advantages compared to SpCas9 such as shorter guide RNAs (CRISPR RNAs-crRNAs), RNase processing of multiple crRNAs from a single transcript^{23,24} and the ability to target T-rich protospacer-adjacent motif (PAM) in contrast to the G-rich PAM used by Cas9^{25–27}. Transcriptional activators based on catalytically inactive *Lachnospiraceae* bacterium Cas12a (dLbCas12a) fused to synthetic activator VPR, and transcriptional repressors deriving from *Acidaminococcus* sp. BV3L6 (dAs-Cas12a) fused to Krab repressor domain, have been recently demonstrated^{28,29} in mammalian cells. Moreover, the development of software algorithms to select gRNA with minimized off-target sites by performing exhaustive searches against genomic sequences, has improved the capability to engineer libraries of devices^{30–33}, paving the way for genome-wide gene-perturbation screens in mammalian cells, as well as sophisticated design of multilayered networks.

Here, we have developed gDesigner, a software pipeline to select appropriate gRNA sequences and associated synthetic promoters with embedded cognate responsive elements for trans-regulation by dLbCas12a protein (Fig. 1). The software takes an initial pool of generated sequences, and automatically selects the largest orthogonal subset that present minimal similarity, low potential for cross-interactions, and minimal off-targets in the host organism genome.

We have then experimentally demonstrated dLbCas12a-based synthetic transcriptional repressors that inhibit gene expression solely by steric hindrance in different mammalian cell lines (Fig. 1), expanding the CRISPR based tools for genetic cross-regulation and allowing combination of dCas9/dCas12a activators and repressors. We have made this pipeline available at <https://github.com/jmacdona/gDesigner>.

¹Section of Structural and Synthetic Biology, Department of Infectious Disease, Imperial College London, London, United Kingdom. ²UK Dementia Research Institute Centre for Care Research and Technology, Imperial College London, London, United Kingdom. ³London Biofoundry, Imperial College Translation and Innovation Hub, White City Campus, 84 Wood Lane, London, United Kingdom. ⁴Istituto Italiano di Tecnologia IIT, Department of Synthetic and Systems Biology for Biomedicine, Genoa, Italy. ✉email: j.macdonald@imperial.ac.uk; p.freemont@imperial.ac.uk; velia.siciliano@iit.it

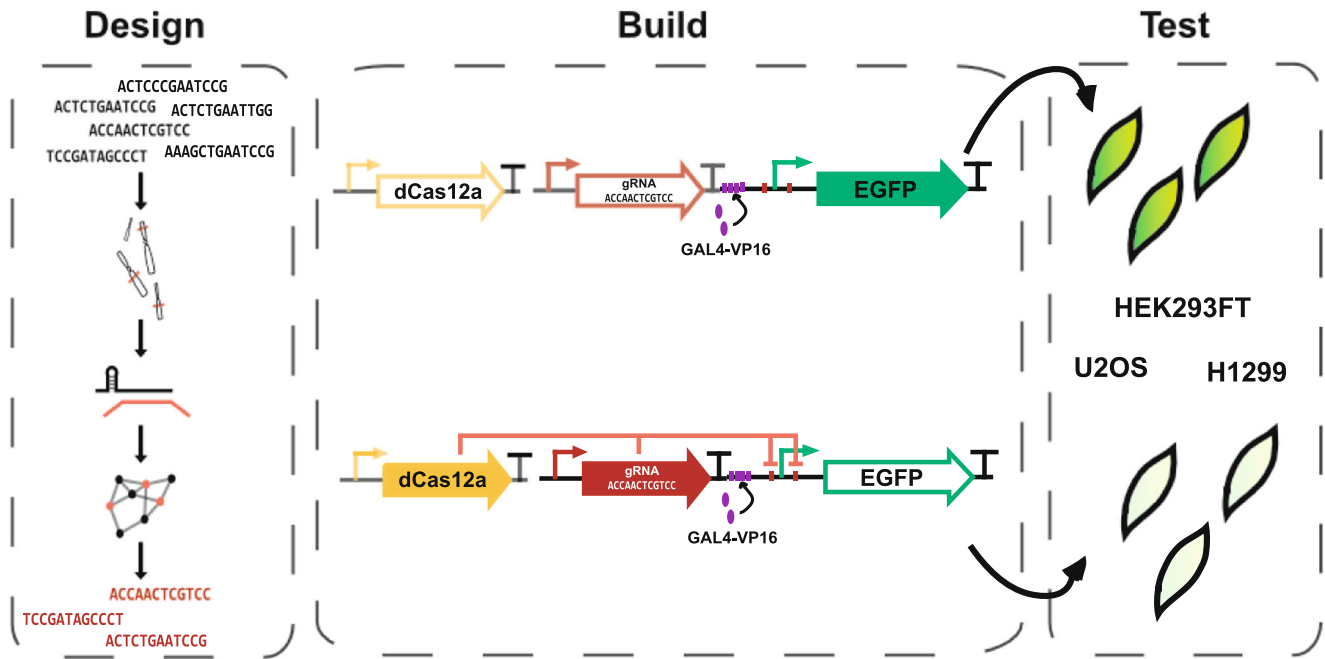


Fig. 1 Framework of the study. gDesigner enables the automated selection of orthogonal gRNAs with minimized off-target effects and promoter crosstalk. We generated a library of orthogonal synthetic dCas12a-repressed promoters that downregulates target gene expression by means of steric hindrance of the cognate promoter. We tested the library in HEK293FT, U2OS and H1299 cells lines.

RESULTS AND DISCUSSION

Generation of synthetic gRNA sequences

An initial pool of synthetic gRNA sequences were generated using R2oDNA Designer, a simulated annealing-based algorithm, previously developed as a general and flexible method to design orthogonal spacer sequences with user-defined constraints^{34,35}. Specifically, we designed random sequences according to the following criteria: i) minimal repeats or inverted repeat sequences, ii) minimal BLAST hits to the human genome and iii) no sequences matching eukaryotic transcription factor binding site motifs (according to the HOCOMOCO database³⁶), or some common restriction site sequences.

This initial pool of gRNA sequences was then fed into our new synthetic gRNA-screening pipeline, gDesigner. The pipeline automatically eliminates sequences with potential off-target hits in specified genomes³⁷, and spacer sequences with estimated hybrid gRNA-spacer:cognate-DNA binding site affinities outside a defined range of values^{38,39}. In the last step, gDesigner retrieves an orthogonal set of gRNA sequences with no pairwise cross-interactions between different gRNA species below a specified minimum free energy, using a simple graph-theoretic algorithm to find the maximum independent set (Fig. 2a, Supplementary Table 2).

Optimization of binding site orientation to maximize repression

To test synthetic gRNA sequences generated by gDesigner, we engineered a promoter architecture framework that allows easy drop-in of synthetic gRNA binding sites and used single point mutations to catalytically deactivate a mammalian codon-optimized Cas12a from *Lachnospiraceae bacterium ND2006* (hdLbCas12a). Repressible promoters are composed of Gal4-VP16-responsive 4x UAS operator sequences, and gRNA binding sites upstream and flanking a super minimal CMV (smCMV) promoter, to regulate the expression of enhanced green fluorescent protein, EGFP (Fig. 2b). Binding of the Cas12a-gRNA complex to the cognate promoter prevents Gal4-VP16-induced transcriptional activation (Fig. 2c) computing a Boolean NAND

logic gate. Similar to previous studies¹⁹ we first investigated the correlation between gRNA binding sites flanking the TATA box directionality and repression efficiency. We designed and tested four promoter configurations with gRNAs facing inwards, outwards, 5', or 3' direction in HEK293FT cells (Supplementary Fig. 1a). We observed stronger hdLbCas12a repression either when gRNA binding sites face outwards the TATA box, or in 5' direction (Supplementary Fig. 1b, Fig. 3a, b). For the next set of experiments, we placed gRNA binding sites to face outwards relative to the TATA box.

Testing orthogonality of repression of generated synthetic gRNAs and minimal promoters

To maximize the activation-repression ratio we tested increasing levels of Gal4-VP16 along with fixed concentration of the other components of the repressor and used the optimized condition (GAL4-VP16:Cas12a:gRNA ratio: 1:1, 5:10) for subsequent tests (Supplementary Fig. 2). Next, we tested hdLbCas12a-dependent repression in HEK293FT cells, using a library of nine computationally designed synthetic gRNA/promoters which were co-transfected along the GAL4-VP16 and in the presence or absence of gRNA (Fig. 3c, Supplementary Table 1). Also, by co-transfecting all combination of reporters and transcription factors into HEK293FT cells (9×9 matrix), we show that our repressors are largely orthogonal (Fig. 3c, Supplementary Fig. 3). Promoters/gRNAs that exhibited different repression activity in HEK293FT were also tested in U2OS (g1-g4) (Fig. 3d, Supplementary Fig. 5) and H1299 (g4) cell lines (Fig. 3e). Interestingly, U2OS cells show similar repression activity to that observed in HEK293FT cells (repression induced by g2 and g1 > g4 > g3) albeit with overall different fold of activation (Supplementary Fig. 5) and repression (Fig. 3b–e, Supplementary Figs. 3 and 4), likely due to cell-dependent differences in the pool of available transcription factors and co-factors⁴⁰.

Our library provides a further expansion to the toolbox of transcriptional regulators that can be layered in complex networks. Minimal cross talk in synthetic promoters is a critical feature of genetic circuits. We demonstrate that gDesigner allows the

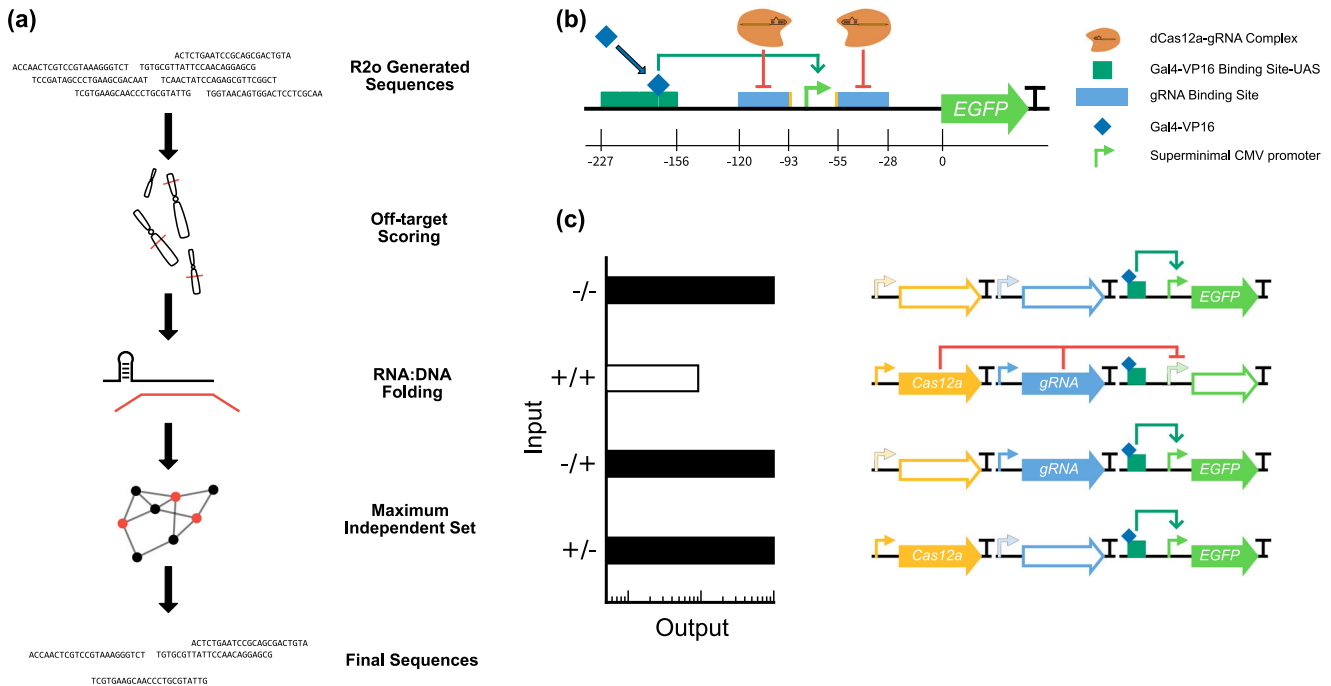


Fig. 2 **gDesigner pipeline, promoter architecture and system dynamics.** **a** Initial sequences are generated using the R2oDNA Designer method. The sequences are subsequently filtered through the gDesigner pipeline (<https://github.com/jmacdona/gDesigner>) to remove sequences with predicted off-target sites, binding affinities outside a specified range, selecting a final set of orthogonal sequences with minimal predicted cross-interactions. **b** Schematics of the promoter architecture. Gal4-VP16 activates the expression of EGFP, while the binding of Cas12a-gRNA controls its repression. 23 bp gRNA-binding sites with their PAM sequences (5'-TTTN-3') flank the smCMV promoter. The 5' binding site is directly adjacent to the TATAA box, and 36 bp downstream the UAS. The 3' binding site is placed 10 bp downstream of the transcription start site (TSS) and 105 bp downstream of UAS. **c** Schematics of the predicted NAND logic gate. EGFP expression is repressed in the presence of hLbCas12a/gRNA-promoter complex.

design of a library of nine gRNA/promoters pairs that show extensive orthogonality. In contrast to recently developed dCas12a based repressors that downregulated endogenous gene expression via fusion to Krab domain, our synthetic repressors effectively repress solely by steric hindrance, potentially overcoming issues related to slow temporal reactivation¹⁷.

A disadvantage to the design of complex regulatory networks is the size limit of the DNA for transient expression or integration, as well as the burden that multiple transcription factors in layered circuits impose to the host cells. CRISPR based transcriptional regulation can overcome this limitation since any circuit topology can be encoded by a single transcriptional repressor along with multiple gRNAs to simultaneously downregulate multiple genes. Here, we provide a proof-of-principle study that dLbCas12a can down-modulate gene expression in a similar way to that of dCas9 in human cells¹⁹. This should enable the engineering of multi-layered networks of increasing complexity. Taken together, our method should facilitate a broad range of applications for biomedical interventions in human cells.

METHOD

Synthetic gRNA design and binding site selection

The new software pipeline implemented as a Python module, was developed to rapidly screen synthetic orthogonal gRNA sequences that conform to a given set of required constraints. Source code and installation instructions are available at <https://github.com/jmacdona/gDesigner>. The user inputs an initial pool of FASTA-formatted candidate synthetic gRNA sequences (here, we used the R2oDNA Designer simulated annealing algorithm but sequences from any source are possible), a user-defined constraints/settings file, and a paths file to instruct the pipeline where software dependencies are installed.

This pipeline was composed of four steps: (i) initial sequence pool generation using the R2oDNA Designer algorithm, (ii) off-target screening using OFF-FINDER, (iii) RNA:DNA hybrid free energy screening using MELTING (iv) calculation of all-against-all pairwise minimum free energies (MFE) using PairFold⁴¹ and determination of the largest orthogonal set of gRNA sequences by retrieving the maximum independent set. The initial pool of DNA sequences (step (i)) was generated using the previously published simulated annealing-based algorithm, R2oDNA Designer, which eliminates repeats above a defined length, inverted repeats above a defined length and other disallowed sequences³⁴. Sequences for LbCas12a were generated in the form of a repeat-spacer (TAATTCTAC-TAAGTGTAGATN₂₃). Certain restriction sites, promoter sequence motifs, and eukaryotic transcription factor binding site motifs from the HOCOMOCO database³⁶ were disallowed. Sequences that generated BLAST hits against the *E. coli* BL21, DH10 β and W3110, *B. subtilis*, *S. cerevisiae* and *H. sapiens* genomes as well a catalog of IGEN parts were excluded below a certain e-value cutoff. The variable region (Nx) of the generated sequences were designed to have a 60 °C DNA melting temperature.

Cas-OffFinder³⁷ was used to search both Human and *E. coli* genomes for potential off-target sequences of up to 7 mismatches with the PAM sequence 5'-TTTN-3'. A previously described off-target scoring scheme for Cas9³⁰ was adapted to score for Cas12a by using the mismatch matrix from a recent publication²³. The scoring scheme calculates the score of each individual off-target sequence (Eq. (1)) and then sums the scores together to give a final quantitative score (Eq. (2)). Sequences with scores below 0.8 were eliminated.

Equation 1 (single mismatch score):

$$\prod_{e \in \mathcal{M}} (1 - W[e]) \times \frac{1}{\left(\frac{22-d}{22} \times 4 + 1\right)} \times \frac{1}{n_{nm}^2} \quad (1)$$

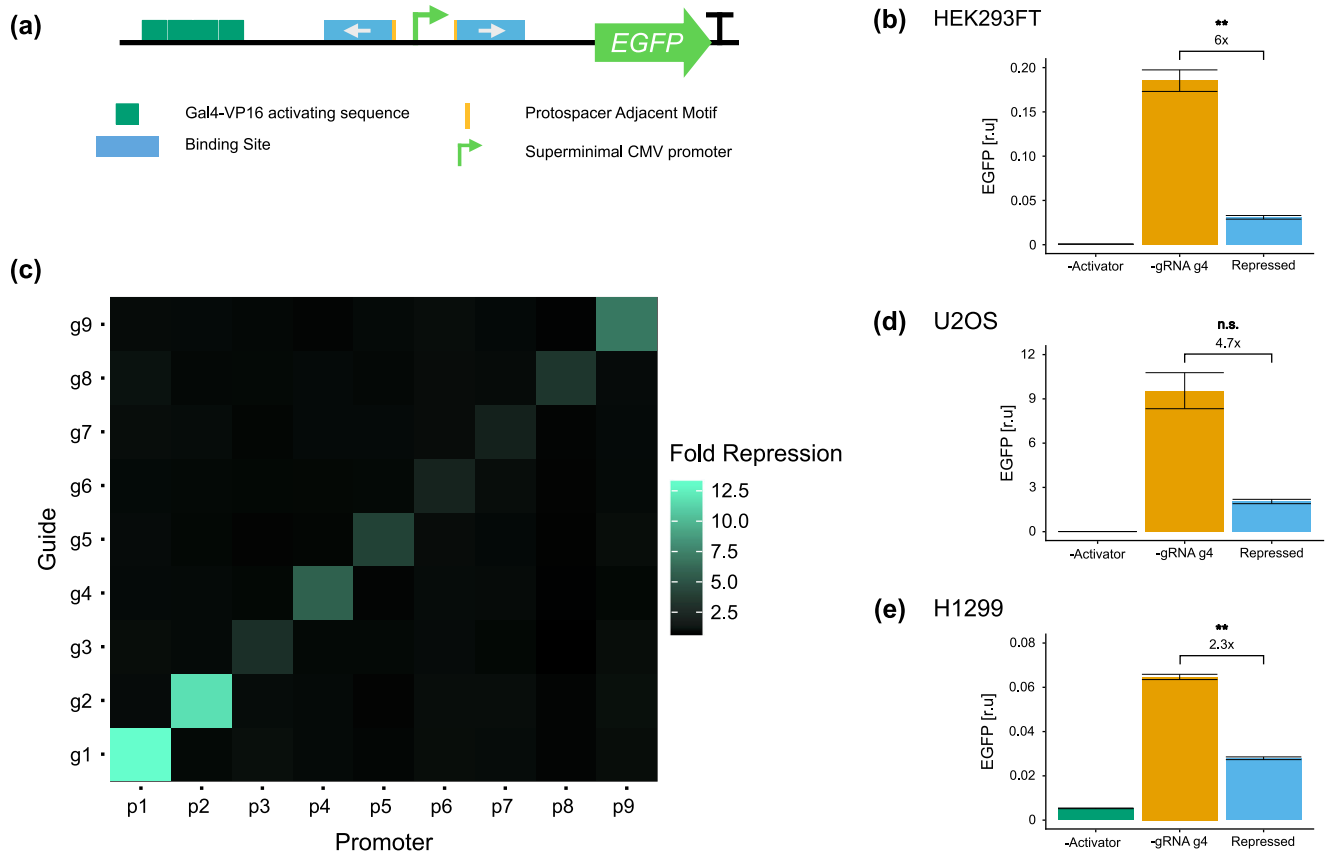


Fig. 3 Test of hdLbCas12a functionality and orthogonality in mammalian cell lines. **a** Design of the best-repressed promoter configuration. Binding sites facing opposite directions flank the smCMV promoter. The 5'-TTTN-3' PAM is proximal to either side of smCMV. **b** EGFP levels in HEK293FT cells in absence of activator as indication of promoter leakiness (-Activator) or gRNA (-gRNA g4), or upon binding of hdLbCas12a/gRNA complex (Repressed). Data represent geometric mean and standard deviation of means of EGFP MEFL normalized by mKate expression for cells expressing $>2 \times 10^4$ MEFL of the transfection marker mKate for at least 3 pooled replicates. **c** Orthogonality matrix of synthetic gRNA/promoters. Repression levels were measured between all combinations of hdLbCas12a/gRNAs and promoters in order to determine levels of repression (with cognate pairs) and crosstalk (with non-cognate pairs). Fold repression is calculated by dividing the normalized geometric mean of EGFP MEFL of unrepressed cells (-gRNA) by the normalized geometric mean of EGFP MEFL of repressed cells (+gRNA), where EGFP MEFL geometric mean is normalized by mKate expression for cells expressing $>2 \times 10^4$ MEFL of transfection marker mKate for at least 2 replicates. Of note, -gRNA samples included hdLbCas12a along with Gal4-VP16 and reporter encoding plasmids to take into account the burden imposed to the cells. **d**, **e** Test of hdLbCas12a/g4 transcription factor/promoter pair-induced repression was tested in U2OS and H1299 cell lines. Promoter output was measured in absence of activator (-Activator), gRNA (-gRNA g4) or in presence of hdLbCas12a/g4 (Repressed). Data represent geometric mean and standard deviation of means of EGFP MEFL normalized by mKate expression for cells expressing $>2 \times 10^4$ (H1299) or $>5 \times 10^3$ (U2OS) MEFL of transfection marker mKate for at least 2 replicates. Statistical significance, calculated using the Student's *t* test, is denoted by asterisks (where ** means $p < 0.01$).

where e is the mismatch position, \mathcal{M} is the mismatch matrix with scores according to mismatch position, \bar{d} is the mean pairwise distance and n_{mm} is the number of mismatches.

Equation 2 (aggregate score):

$$S_{\text{guide}} = \frac{100}{100 + \sum_{i=1}^{n_{\text{mm}}} S_{\text{hit}}(h_i)} \quad (2)$$

Where S_{guide} is the aggregate score for the guide RNA and $S_{\text{hit}}(h_i)$ is the score for each individual genomic hit.

Equation 3 (mismatch matrix):

$$\mathcal{M} = [0.63 \ 0.590 \ 0.780 \ 0.700 \ 0.670 \ 0.720 \ 0.610 \ 0.430 \ 0.310 \ 0.230 \ 0.130 \ 0.320 \ 0.2100 \ 0.350 \ 0.210 \ 0.080 \ 0.250 \ 0.450 \ 0.380 \ 0.340 \ 0.360] \quad (3)$$

In an attempt to maintain similar repression strengths between different CRISPRi promoters, only gRNA spacer sequences with predicted hybrid gRNA-spacer/cognate-DNA binding free energies with a specified range were retained. Hybrid DNA-RNA free energies were calculated using MELTING^{38,39}.

Finally, all-against-all pairwise minimum free energies (MFE) were calculated for all synthetic gRNA sequences using PairFold⁴¹. The largest orthogonal set of synthetic gRNA sequences was determined by retrieving the maximum independent set using the Python graph library NetworkX⁴². An undirected graph was constructed, where each node corresponded to a gRNA sequence. The nodes were connected by edges if the sequences had Smith-Waterman scores above a user defined value, sub-sequence matches above a user defined length or MFES below a user defined value. The maximum clique of the complement graph then gives the maximum independent set. This was carried out using the *find_cliques* function of NetworkX which implements existing algorithms⁴³⁻⁴⁵.

Plasmid assembly

Circular Polymerase Extension Cloning (CPEC) was used for the construction of most of the plasmids used in this paper. Assembly conditions and primer design followed recommendations from Quan et al.⁴⁶.

pMC085-pcDNA3.1-hdLbCas12a was created using site directed mutagenesis and CPEC. pcDNA3.1-hLbCas12a (Addgene #69988) was amplified in three fragments using primers to obtain the catalytically inactive mutant LbCas12a (D832A). Fragments were digested with DpnI (NEB), gel purified using the Zymoclean™ Gel DNA recovery kit (Zymo Research) and reassembled in a thermocycler (Annealing: 55 °C, Extension: 5:00, Cycles: 15). Positive clones were then confirmed by DNA sequencing.

Drop-out constructs pMC102-Prom-RFP and pMC088-U6-Cas12agRNA-RFP were created using CPEC to allow for Type-IIs restriction site cloning of promoter binding sites and gRNAs respectively.

pMC102-Prom-RFP backbone (containing UAS binding site, mammalian EGFP, bacterial origin of replication, Kanamycin resistance and polyA signal) was PCR amplified from pLS1-BoxCDGC_2xKMet⁴⁷ and the dropout was amplified from an Integrated DNA Technologies gBlock containing a constitutively expressed RFP.

pMC088-U6-Cas12agRNA-RFP backbone (containing U6 promoter, bacterial origin of replication and ampicillin resistance) was PCR amplified from pSIREN_U6-shRNA_FF5-CMV-iRFP (courtesy of Dr. Wroblewska), and dropout was PCR amplified from the same Integrated DNA Technologies gBlock containing constitutively expressed RFP. Fragments were then digested with DpnI (NEB), gel purified using the Zymoclean™ Gel DNA recovery kit (Zymo Research) and assembled in a thermocycler using the following conditions: Annealing: 55 °C, Extension: 2:30 (pMC102-Prom-RFP) 2:00 (pMC088-U6-Cas12agRNA-RFP), Cycles: 15.

Colonies were screened according to their phenotype (red color) upon transformation and confirmed with DNA sequencing.

gRNA constructs were created using Typells assembly⁴⁸. Annealed, phosphorylated oligonucleotides with the LbCas12a repeat sequence and specified spacer sequence, were combined in a reaction with T4 DNA Ligase (NEB), BspMI (NEB), pMC088-U6-Cas12agRNA-RFP, Ligase Buffer (NEB) and Bovine Serum Albumin (NEB). Reactions were then thermocycled for 30 cycles of 37 °C for 5:00 followed by 16 °C for 10:00 as described previously⁴⁹. Colonies were screened according to their phenotype (white color) and confirmed with DNA sequencing.

Promoter constructs were created using Typells assembly⁴⁸. Annealed phosphorylated oligonucleotides with flanking binding sites and a central superminimal CMV promoter (TATATAAGCA-GAGCTCGTTTAGTGAACCGTCAGATCGC) were ligated to respective plasmidic backbones in a reaction mix including with T4 DNA Ligase (NEB), Bpil (Thermo Fisher Scientific), pMC102-Prom-RFP, Ligase Buffer (NEB) and Bovine Serum Albumin (NEB). Reactions were then thermocycled for 30 cycles of 37 °C for 5:00 followed by 16 °C for 10:00 as described previously⁴⁹. Colonies were screened according to their phenotype (white color) and confirmed with DNA sequencing.

Transformations were performed in *E. coli* DH5 α cells cultured at 37 degrees in Lysogeny Broth, supplemented with the appropriate antibiotic (kanamycin 50 μ g/ml, carbenicillin 100 μ g/ml) (Sigma).

All plasmids were sequence verified (Eurofins) using appropriate sequencing primers (Supplementary Table 3). Sequence verified transformants were grown in lysogeny broth and stored as glycerol stocks at -80 °C in 20% glycerol. For transfections, plasmids were prepared using the QIAfilter plasmid midi kit (Qiagen) following the manufacturer protocol.

Cell culture

U2OS and HEK293FT cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific) and H1299 cells were maintained in Roswell Park Memorial Institute Medium (RPMI-1640) (Thermo Fisher Scientific). Media was supplemented with 10% Fetal Bovine Serum (Thermo Fisher Scientific), Penicillin-Streptomycin (Thermo Fisher Scientific), L-Glutamine (Thermo Fisher Scientific) and MEM Non-Essential

Amino Acids (Thermo Fisher Scientific) and cells were grown at 37 °C in a 5% CO₂ incubator.

Transfections

All cells were transfected using a "fast forward" protocol according to manufacturer instructions. Cells were trypsinized and plated immediately before adding plasmid/liposome complexes. All cell lines were transfected in 24 well plates with an initial cell number of 2×10^5 cells per well.

A total of 540 ng of DNA was transfected: 20 ng gRNA-responsive promoter, 60 ng hdLbCas12a, 400 ng U6-gRNA, 40 ng Gal4VP16, pL-A2 to 540 ng.

HEK293FT cells were transfected using Lipofectamine 3000 (Thermo Fisher Scientific). H1299 cells were transfected using Lipofectamine 2000 (Thermo Fisher Scientific), and U2OS cells were transfected using FuGENE6 (Promega Corporation) following manufacturer's recommendations. 1 ml of media was added to cells 24 h post-transfection.

For flow cytometer analysis, 48 h post transfection media was removed and cells were washed with PBS before being trypsinized and resuspended in DMEM without phenol red. Cells were then transferred to flow cytometry tubes on ice for sample processing. Cells were analyzed using a BD LSR Fortessa™ flow cytometer.

Inclusion and ethics statement

Research was conducted in compliance with the Global Code of Conduct.

QUANTIFICATION AND STATISTICAL ANALYSIS

Flow cytometry and data analysis

Cells were analyzed with LSR Fortessa flow cytometer, equipped with 488 and 561 nm lasers (BD Biosciences). We collected 10,000–20,000 events per sample and fluorescence data were acquired with the following cytometer settings: 488 nm laser and a 525/20 nm (H1299 and U2OS) or 530/30 nm (HEK293FT) bandpass filter for EGFP, 561 nm laser and a 610/20 nm (H1299 and U2OS) or 660/20 nm (HEK293FT) bandpass filter for mKate. The population of live cells was selected according to FSC/SSC parameters (Supplementary Fig. 6 and Supplementary Table 4). Data analysis was performed with FlowJo and TASBE using Rainbow Calibration Particles (8 peaks, Spherotech)⁵⁰.

Flow cytometry data was converted from arbitrary units to compensated MEFL (Molecules of Equivalent Fluorescein) using the TASBE characterization method⁴¹. The TASBE method uses a strong constitutively expressed fluorophore, which serves as both a transfection marker and an indicator of relative circuit copy count. An affine compensation matrix is computed from single positive and blank controls. FITC measurements are calibrated to MEFL using RCP-30-5-A beads (SpheroTech) and mappings from other channels to equivalent FITC are computed from co-transfection of constitutively expressed mKate, controlled by a CMV promoter on its plasmid.

Cells were manually gated during TASBE analysis to select viable cells according to their Side Scatter (SSC-A) and Forward Scatter (FSC-A) characteristics. Single cells were then selected using their Forward Scatter (FSC-H and FSC-W) properties.

Fold repression was calculated by dividing the geometric mean of EGFP MEFL (normalized by the MEFL of the mKate transfection marker) of unrepressed control cells (transfected with the EGFP reporter, Gal4-VP16 and hdLbCas12a expression plasmids) by the transfection marker normalized geometric mean of EGFP MEFL of unrepressed cells (transfected with plasmids encoding the EGFP reporter, Gal4-VP16, hdLbCas12a and gRNA). Based on the observed constitutive fluorescence distributions, a transfection marker threshold was selected for each data set, below which data

were excluded as being too close to the non-transfected population while also retaining a significant number of cells: 5×10^3 MEFL for U2OS, 1.26×10^4 MEFL for H1299 and 2×10^4 MEFL for HEK293FT data sets. Fold repression for a range of transfection marker thresholds are shown in Supplementary Figs. S3 and S4(b). High outliers are removed by excluding all bins without at least 100 data points. Population geometric statistics are computed over this filtered set of data. All experiments included at least 2 technical replicates and error bars indicate standard deviation. Variance for all groups is generally similar; any differences are reflected in the displayed standard deviation.

DATA AVAILABILITY

Requests for datasets should be directed to the Lead Contact, Velia Siciliano (velia.siciliano@iit.it).

CODE AVAILABILITY

Code is open source and publicly available from Github (<https://github.com/jmacdona/gDesigner>).

Received: 28 April 2022; Accepted: 28 July 2022;
Published online: 16 September 2022

REFERENCES

- Ausländer, S. & Fussenegger, M. Engineering gene circuits for mammalian cell-based applications. *Cold Spring Harb Perspect Biol* **8**, a023895 (2016).
- Ollé-Vila, A., Duran-Nebreda, S., Conde-Pueyo, N., Montañez, R. & Solé, R. A morphospace for synthetic organs and organoids: the possible and the actual. *Integr. Biol.* **8**, 485–503 (2016).
- Schukur, L., Geering, B., Hamri, G. C.-E. & Fussenegger, M. Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis. *Sci. Transl. Med.* **7**, 318ra201 (2015).
- Siciliano, V. et al. MiRNAs confer phenotypic robustness to gene networks by suppressing biological noise. *Nat. Commun.* **4**, 2364 (2013).
- Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
- Mathur, M., Xiang, J. S. & Smolke, C. D. Mammalian synthetic biology for studying the cell. *J. Cell Biol.* **216**, 73–82 (2017).
- Siciliano, V. et al. Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *Plos Comput. Biol.* **7**, e1002074 (2011).
- Bonfá, G., Cella, F. & Siciliano, V. Engineering protein-based parts for genetic devices in mammalian cells. *Methods Mol. Biol.* **2229**, 331–346 (2021).
- Cella, F. & Siciliano, V. Protein-based parts and devices that respond to intracellular and extracellular signals in mammalian cells. *Curr. Opin. Chem. Biol.* **52**, 47–53 (2019).
- Stanton, B. C. et al. Systematic transfer of prokaryotic sensors and circuits to mammalian cells. *ACS Synth. Biol.* **3**, 880–891 (2014).
- Li, Y. et al. Modular construction of mammalian gene circuits using TALE transcriptional repressors. *Nat. Chem. Biol.* **11**, 207–213 (2015).
- Culler, S. J., Hoff, K. G. & Smolke, C. D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **330**, 1251–1255 (2010).
- Bird, J. *Engineering mathematics* (Elsevier, 2007).
- Cong, L., Zhou, R., Kuo, Y., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA binding modules and transcriptional repressor domains. *Nat. Commun.* **3**, 968–968 (2012).
- Lohmueller, J. J., Armel, T. Z. & Silver, P. A. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res.* **40**, 5180–5187 (2012).
- Deuschle, U., Meyer, W. K. & Thiesen, H. J. Tetracycline-reversible silencing of eukaryotic promoters. *Mol. Cell Biol.* **15**, 1907–1914 (1995).
- Kramer, B. P., Fischer, C. & Fussenegger, M. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol. Bioeng.* **87**, 478–484 (2004).
- Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
- Kiani, S. et al. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11**, 723–726 (2014).
- Gu, M. B., Todd, P. & Kompala, D. S. Metabolic burden in recombinant CHO cells: effect of hfr gene amplification and lacZ expression. *Cytotechnology* **18**, 159–166 (1995).
- Niklas, J., Priesnitz, C., Rose, T., Sandig, V. & Heinze, E. Metabolism and metabolic burden by α 1-antitrypsin production in human AGE1.HN cells. *Metab. Eng.* **16**, 103–114 (2013).
- Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–261 (1995).
- Fonfara, I., Richter, H., Bratovič, M., Rhun, A. L. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
- Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
- Kim, H. K. et al. In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nat. Methods* **14**, 153–159 (2017).
- Gao, L. et al. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792 (2017).
- Tak, Y. E. et al. Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. *Nat. Methods* **14**, 1163–1166 (2017).
- Liu, Y. et al. Engineering cell signaling using tunable CRISPR–Cpf1-based transcription factors. *Nat. Commun.* **8**, 2095 (2017).
- Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
- O’Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* **43**, 3389–3404 (2015).
- Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**, 1120–1123 (2015).
- Xie, S., Shen, B., Zhang, C., Huang, X. & Zhang, Y. sgRNACas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *Plos One* **9**, e100448 (2014).
- Casini, A. et al. R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol.* **3**, 525–528 (2014).
- MacDonald, J. T. & Siciliano, V. Computational sequence design with R2oDNA designer. *Methods Mol. Biol. Clifton N J* **1651**, 249–262 (2017).
- Kulakovskiy, I. V. et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.* **44**, D116–D125 (2016).
- Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
- Novère, N. L. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinform. Oxf. Engl.* **17**, 1226–1227 (2001).
- Dumousseau, M., Rodriguez, N., Juty, N. & Novère, N. L. MELTING, a flexible program to predict the melting temperatures of nucleic acids. *BMC Bioinformatics* **13**, 101–101 (2012).
- Karreth, F. A., Tay, Y. & Pandolfi, P. P. Target competition: transcription factors enter the limelight. *Genome Biol.* **15**, 114–114 (2014).
- Andronescu, M., Zhang, Z. C. & Condon, A. Secondary structure prediction of interacting RNA molecules. *J. Mol. Biol.* **345**, 987–1001 (2005).
- Hagberg, A., Swart, P. & Schult, D. Exploring network structure, dynamics, and function using networkx. United States: N. p., 2008. Web.
- Bron, C. & Kerbosch, J. Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM* **16**, 575–577 (1973).
- Tomita, E., Tanaka, A. & Takahashi, H. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theor. Comput. Sci.* **363**, 28–42 (2006).
- Cazals, F. & Karande, C. A note on the problem of reporting maximal cliques. *Theor. Comput. Sci.* **407**, 564–568 (2008).
- Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6**, 242–251 (2011).
- Wroblewska, L. et al. Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nat. Biotechnol.* **33**, 839–841 (2015).
- Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
- Moore, S. J. et al. EcoFlex: a multifunctional MoClo kit for E. coli synthetic biology. *ACS Synth. Biol.* **5**, 1059–1069 (2016).
- Beal, J. et al. An end-to-end workflow for engineering of biological networks from high-level specifications. *ACS Synth Biol* **1**, 317–331 (2012).

ACKNOWLEDGEMENTS

We thank Kirsten Jensen and Ciaran Mckeown for lab support. This project was supported by the Imperial College Research Fellowship (ICRF), the Istituto Italiano di

Tecnologia (IIT), the ERC Starting grant Synthetic T-rex (852012), the EPSRC, UK (EP/K034359/1 and EP/M006700/1).

AUTHOR CONTRIBUTIONS

V.S. and J.T.M. conceived the idea. V.S., J.T.M. and P.F. directed the research. M.A.C. and J.T.M. developed the software. M.A.C. and V.S. conducted experiments. J.T.M., M.A.C. and V.S. analyzed the data. V.S., J.T.M. and M.A.C. wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-022-00241-w>.

Correspondence and requests for materials should be addressed to James T. MacDonald, Paul S. Freemont or Velia Siciliano.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022