





# A synthetic transcription platform for programmable gene expression in mammalian cells

Received: 1 April 2021

Accepted: 13 September 2022

Published online: 18 October 2022

 Check for updates

William C. W. Chen <sup>1,2,3,4,5,14</sup> ✉, Leonid Gaidukov<sup>1,3,14</sup>, Yong Lai<sup>1,2,14</sup>, Ming-Ru Wu<sup>1,11</sup>, Jicong Cao<sup>1</sup>, Michael J. Gutbrod<sup>6,7</sup>, Gigi C. G. Choi <sup>1,12</sup>, Rachel P. Utomo<sup>1,8</sup>, Ying-Chou Chen<sup>1,2,13</sup>, Liliana Wroblewska<sup>9</sup>, Manolis Kellis <sup>6,7</sup>, Lin Zhang<sup>10</sup>, Ron Weiss <sup>1,3</sup> & Timothy K. Lu<sup>1,2,3</sup> ✉

Precise, scalable, and sustainable control of genetic and cellular activities in mammalian cells is key to developing precision therapeutics and smart bio-manufacturing. Here we create a highly tunable, modular, versatile CRISPR-based synthetic transcription system for the programmable control of gene expression and cellular phenotypes in mammalian cells. Genetic circuits consisting of well-characterized libraries of guide RNAs, binding motifs of synthetic operators, transcriptional activators, and additional genetic regulatory elements express mammalian genes in a highly predictable and tunable manner. We demonstrate the programmable control of reporter genes epically and chromosomally, with up to 25-fold more activity than seen with the EF1 $\alpha$  promoter, in multiple cell types. We use these circuits to program the secretion of human monoclonal antibodies and to control T-cell effector function marked by interferon- $\gamma$  production. Antibody titers and interferon- $\gamma$  concentrations significantly correlate with synthetic promoter strengths, providing a platform for programming gene expression and cellular function in diverse applications.

The regulation of gene expression in complex organisms has been a central focus for characterizing disease variation<sup>1</sup>, producing monoclonal antibodies (mAbs)<sup>2</sup>, developing gene and cell therapies<sup>3</sup>, and investigating other biological phenomena<sup>4</sup>. Synthetic biology offers powerful ways to harness artificial gene regulatory tools in mammalian systems<sup>5–9</sup>. For example, tumor-specific synthetic gene circuits applied to cancer immunotherapy yield significant anti-tumor responses in mice<sup>10</sup>. However, the use of strong constitutive promoters in gene expression platforms can increase the expression of target genes to the point of causing unwanted, dose-dependent side effects, raising safety concerns<sup>11</sup>. For instance, the use of the cytomegalovirus promoter (CMVp) to express a tumor-killing gene markedly increases apoptosis in normal cells and induces acute systemic toxicity in vivo<sup>12</sup>. Although promoter substitution is a simple and commonly implemented strategy for altering gene expression<sup>13</sup>, optimizing cell type-

specific or gene-specific natural promoters demands extensive effort<sup>14</sup>. Thus, controlling target gene expression with natural promoters has had only limited success in achieving desired biological phenotypes or therapeutic outcomes.

Another approach to regulating gene expression is to engineer transcription factors (TFs) and transcriptional activation domains (TADs) to control transcriptional activities<sup>15</sup>. Artificial transcription factors (aTFs), which can be rationally designed in silico, have been derived from zinc fingers (ZFs)<sup>16</sup>, transcription activator-like effectors (TALEs)<sup>17</sup>, and clustered regularly interspaced short palindromic repeats associated protein (CRISPR-Cas)<sup>18,19</sup>. By conjugating DNA-binding proteins with various TADs, such as VP16<sup>20</sup>, VP64<sup>21</sup>, and VPR (VP64-p65-RTA)<sup>22</sup>, researchers have demonstrated the utility of tuning promoter activity with aTFs. Particularly, CRISPR-based aTFs (crisprTFs) are simpler to customize and target to genomic loci of

A list of author affiliations appears at the end of the paper. ✉ e-mail: [William.Chen@usd.edu](mailto:William.Chen@usd.edu); [tim@lugroup.org](mailto:tim@lugroup.org)

interest, using guide RNA (gRNA) with homology, than complex ZFs and TALEs; thus, crisprTFs are rapidly gaining popularity in biomedical research<sup>23</sup>. For example, several types of crisprTFs and compound activation mediators, based on deactivated CRISPR-associated protein 9 (dCas9), have been widely used in mammalian cells, such as dCas9-VP64<sup>24</sup>, dCas9-VPR, SunTag<sup>25</sup>, and synergistic activation mediator (SAM)<sup>26</sup>. dCas9-VPR, SunTag, and SAM can strongly activate genes in multiple species<sup>27</sup>. Our group has also developed crisprTFs to regulate gene expression driven by natural or artificial eukaryotic promoters<sup>28</sup>.

Here, we build upon these aTF platforms by creating a comprehensive crisprTF promoter system for the programmable regulation of gene expression in mammalian cells. Our goal is to engineer a universal platform for tunable, scalable, and consistent transcriptional control in a wide variety of contexts, applicable to various cell types or target genes. Specifically, through mimicking natural mammalian promoters, we have created modular libraries of both crisprTFs and synthetic operators by: (1) altering gRNA sequences; (2) adjusting the number of gRNA binding sites (BS) in the synthetic operator; (3) incorporating additional control elements in the operator or crisprTF to augment expression; and (4) designing multiple orthogonal crisprTFs. Because it implements a multi-tier gene circuit assembly design, this system has the advantage of operating at both epigenomic and genomic levels with precise tunability, versatile modularity, and high scalability.

To demonstrate the utility of this synthetic transcription platform, we first validated the precise control of two fluorescent reporter genes and then programmed the production of recombinant human mAbs, including a functional checkpoint antibody, anti-human programmed cell death protein 1 (anti-hPD1)<sup>29</sup>. High-yield, stable production was achieved by using crisprTF promoters within a recombinase-mediated, multi-landing pad (multi-LP) DNA integration platform<sup>30</sup>. Multi-LP DNA integration in genomic safe harbor loci (e.g., the *Rosa26* locus) enables predictable single-copy integration, limited transgene silencing, stable gene expression, and consistent long-term protein production<sup>30–32</sup>. Anti-hPD1 gene circuits expressed chromosomally with this system modulated certain anti-tumor phenotypes of human T cells. These results indicate that highly tunable, sustainable, and predictable protein expression over a wide dynamic range can be achieved with our scalable, modular synthetic transcription system.

## Results

### Construction of a programmable, modular synthetic transcription system with crisprTFs

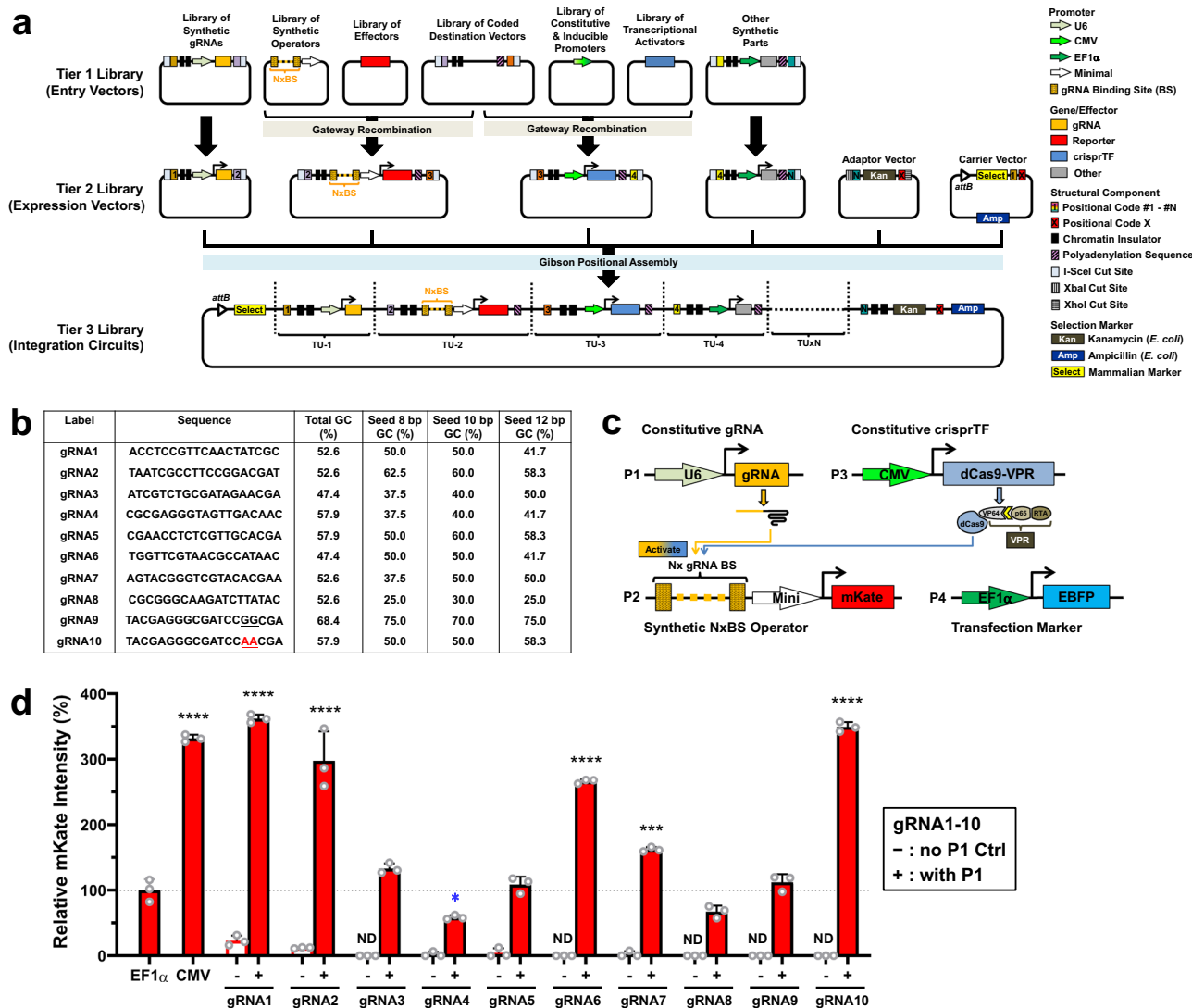
To enable high tunability and versatility of the synthetic transcription system for a wide spectrum of applications, we adopted a 3-tiered modular library design<sup>33</sup>. The Tier 1 entry vector library encodes a variety of interchangeable gene regulatory parts and effector genes, including crisprTFs, gRNAs, operators, and other components (Fig. 1a). Tier 1 parts can be assembled into the Tier 2 expression vector library for transient expression, followed by modular assembly into the Tier 3 integration gene circuit library (Fig. 1a). We built a library of guide RNAs (gRNAs) that were orthogonal to the CHO genome and first selected eight of the scored ones for evaluation (Fig. 1b). For each gRNA, we designed a corresponding operator containing 8× complementary gRNA BS to drive the expression of a far-red fluorescent reporter gene, *mKate*, and evaluated their performance episomally (Fig. 1c and Supplementary Fig. 1). Flow cytometry results showed a wide range of *mKate* expression levels among different gRNAs (Fig. 1d). As active natural mammalian promoter sequences typically have a high GC content (57%)<sup>34</sup>, we reasoned that the GC content of the protospacer adjacent motif (PAM)-proximal seed region (8–12 bases at the 3'-end of a gRNA and its BS sequence) plays a role in regulating gene expression. gRNAs with a GC content of ~50–60% in seed sequences appeared to express *mKate* at higher levels than gRNAs with a lower or higher GC content (Fig. 1b, d). Therefore, we selected a relatively weak gRNA (gRNA9) with a high GC content (≥70%) and

introduced mutations in two consecutive bases within the seed sequence to alter its GC content. The resulting gRNA10 and its matching operator, with ≥50% GC and still orthogonal to the CHO genome, yielded much higher *mKate* expression than its ancestor, gRNA9 (Fig. 1d,  $p = 0.0004$ ; one-way analysis of variance (ANOVA) with Dunnett's Multiple Comparison). The CMV control and gRNA1, 2, 6, 7, and 10 had higher expression (Fig. 1d, all  $p \leq 0.0005$ ), while only gRNA4 had lower expression, than the EF1 $\alpha$  control (Fig. 1d,  $p < 0.05$ ). The expression obtained with gRNA1, 2, and 10 was not significantly different from that obtained with the CMV control (all  $p > 0.05$ ). Overall, the top four gRNAs, with more than twofold higher expression than EF1 $\alpha$ , had 50–60% GC in the first 8–10 bases of the seed sequences (Supplementary Fig. 2). We selected three representative gRNAs that yielded weak (gRNA4), medium (gRNA7), and strong expression (gRNA10) to investigate further. A comparison of three crisprTFs (dCas9-VP16, dCas9-VP64, dCas9-VPR) showed that dCas9-VPR yielded a markedly higher expression level than dCas9-VP16 or dCas9-VP64 (Supplementary Fig. 3, both  $p < 0.0001$ ; one-way ANOVA with Dunnett's Multiple Comparison), consistent with a previous report<sup>27</sup>. Thus, to enable the largest dynamic expression range possible, we selected dCas9-VPR for subsequent investigation.

### Transcriptional programming in multiple mammalian cell types

To control gene expression at distinct levels, we built a library of synthetic operators containing 2×–16× gRNA BS for gRNA4, 7, and 10 (Fig. 1c). Transient expression of *mKate* was assessed in CHO cells at 48 hours post-transfection (Supplementary Fig. 4). Flow cytometry data showed dramatically different patterns of *mKate* expression among the three gRNA series (Fig. 2a). Comparison of each gRNA series with expression levels observed with the EF1 $\alpha$  promoter control showed: reporter activities for the gRNA4 series that ranged from 15% (2× BS) to 270% (16× BS), with 2×–8× BS driving notably lower, and 16× BS driving significantly higher, expression; reporter activities for the gRNA7 series that ranged from 26% (2× BS) to 760% (16× BS), with 2×–4× BS driving notably lower, and 8×–16× BS driving significantly higher, expression; and reporter activities for the gRNA10 series that ranged from 30% (2× BS) to 1107% (16× BS), with 6×–16× BS having significantly higher expression (Fig. 2a). Significant correlations between expression levels and the number of BS were found with all three gRNA series (Supplementary Fig. 5a, all  $p < 0.05$ ; Pearson correlation analysis). Overall, with differences in gRNA sequences and the number of complementary gRNA BS in the operators, we achieved a wide dynamic range of ~74-fold change in the intensity of the reporter signals with these gRNA series in CHO cells. Gene expression was proportionate to the number of BS, indicating consistent tunability.

To translate these results into other mammalian cell types (mouse, rat, and human), we selected the gRNA10 series because it had the lowest leakage and the highest expression among the three comprehensively tested gRNAs. Interestingly, we observed dramatic differences in *mKate* expression with the CMV control in mouse C2C12 myoblasts (Fig. 2b), human embryonic kidney (HEK)-293T cells (Fig. 2c), rat H9c2 cardiac myoblasts (Fig. 2d), and human induced pluripotent stem cells (hiPSCs) (Fig. 2e). The gRNA10 series behaved in a similar order in these cells as it did in CHO cells. In C2C12 myoblasts, *mKate* expression ranged from 11% (2× BS) to 316% (16× BS) of EF1 $\alpha$ , with 2×–6× BS being notably lower and 8×–16× BS being significantly higher than that observed with EF1 $\alpha$  (Fig. 2b). In HEK cells, *mKate* expression ranged from 60% (2× BS) to 1026% (16× BS) of EF1 $\alpha$ , with 6×–16× BS being significantly stronger than EF1 $\alpha$  (Fig. 2c). In rat H9c2 cells, *mKate* expression ranged from 27% (2× BS) to 688% (16× BS) of EF1 $\alpha$ , with 6×–16× BS being significantly stronger than EF1 $\alpha$  (Fig. 2d). In hiPSC cells, *mKate* expression ranged from 80% (2× BS) to 950% (16× BS) of EF1 $\alpha$ , with 8×–16× BS being significantly stronger than EF1 $\alpha$  (Fig. 2e). Similarly, expression levels and the number of BS markedly correlated in all four rodent and human cell types tested with the



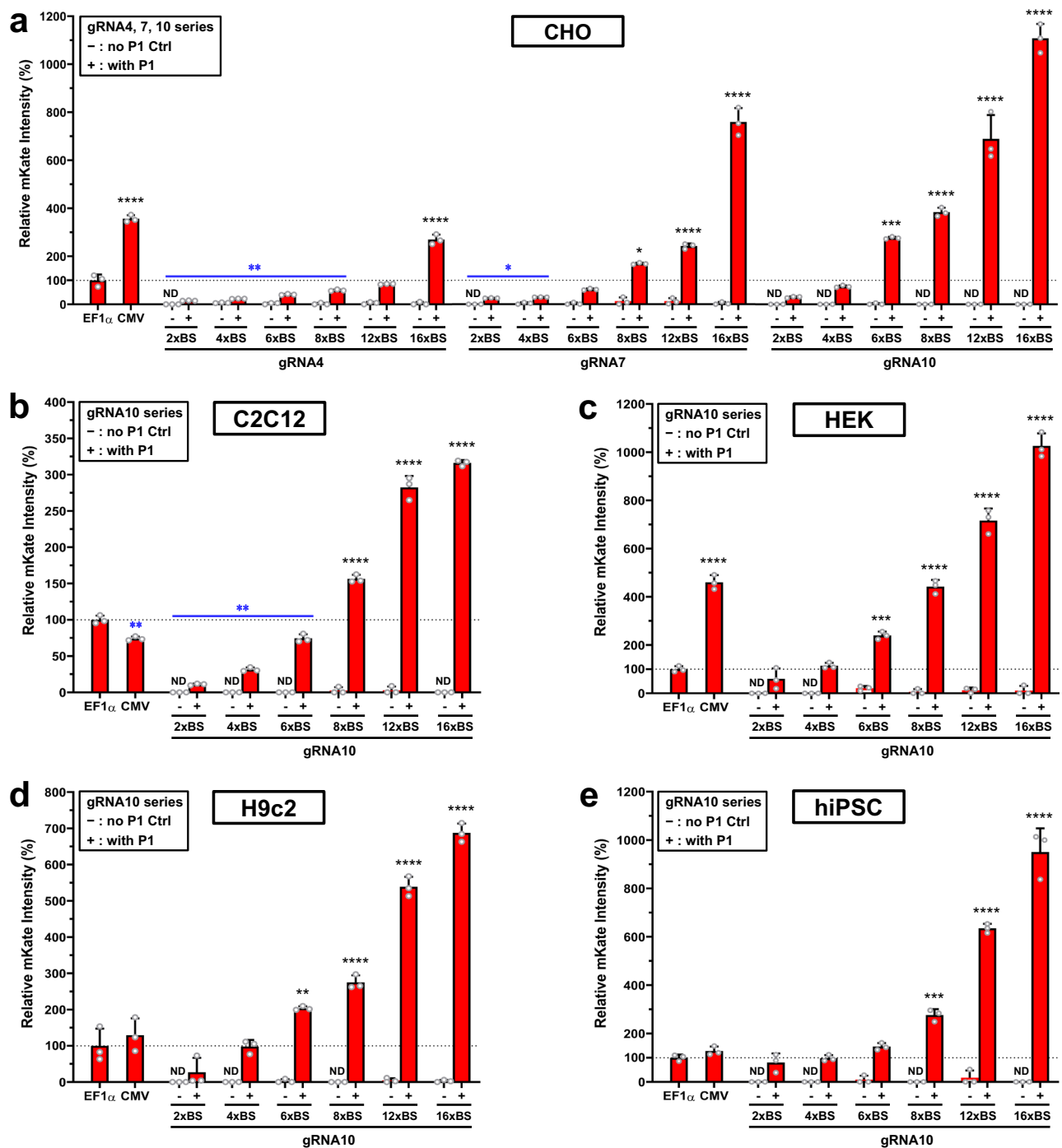
**Fig. 1 | Design and development of the crisprTF promoter system.** **a** Schematic illustration of the programmable and modular design of the crisprTF-based transcription system. To increase its programmability, this platform was modularly divided into three tiers of libraries constructed with the Gateway-Gibson cloning approach. The Tier 1 library was composed of entry vector modules separately encoding gRNAs, synthetic operators with gRNA binding sites (BS) upstream of a minimal promoter, effector genes, crisprTFs and their associated promoters, and other transcriptional control elements. Tier 1 library units were assembled into positional expression vectors with pre-defined orders by Gateway cloning, forming the Tier 2 library. Positional assembly by Gibson cloning was performed to connect independent transcriptional units (TUs), derived from positional expression vectors in the Tier 2 library by I-SceI restriction digestion, into complete gene circuits. The Tier 3 library comprised integration circuits enabling precision control of the target gene(s) when integrated into a landing pad (i.e., a designated chromosomal safe harbor) with BxB1 integrase-mediated, site-specific chromosomal integration. **b** Ten gRNAs (gRNA1-10) orthogonal to the CHO genome were screened for expression. gRNA10 was modified from gRNA9 with GG-to-AA mutations to reduce the GC content of the seed and the entire sequences. **c** To evaluate episomal gene expression levels, CHO-K1 cells were transiently transfected with four plasmids: plasmid #1 (P1)

constitutively expressing gRNA; plasmid #2 (P2) encoding the synthetic operator with some number (x) of gRNA BS to drive mKate expression; plasmid #3 (P3) constitutively expressing a crisprTF; and plasmid #4 (P4) constitutively expressing the transfection marker (EBFP). mKate signals were assessed at 48 hours post-transfection. **d** For the gRNA screening, each gRNA was paired with a matching synthetic operator containing 8× gRNA BS to control mKate expression. EF1 $\alpha$  and CMV promoters driving mKate expression served as positive controls. Experimental groups (+), represented by red solid bars, were transfected with four plasmids (P1–P4). Control groups (–), represented by red hollow bars, to detect baseline operator leakage were transfected without P1 (P2–P4). Data were normalized to the EF1 $\alpha$  control and are presented as relative median mKate intensity (%). gRNA1 and gRNA2 operators exhibited notable leakage without gRNA, suggesting non-specific transcriptional activities that were not associated with targeted crisprTF binding (CMV, gRNA1, gRNA2, gRNA4, gRNA6, gRNA7, and gRNA10 P1 + vs. EF1 $\alpha$ ;  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.0264$ ,  $p < 0.0001$ ,  $p = 0.0005$ , and  $p < 0.0001$ , respectively). Data represent the mean  $\pm$  SD ( $n = 3$ ) (one-way ANOVA with multiple comparisons corrected by Dunnett test; for increased expression marked in black: \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ; for decreased expression marked in blue: \* $p < 0.05$ ; ND not detected). Source data are provided as a Source Data file.

gRNA10 series (Supplementary Fig. 5b–e, all  $p < 0.001$ ; Pearson correlation analyses). Collectively, with the gRNA10 series alone, we achieved up to -29-fold and 17-fold changes in mKate expression in rodent and human cells, respectively.

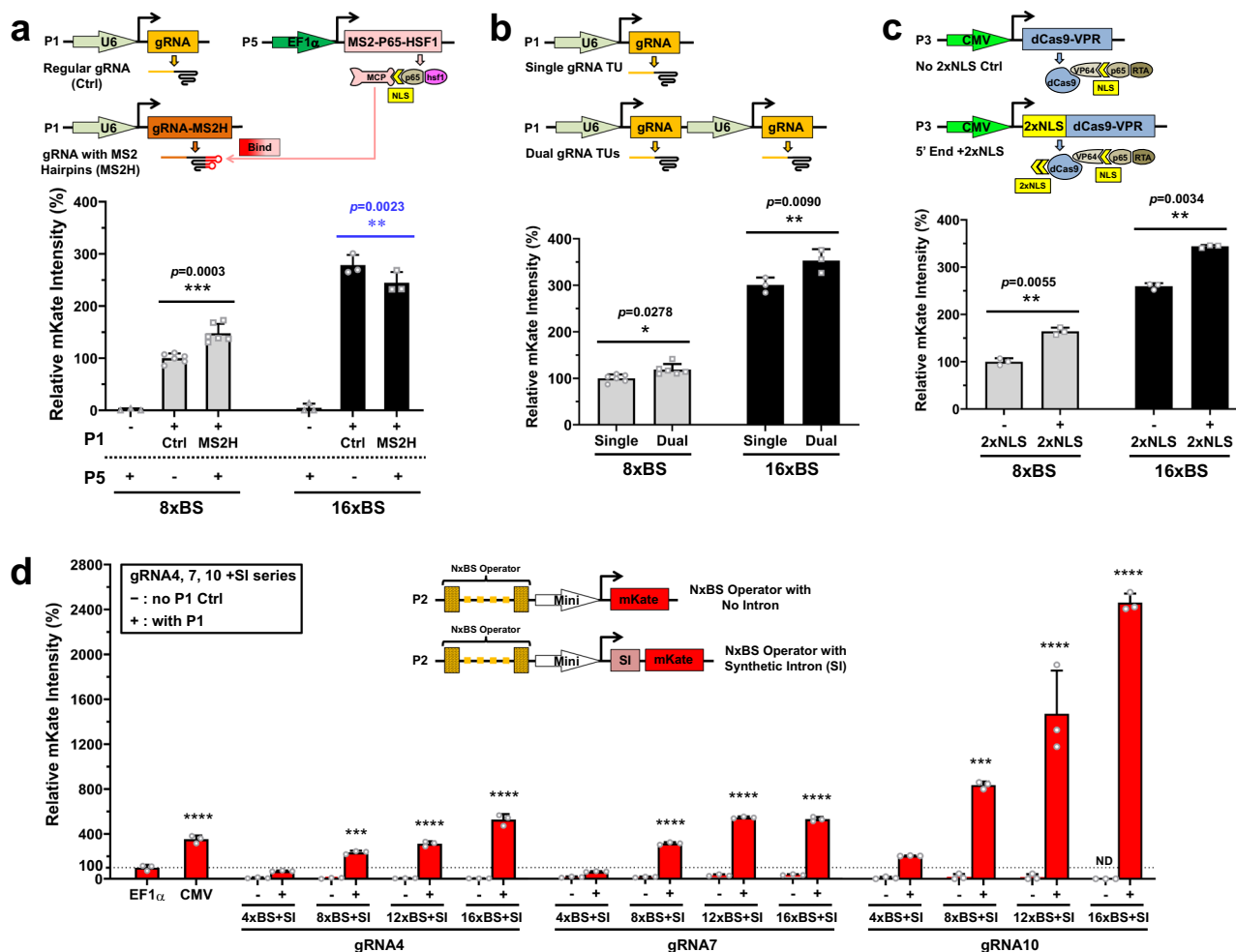
To compare the programmable functionality of our crisprTF-based transcription platform to another orthogonal synthetic TF-

based system, we constructed a Gal4-VPR-based platform containing 2×–8× UAS-BS with the same architecture (Supplementary Fig. 6a). Although the Gal4-VPR/UAS system achieved mKate expression ranging from 11% (2× UAS-BS) to 745% (4× UAS-BS) of crisprTF 8× gRNA10-BS in CHO-K1 cells (Supplementary Fig. 6b), we observed no correlation between gene expression levels and the number of UAS-BS in the



**Fig. 2 | Gene expression programmed by the number of gRNA BS in synthetic operators.** A library of gRNA4, 7, and 10 synthetic operators containing 2 $\times$ –16 $\times$  gRNA BS was built to assess the number of gRNA BS that would be effective as a programmable parameter in controlling gene expression levels. Experimental groups (+), represented by red solid bars, were transiently transfected with four plasmids (P1–P4 in Fig. 1c) whereas gRNA-free control groups (–), represented by paired red hollow bars, were transfected without P1 (only P2–P4 in Fig. 1c) to detect operator leakage. EF1 $\alpha$  and CMV promoters served as positive controls. Median mKate signals relative to the EF1 $\alpha$  control were analyzed at 48 hours post-transfection with flow cytometry. **a** Synthetic operators of the gRNA4, 7, and 10 series with 2 $\times$ –16 $\times$  BS were examined in CHO-K1 cells (one-way ANOVA with multiple comparisons corrected by Dunnett test for each gRNA series with the same EF1 $\alpha$  and CMV control groups; CMV, gRNA4 P1(+) 2 $\times$ , 4 $\times$ , 6 $\times$ , 8 $\times$ , and 16 $\times$  BS, gRNA7 P1(+) 2 $\times$ , 4 $\times$ , 8 $\times$ , 12 $\times$ , and 16 $\times$  BS, and gRNA10 P1(+) 6 $\times$ , 8 $\times$ , 12 $\times$ , and 16 $\times$  BS vs. EF1 $\alpha$ :  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.0002$ ,  $p = 0.0051$ ,  $p < 0.0001$ ,

$p = 0.0069$ ,  $p = 0.0106$ ,  $p = 0.0122$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.0007$ ,  $p < 0.0001$ ,  $p < 0.0001$ , and  $p < 0.0001$ , respectively). **b–e** Synthetic operators of the gRNA10 series with 2 $\times$ –16 $\times$  BS were tested in **b** mouse C2C12 myoblasts (CMV and gRNA10 P1(+) 2 $\times$ , 4 $\times$ , 6 $\times$ , 8 $\times$ , 12 $\times$ , and 16 $\times$  BS vs. EF1 $\alpha$ :  $p = 0.0016$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.0018$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ , respectively), **c** human HEK293T cells (CMV and gRNA10 P1(+) 6 $\times$ , 8 $\times$ , 12 $\times$ , and 16 $\times$  BS vs. EF1 $\alpha$ :  $p < 0.0001$ ,  $p = 0.0007$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ , respectively), **d** rat H9C2 cardiomyoblast cells (gRNA10 P1(+) 6 $\times$ , 8 $\times$ , 12 $\times$ , and 16 $\times$  BS vs. EF1 $\alpha$ :  $p = 0.0063$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ , respectively), and **e** hiPSCs (gRNA10 P1(+) 8 $\times$ , 12 $\times$ , and 16 $\times$  BS vs. EF1 $\alpha$ :  $p = 0.0004$ ,  $p < 0.0001$ ,  $p < 0.0001$ , respectively). Data represent the mean  $\pm$  SD ( $n = 3$ ) (one-way ANOVA with multiple comparisons corrected by the Dunnett test). For increased expression (marked in black): \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ; \*\*\*\* $p \leq 0.0001$ ; for decreased expression (marked in blue): \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ ; ND not detected. Source data are provided as Source Data files.



**Fig. 3 | Genetic control elements added to maximize the expression level of the target gene.** Schematic illustrations depicting individual experiments exhibit only plasmid constructs that were different from P1–P4 in Fig. 1c. Additional genetic control elements were incorporated into the crisprTF promoters (a–d). Experimental groups (+) were transiently transfected with four or five plasmids (P1–P4 or P1–P5), whereas negative controls (–) were transfected without P1. a Synergistic activation mediator (SAM). dCas9-VPR was combined with SAM, another strong synthetic transcriptional activator. Two-tailed paired Student’s *t* test was performed to compare two P1(+) groups with the same number of gRNA BS (with the 8× BS operator, MS2H gRNA vs. control gRNA,  $p = 0.0003$ ; with the 16× BS operator, MS2H gRNA vs. control gRNA,  $p = 0.0023$ ) b Dual gRNA transcriptional units (TUs). An extra gRNA10 TU was added to increase gRNA expression (two-tailed paired Student’s *t* test; with the 8× BS operator, dual gRNA vs. single gRNA,  $p = 0.0278$ ; with the 16× BS operator, dual gRNA vs. single gRNA,  $p = 0.0090$ ). c An additional 2× nuclear localizing sequence (NLS) in the crisprTF was added at the 5’-end of

dCas9-VPR (two-tailed paired Student’s *t* test; with the 8× BS operator, 2× NLS vs. control,  $p = 0.0055$ ; with the 16× BS operator, 2× NLS vs. control,  $p = 0.0034$ ). d A synthetic intron (SI) was added at the 5’ UTR of the target gene. The SI was incorporated into four synthetic operators of the gRNA4, 7, and 10 series (4×, 8×, 12×, and 16× BS), respectively. Red solid bars represent experimental groups (+); paired red hollow bars represent corresponding control groups without gRNA (–) (one-way ANOVA with multiple comparisons corrected by Dunnett test for each gRNA series with the same EF1 $\alpha$  and CMV control groups; CMV, gRNA4 P1(+) 8×, 12×, and 16× BS with SI, gRNA7 P1(+) 8×, 12×, and 16× BS with SI, and gRNA10 P1(+) 8×, 12×, and 16× BS with SI vs. EF1 $\alpha$ :  $p < 0.0001$ ,  $p = 0.0003$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p < 0.0001$ ,  $p = 0.0006$ ,  $p < 0.0001$ , and  $p < 0.0001$ , respectively). All data represent the mean  $\pm$  SD ( $n \geq 3$ ). For increased expression (marked in black): \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ ; for decreased expression (marked in blue): \*\* $p < 0.01$ ; ND not detected. Source data are provided as Source Data files.

operators (Supplementary Fig. 6c,  $p > 0.05$ ; Pearson correlation analysis). Altogether, these results underscore the predictability and tunability of our crisprTF-based system in mammalian cells.

### Incorporation of additional genetic control elements to enhance gene expression

To extend tunability over a greater range, we incorporated the following additional genetic control elements in the crisprTF-based system: (i) a strong synthetic transcriptional activator-SAM; (ii) dual gRNA transcriptional units (TUs) to increase gRNA expression; (iii) an additional 2× nuclear localizing sequence (NLS) at the 5’-end of the dCas9-VPR gene; and (iv) a synthetic intron (SI) to enhance gene expression<sup>35</sup> (Fig. 3). Transient mKate expression in CHO-K1 cells was assessed at 48 hours post-transfection. The combination of SAM, dCas9-VPR, and

the gRNA10 8× BS operator resulted in a 48% increase in gene expression ( $p = 0.0003$ ); however, gene expression decreased by 12.3% when SAM was combined with dCas9-VPR and the 16× BS operator ( $p = 0.0023$ ), suggesting a limitation of gene activation by conjoining various TADs (Fig. 3a; two-tailed paired Student’s *t* test). With the addition of an extra gRNA10 TU to boost the amount of gRNA, we observed modest increases, of 18.7% ( $p = 0.0278$ ) and 17.5% ( $p = 0.009$ ), with gRNA10 8× and 16× BS operators, respectively (Fig. 3b; two-tailed paired Student’s *t* test). Thus, a single gRNA TU may be sufficient for many applications. With an additional 2× NLS incorporated at the 5’-end of dCas9-VPR, mKate expression increased by 64.0% ( $p = 0.0055$ ) and 33.0% ( $p = 0.0034$ ) with gRNA10 8× and 16× BS operators, respectively, suggesting that the nuclear localization of the original dCas9-VPR may be slightly insufficient to gain maximum

activation of gene expression (Fig. 3c; two-tailed paired Student's *t* test).

On the other hand, the addition of SI at the 5' untranslated region (UTR) of the target gene led to an ~200–300% elevation in mKate signals with nearly all operators in all three gRNA series when compared with their original counterparts, even with the strongest operator (i.e., 16× BS of gRNA10; Supplementary Fig. 7). The gRNA4-SI series yielded expression levels that ranged from 70% (4× BS + SI) to 528% (16× BS + SI) of those observed with the EF1α promoter, with 8×–16× BS + SI being notably higher than the EF1α control; the gRNA7-SI series ranged from 65% (4× BS + SI) to 532% (16× BS + SI) of EF1α, with 8×–16× BS + SI being significantly higher than the EF1α control; the gRNA10-SI series ranged from 205% (4× BS + SI) to 2463% (16× BS + SI) of EF1α, with 8×–16× BS + SI showing significantly higher expression than that observed with the EF1α promoter (Fig. 3d). We did not observe any substantial increases of fluorescent signals in the absence of gRNA in any of the SI constructs (Fig. 3d). These results indicated that SI can be fully compatible with synthetic crisprTF promoter-based gene regulation and is an efficient control element to increase gene expression. Notable correlations were found between expression levels and the number of BS with the gRNA4-SI and gRNA10-SI series (Supplementary Fig. 8, both  $p < 0.05$ ; Pearson correlation analysis). Taken together, by adding SI to the original gRNA operators, we expanded the achievable dynamic expression range in CHO cells to an ~167-fold change, offering analog precision control of gene expression with a near-continuous spectrum (Supplementary Fig. 9).

### Precision control of gene expression via genomic integration

We designed our crisprTF promoter platform to be fully compatible with our multi-LP DNA integration platform<sup>30</sup> for the stable expression of large gene circuits in engineered mammalian cells. We built gene circuits that contained insulated TUs encoding different gRNAs, gRNA operator-target gene pairs, and a crisprTF gene, together with an integration-enabled circuit selection marker gene, *puromycin* (Fig. 4a). Chromosomal integration of a single DNA copy of the gene circuit was mediated by BxB1 integrase into a single-LP (sLP) locus in engineered adherent CHO cells (Fig. 4a). After selection with puromycin for integration in engineered sLP-CHO cells, indicated by the disappearance of the EYFP signal, we analyzed target gene expression by flow cytometry. Both the gRNA4 and gRNA10 series exhibited tunable control of expression profiles in sLP-CHO cells, similar to that seen with their episomal counterparts (Fig. 4b). Gene expression of the gRNA4 series ranged from 7% (2× BS) to 56% (16× BS) of that observed with the integrated EF1α control, with 2×–6× BS exhibiting markedly lower expression than with EF1α; the gRNA10 series ranged from 13% (2× BS) to 207% (16× BS) of EF1α, with 2×–4× BS also exhibiting significantly lower expression than with EF1α. Adding SI at the 5' UTR resulted in 200–250% increases in target gene expression with representative integration circuits (Fig. 4b, right side of the panel). With a single genomically integrated copy, only gRNA10 16× BS without SI (207%) and 8× and 16× BS with SI (223% and 521%, respectively) exhibited significantly stronger expression than seen with the integrated EF1α control (Fig. 4b). Similar to what was observed with their episomal counterparts, when both the gRNA4 and gRNA10 series were chromosomally integrated into sLP, the correlations between the number of BS in the synthetic operator and gene expression levels were highly significant (Supplementary Fig. 10a, both  $p < 0.005$ ; Pearson correlation analysis).

To provide evidence that our crisprTF-based transcription system directly programs gene transcription, we measured the transcription levels of mKate in chromosomally integrated CHO cells. In line with mKate expression determined by flow cytometry, gene transcription of the gRNA10 series ranged from 5.7% (2× BS) to 833% (16× BS with SI) of EF1α (Supplementary Fig. 10b). Intriguingly, with the addition of the SI, the transcription of mKate increased 6.9 folds (8× BS) and 4.3 folds (16× BS), indicating that the SI may be able to upregulate gene

transcription. The correlation between the number of BS and gene transcription levels was markedly significant (Supplementary Fig. 10c,  $p = 0.0002$ ; Pearson correlation analysis).

To demonstrate that crisprTF circuit integration functions in human cells, we engineered an sLP in HEK293 cells and performed BxB1-mediated chromosomal integration with the gRNA10 series (Supplementary Fig. 11a). In sLP-HEK293 cells, gene expression of the gRNA10 series ranged from 17% (2× BS) to 401% (16× BS) of EF1α (Supplementary Fig. 11b), similar to what we recorded in sLP-CHO cells. The number of BS correlated strongly with gene expression levels in sLP-HEK293 cells integrated with the gRNA10 series (Supplementary Fig. 11c,  $p = 0.0007$ ; Pearson correlation analysis).

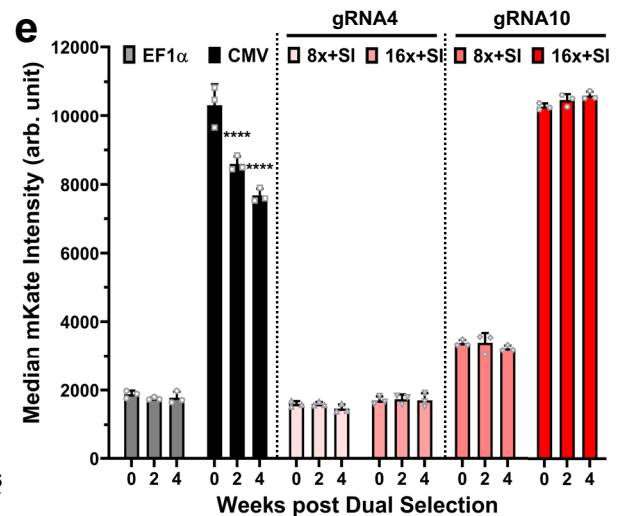
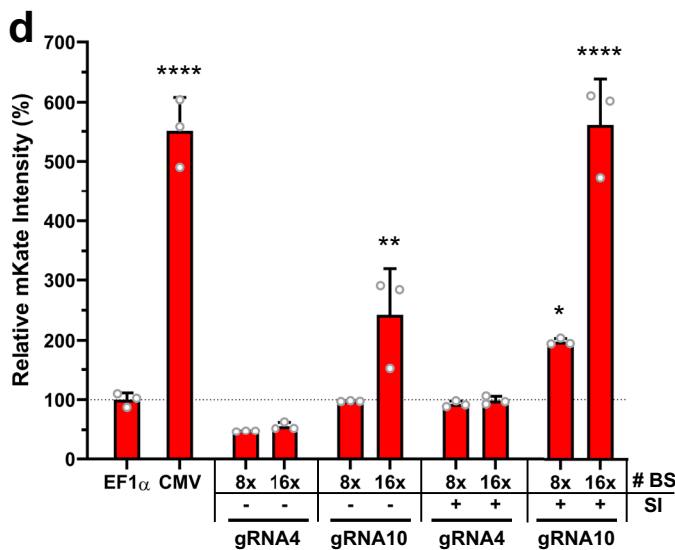
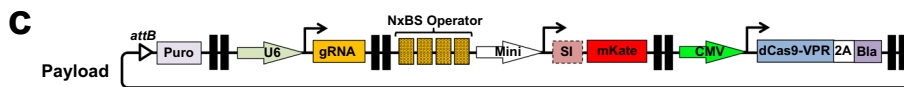
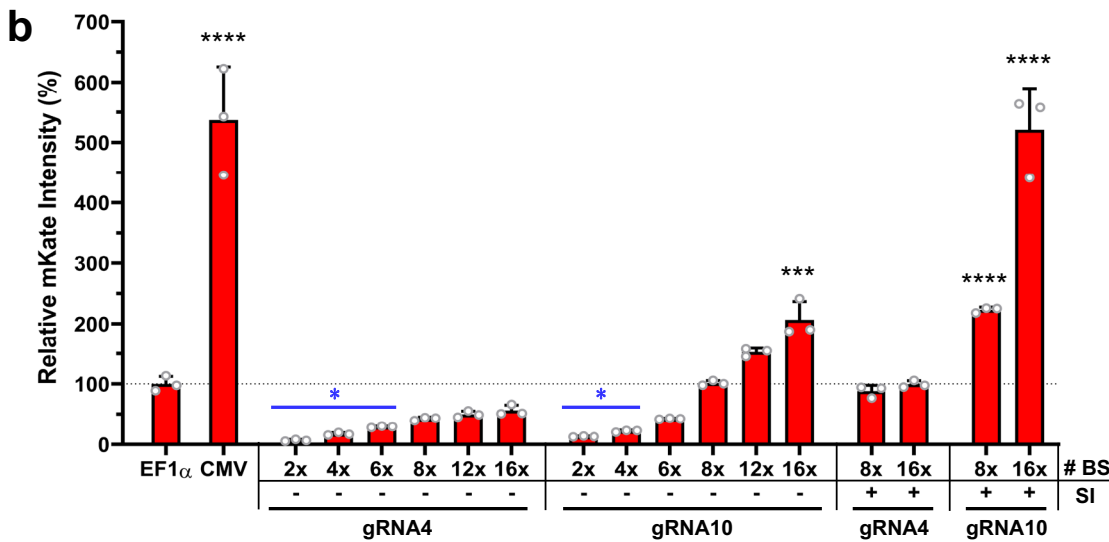
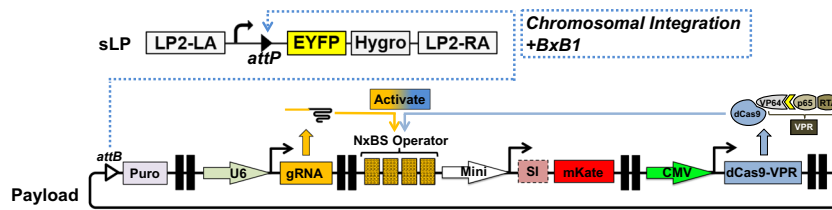
Nonetheless, by observing the expression profiles of puromycin-selected pools of crisprTF circuit integrants in sLP-CHO cells, we found that after 4 weeks of culturing, all four circuits from both the gRNA4 and gRNA10 series had notably decreased expression levels, suggesting the instability of gene expression (Supplementary Fig. 10d, all  $p < 0.01$  at 4 weeks post selection; two-way ANOVA with Dunnett's Multiple Comparison). To sustain long-term expression profiles, we incorporated an additional 3' flanking selection marker, blasticidin, into dCas9-VPR, linked by a 2A self-cleavage peptide (Fig. 4c)<sup>36</sup>. Flow cytometry results demonstrated gene expression levels that were similar or even slightly higher with most integration circuits immediately after one-time dual selection, compared with their unmodified counterparts (Fig. 4d). Unlike what was observed with the CMV control (Fig. 4e,  $p < 0.0001$  at 2 and 4 weeks; two-way ANOVA with Dunnett's Multiple Comparison), observations over four weeks revealed improved stability in expression levels, with no marked change for any of the four circuits examined (Fig. 4e, all  $p > 0.05$  at 2 and 4 weeks). Collectively, these data suggest that a flanking selection marker gene co-expressed with the *dCas9-VPR* gene may stabilize the crisprTF circuit post-integration in sLP-CHO cells.

### Modulation of human monoclonal antibody production

The controllable production of mAbs is desirable for many biomedical applications. To determine whether our transcriptional platform could be used for the precise control of antibody synthesis, we built integration gene circuits expressing a human mAb, JUG444, with the immunoglobulin kappa light chain (LC) and immunoglobulin gamma heavy chain (HC) genes separately expressed by the same gRNA10 operator as independent TUs (Fig. 5a). To avoid the potential instability of dCas9-VPR expression under the control of CMVp during long-term culture, we used adherent double LP (dLP) CHO cells, engineered with two distinct wild-type LPs: dLPI-1 and dLPI-2, to accommodate an additional copy of the *dCas9-VPR* gene (Fig. 5a)<sup>31,37</sup>. We first performed a targeted integration into dLPI-1 alone in dLP-CHO cells with a DNA payload encoding dCas9-VPR (Fig. 5a). Cells were selected with three antibiotics and then subjected to single-cell fluorescence-activated cell sorting (FACS) to isolate EYFP-/EBFP+ clones with the stably integrated dLPI-1 site and the free dLPI-2 site.

The functionality of clonally expanded cells was examined by integrating individual gRNA10 control circuits that co-expressed mKate and a monomeric blue fluorescent reporter, TagBFP, as well as another copy of the *dCas9-VPR* gene, as independent TUs into the dLPI-2 site (Fig. 5a). Dually integrated cells were selected with four antibiotics and then FACS-sorted into pools to expand them and evaluate reporter expression (Fig. 5a). We found that mKate and TagBFP driven by the same gRNA10 operators were simultaneously expressed at similar levels in two distinct configurations (8× BS without SI and 16× BS with SI), suggesting uniform transcriptional activation of both reporter TUs by two separate copies of the *dCas9-VPR* gene (Fig. 5b and Supplementary Fig. 12a). Next, we replaced the control circuit in the dLPI-2 site with individual gRNA10 mAb circuits that differentially express JUG444 (Fig. 5a). Similarly, cells with mAb circuits were selected and FACS-sorted into pools.

**a** Adherent CHO Cells with Single Landing Pad (sLP)



To examine the precision regulation and long-term stability of mAb production, we maintained cultures of sorted cell pools over five weeks and measured mAb concentration weekly, without constant antibiotic selection. We recorded differential JUG444 production in Week 1, driven by four distinct configurations: 8x or 16x BS, with or without SI (JUGAb1-4, Fig. 5c). Stable mAb production from a single mAb gene copy was sustained throughout the experimental duration with all four configurations (Fig. 5c). Strikingly, mAb production levels significantly correlated with the gRNA10 operator strengths documented in Fig. 4d (Fig. 5d,  $p < 0.05$

at all time points; Pearson correlation analysis). The strongest operator (16x BS with SI) yielded the highest mAb level, which was comparable to the level previously observed with the strong constitutive CMVp expressing two mAb gene copies<sup>30</sup>. The doubling time of JUGAb3 (16x BS) and JUGAb4 (16x BS with SI) increased to 22 and 34 hours, respectively, suggesting the influence of increased antibody production on cell growth in engineered dLPCHO cells (Fig. 5e). Overall, these data indicated that tunable and lasting control can be applied to express human antibodies and potentially other therapeutic proteins.

**Fig. 4 | Genomic integration and long-term precision gene expression in CHO landing pad cells.** **a** A schematic illustration of an integration gene circuit and BxB1 recombinase-mediated, site-specific integration in an engineered, adherent CHO cell line with a single landing pad (sLP). Positive integration control circuits had a central TU with an EF1 $\alpha$  or CMV promoter driving mKate expression and two flanking dummy TUs with no gene expression in the same architecture. **b** The mKate signal intensities of the chromosomally integrated crisprTF promoter circuits in sLP-CHO cells relative to the integrated EF1 $\alpha$  control circuit (one-way ANOVA with multiple comparisons corrected by Dunnett test; CMV, gRNA4 2 $\times$ , 4 $\times$ , and 6 $\times$  BS without SI, and gRNA10 2 $\times$ , 4 $\times$ , and 16 $\times$  BS without SI vs. EF1 $\alpha$ :  $p < 0.0001$ ,  $p = 0.0031$ ,  $p = 0.0111$ ,  $p = 0.0414$ ,  $p = 0.0066$ ,  $p = 0.0191$ , and  $p = 0.0006$ , respectively; gRNA10 8 $\times$  and 16 $\times$  BS with SI vs. EF1 $\alpha$ :  $p < 0.0001$  and  $p < 0.0001$ , respectively). **c** A schematic illustration of an integration circuit with the addition of a 3' flanking selection marker (blasticidin) into dCas9-VPR, linked by a 2A self-cleavage

peptide, to increase stability of target gene expression in long-term culture. A number of integration circuits from the gRNA4 and gRNA10 series were selected for the modification. **d** The mKate signal intensities of the chromosomally integrated crisprTF promoter circuits after dual selection with puromycin and blasticidin (one-way ANOVA with multiple comparisons corrected by Dunnett test; CMV and gRNA10 16 $\times$  BS without SI vs. EF1 $\alpha$ :  $p < 0.0001$  and  $p = 0.0020$ , respectively; gRNA10 8 $\times$  and 16 $\times$  BS with SI vs. EF1 $\alpha$ :  $p = 0.0446$  and  $p < 0.0001$ , respectively). **e** The mKate expression levels with two of the strongest circuits from each gRNA series over the course of 4 weeks following dual selection. All four circuits examined displayed no noteworthy change in mKate signals at 2 or 4 weeks (all  $p > 0.05$ ), similar to the EF1 $\alpha$  control (two-way ANOVA with multiple comparisons corrected by Dunnett test; CMV: 0 vs. 2 weeks:  $p < 0.0001$ , 0 vs. 4 weeks:  $p < 0.0001$ ). All data represent the mean  $\pm$  SD ( $n = 3$ ) (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ). Source data are provided as Source Data files.

## Programmable control of the human T-cell immune response against tumor cells

To explore the applications of our platform for cellular therapy, we built gRNA10 integration circuits with various configurations to program the secretion of anti-hPD1, an important immune checkpoint inhibitor widely used as a therapy in immuno-oncology<sup>29</sup>. The light chain and heavy chain of anti-hPD1 were separately expressed by the same gRNA10 operators as independent TUs (Fig. 6a). Similar to the JUG444 production, a copy of dCas9-VPR was pre-integrated into the dLPI-1 site of the dLP-CHO cells to avoid its complete silencing. Each of the four circuits, designed to differentially express anti-hPD1 (8 $\times$  or 16 $\times$  BS, with or without SI), was then integrated into the dLPI-2 site of the dLP-CHO cells, which had been selected and clonally sorted for dLPI-1 occupancy (PDIAb1-4, Fig. 6b). Dually integrated cells were selected and then sorted into pools with FACS for expansion. Quantification of anti-hPD1 secretion by sorted cells showed that the PDIAb4 group (16 $\times$  BS with SI) had a significantly higher titer of anti-hPD1 than all the other groups (Fig. 6b, all  $p < 0.001$ ; one-way ANOVA mixed-effects analysis with Tukey's multiple comparisons) and that the PDIAb3 group (16 $\times$  BS without SI) had a notably higher titer than the PDIAb1 group (8 $\times$  BS without SI) (Fig. 6b,  $p < 0.05$ ). A strong correlation was found between anti-hPD1 titers and crisprTF promoter strengths (Fig. 6c,  $p < 0.005$ ; Pearson correlation analysis).

To demonstrate the functionality of actively secreted anti-hPD1, we developed a three-way CHO-tumor-human T-cell co-culture system (Fig. 6d). We hypothesized that differentially programmed anti-hPD1 secretion would enhance T-cell effector function by correspondingly blocking the interactions between tumor cells and T cells, mediated by the engagement of PD1 with programmed death ligands. Sorted, dually integrated dLP-CHO cells with individual configurations expressing anti-hPD1 were first seeded. Pre-activated human T cells and ovarian cancer cells expressing a surface-displayed T-cell engager were then added to the attached dLP-CHO cells<sup>10</sup>. IFN- $\gamma$  production by T cells was measured at 24 hours post-co-culture as a marker of T-cell activation (Fig. 6d). T cells in the PDIAb3 and PDIAb4 groups produced significantly more IFN- $\gamma$  than the control group (Fig. 6e,  $p < 0.05$  and  $p < 0.01$ , respectively; one-way ANOVA with Dunnett's Multiple Comparison), corresponding to the higher titers of anti-hPD1 secretion. IFN- $\gamma$  production also notably correlated with anti-hPD1 titers in dLP-CHO cell cultures prior to the start of co-culturing (Supplementary Fig. 12b,  $p < 0.05$ ; Pearson correlation analysis). Thus, this transcription system can be used to produce functional proteins of clinical interest and perturb cellular phenotypes in a highly regulated and precise manner.

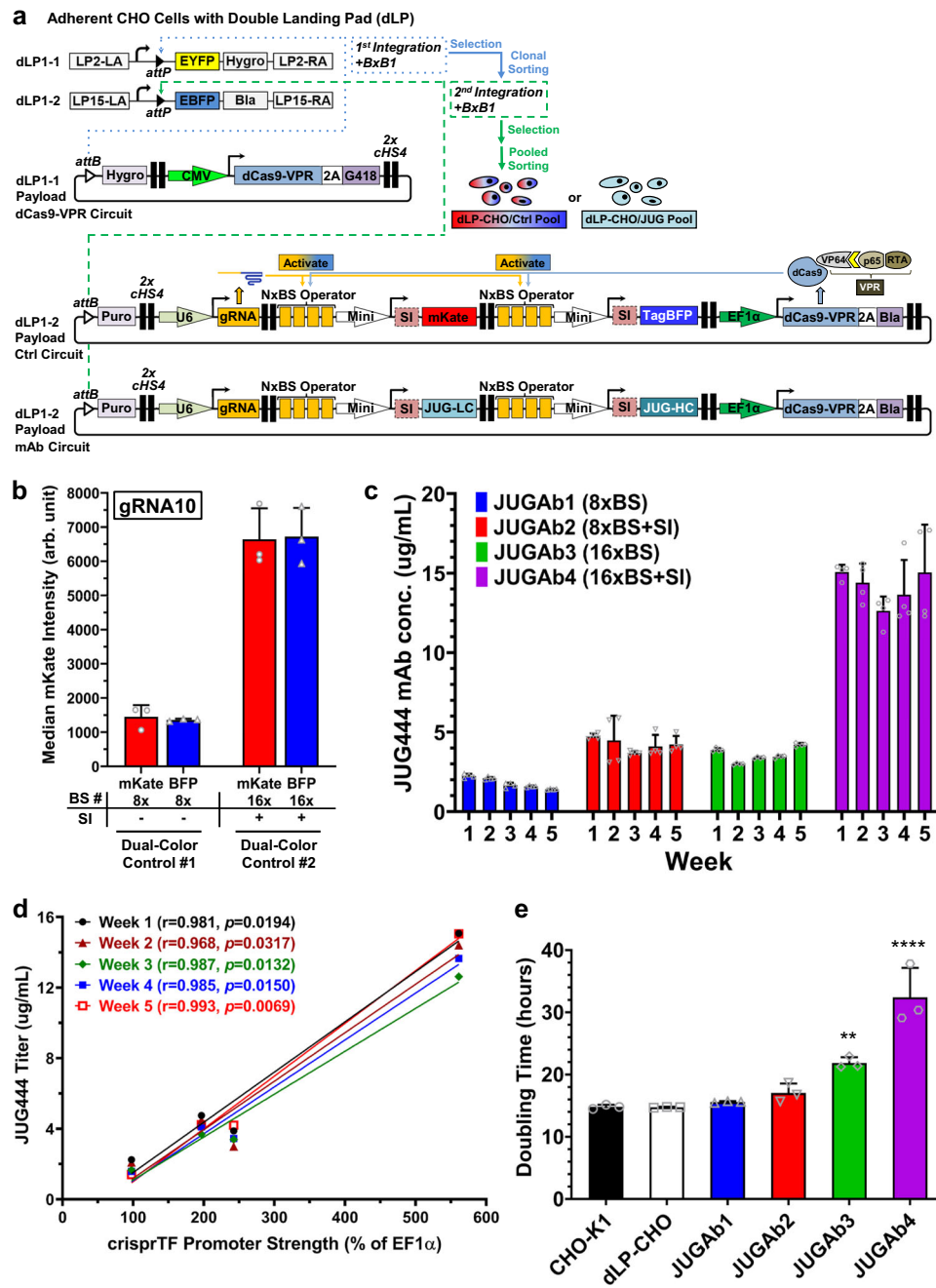
## Discussion

We have built and characterized a crisprTF promoter system for the programmable regulation of gene expression in mammalian cells. This system functions consistently across multiple mammalian cell types and is, therefore, sufficiently versatile to be used in a variety of cellular models for biomedicine and biomanufacturing.

To demonstrate the generalizability of the system's gene regulatory activity, we used both episomal vectors in transiently transfected cells and site-specific, stably integrated chromosomal constructs in LP-engineered cells. We modulated three key parameters: (1) the gRNA sequence; (2) the number of gRNA BS repeats in the operator; and (3) the CRISPR-based transcriptional activator (different crisprTFs, different selection markers in the crisprTF, or an additional copy of the crisprTF in LP-engineered cells). We also incorporated genetic elements to enhance gene expression, including SAM, an extra gRNA TU, extra NLS, and SI. Systematic characterization of these constructs resulted in >1000-fold range of gene expression levels. The strongest of our synthetic promoters, composed of 16 $\times$  gRNA BS repeats, was significantly stronger than CMVp, one of the strongest constitutive promoters currently used for mammalian applications<sup>13</sup>. Recent systematic investigation of mismatched single gRNAs (sgRNAs) has revealed sgRNA-DNA interaction rules controlling endogenous gene expression and correlated phenotype<sup>38</sup>. In agreement with previous findings, our introduction of a 2-bp mutation that moderately altered the GC ratio in the PAM-proximal seed region of gRNA10 and its BS markedly changed its gene expression profiles, suggesting the importance of the gRNA seed sequence in controlling our crisprTF promoter-driven transcriptional activities. We anticipate that additional large-scale experimental screening, combined with computational modeling or machine learning, could be used to program gene expression even more precisely.

Various mechanisms of controlling gene expression at the transcriptional level in eukaryotes have been identified<sup>39,40</sup>. Core promoters and neighboring genetic elements play instrumental roles in consolidating complex cascades of signaling events involved in transcription, impacting gene expression<sup>39</sup>. The complexity of combinatorial interactions among constituent TF regulatory elements hinders the creation of synthetic mammalian promoters using natural motifs. Although this issue has been gradually resolved in silico, it remains challenging to overcome the context- or species-dependency for the de novo design of mammalian promoters<sup>40</sup>. Positive transcriptional elements have been found between -350 and -40 bp relative to the transcription start sites (TSS) in many of the tested human promoters, whereas negative regulatory elements are more likely to be located -350 to -1000 bp upstream of the TSS<sup>34</sup>. Therefore, we positioned our synthetic operators at up to roughly -400 bp upstream of the TSS. Our data are consistent with previous reports suggesting that longer transcriptional bursts and thus higher expression levels might be achieved with synthetic transcriptional activators by arraying multiple BS upstream of a given promoter<sup>28,41</sup>. The results of our episomal tests with all three gRNA series established strong linear correlations between gene expression levels and the number of BS (up to 16 $\times$ ) in the operators in multiple mammalian species and cell types (Supplementary Figs. 5 and 8). We found similar significant correlations with two gRNA series when they were genomically integrated in LP-engineered cells (Supplementary Figs. 10a and 11c).

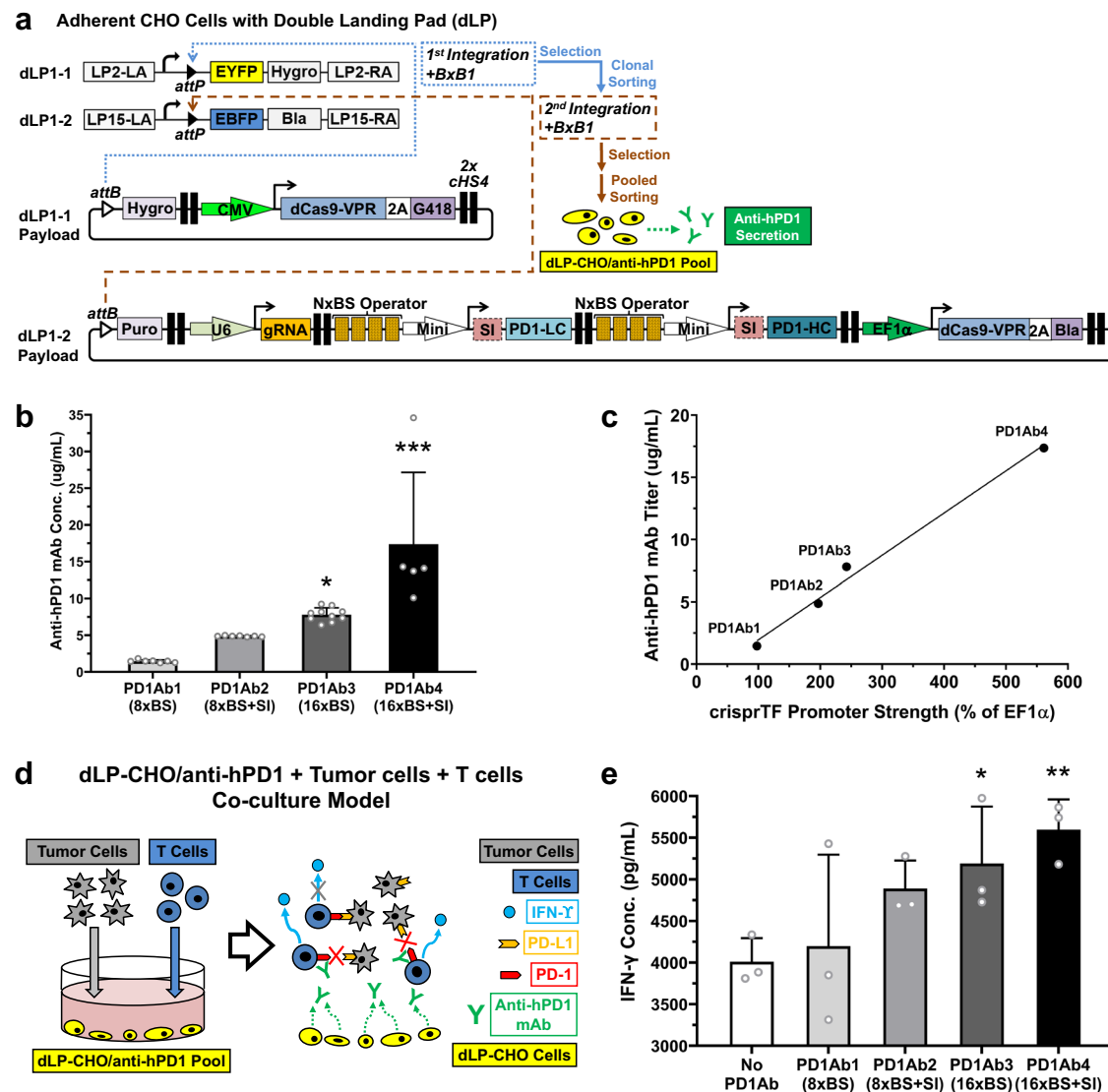




**Fig. 5 | Precision control of human monoclonal antibody (mAb) production.**

**a** Schematic illustrations of the sequential, site-specific genomic integration of two payload gene circuits into the CHO cells engineered with a double landing pad (dLP) for human mAb production. First, a synthetic gene circuit encoding one copy of *dCas9-VPR* gene and 2 flanking mammalian selection marker genes, *hygromycin* (5'-end) and *G418* (3'-end), was integrated into dLP1-1 site with BxB1 integrase. Selected cells with single dLP1-1 occupancy were clonally sorted based on EYFP<sup>-</sup>/EBFP<sup>+</sup> signals. A single clone with the most consistent outputs of dCas9-VPR and EBFP was chosen for the second BxB1-mediated integration targeting the free dLP1-2 site. The second integration gene circuit contained independent TUs encoding either *mKate* and *TagBFP* reporter genes (control circuit) or the light chain and heavy chain genes of a human mAb *JUG444* (mAb circuit) as well as gRNA10, one copy of *dCas9-VPR* gene, and two additional flanking selection marker genes: *puromycin* (5'-end) and *blasticidin* (3'-end). Integrated cells selected with four antibiotics were then pool-sorted based on EYFP<sup>-</sup>/EBFP<sup>-</sup> signals. **b** The mKate and TagBFP expression of the integrated payload control circuits in dLP-CHO cells with

two distinct configurations (8 $\times$  BS without SI and 16 $\times$  BS with SI). **c** The mAb production of the integrated payload circuits to express the light chain and heavy chain of *JUG444* with four gRNA10 operator configurations: *JUGAb1* (8 $\times$  BS), *JUGAb2* (8 $\times$  BS with SI), *JUGAb3* (16 $\times$  BS), and *JUGAb4* (16 $\times$  BS with SI). Octet mAb titer quantitation over five weeks showed stable, differential *JUG444* production by all four integration circuits. **d** Pearson correlation analysis to determine the relationship between *JUG444* mAb titers and crisprTF promoter strengths over the course of five weeks. The Pearson correlation coefficients (*r*) at Week 1 through Week 5 were:  $r = 0.98$  ( $R^2 = 0.96$ ,  $p = 0.0194$ ),  $r = 0.97$  ( $R^2 = 0.94$ ,  $p = 0.0317$ ),  $r = 0.99$  ( $R^2 = 0.97$ ,  $p = 0.0132$ ),  $r = 0.99$  ( $R^2 = 0.97$ ,  $p = 0.015$ ), and  $r = 0.99$  ( $R^2 = 0.99$ ,  $p = 0.0069$ ), respectively. **e** The doubling time of mAb-producing cell lines (*JUGAb3* and *JUGAb4* vs. dLP-CHO:  $p = 0.0049$  and  $p < 0.0001$ , respectively). All data represent the mean  $\pm$  SD ( $n = 3$ ) (one-way ANOVA with multiple comparisons corrected by Dunnett test; \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ ). Source data are provided as Source Data files.



**Fig. 6 | Programmable control of anti-hPD1 secretion and the human T-cell anti-tumor response.** **a** Schematic illustration of engineering the anti-hPD1-secreting dLP-CHO cells with sequential and site-specific integration of two payload gene circuits with BxB1 integrase. Clonally sorted EYFP<sup>+</sup>/EBFP<sup>+</sup> dLP-CHO cells stably integrated with a gene circuit encoding one copy of *dCas9-VPR* gene and flanking selection marker genes in the dLP1-1 site were used for the second BxB1-mediated integration. The free dLP1-2 site was integrated with a gene circuit containing independent TUs that encoded: the 5' flanking *puromycin* gene, gRNA10, the light chain and heavy chain genes of *anti-hPD1* driven by the same gRNA10 operators, one copy of *dCas9-VPR* gene, and the 3' flanking *blastidicin* gene linked to *dCas9-VPR* gene by a 2A self-cleavage peptide sequence. Dually integrated cells were selected with four antibiotics and then subjected to pooled cell sorting based on EYFP<sup>+</sup>/EBFP<sup>+</sup> signals. **b** Octet mAb titer quantitation showed differential anti-hPD1 secretion programmed by four distinct configurations of gRNA10 operators: PD1Ab1 (8 × BS), PD1Ab2 (8 × BS with SI), PD1Ab3 (16 × BS), and PD1Ab4 (16 × BS with SI) (one-way ANOVA mixed-effects analysis with multiple comparisons corrected by

Tukey test; PD1Ab3 and PD1Ab4 vs. PD1Ab1:  $p = 0.0160$  and  $p < 0.0001$ , respectively; PD1Ab4 vs. PD1Ab2:  $p < 0.0001$ ; PD1Ab4 vs. PD1Ab3:  $p = 0.0009$ ). **c** Pearson correlation analysis revealed that anti-hPD1 titers strongly correlated with crisprTF promoter strengths ( $r = 0.99$ ,  $R^2 = 0.99$ ,  $p = 0.0051$ ). **d** Schematic diagram of CHO-tumor-T-cell co-culture system to evaluate the functionality of anti-hPD1 and to explore the utility of our crisprTF promoter platform for cellular therapy. dLP-CHO cells engineered with one of the above four configurations for anti-hPD1 secretion were first seeded for 48 hours. The control group was seeded with EYFP<sup>+</sup>/EBFP<sup>+</sup> dLP-CHO cells with no anti-hPD1 payload circuit. Pre-activated human T cells and human ovarian cancer cells (OVCAR8) expressing a surface T-cell engager were subsequently seeded in each well with the attached dLP-CHO cell populations. **e** Quantification of IFN- $\gamma$  concentrations in the media at 24 hours post-co-culture by ELISA revealed tunable IFN- $\gamma$  production by T cells (one-way ANOVA with multiple comparisons corrected by Dunnett test; PD1Ab3 and PD1Ab4 vs. No PD1Ab:  $p = 0.0248$  and  $p = 0.0049$ , respectively). All data represent the mean  $\pm$  SD ( $n \geq 3$ ) ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ). Source data are provided as Source Data files.

Besides tuning transcriptional activities by altering gRNA sequences and operator strengths, we discovered several compatible genetic control elements that synergistically augment gene expression. When combined with *dCas9-VPR*, SAM moderately increased expression at medium, but not high, levels episomally. We examined the compatibility of SAM with *dCas9-VPR* when genomically integrated, by assembling large gRNA10 circuits that contained independent TUs encoding both systems (Supplementary Fig. 13a). Surprisingly, the VPR-SAM hybrid system at medium expression level generated no further

increase in gene expression; in fact, at high expression level, VPR-SAM actually decreased expression (Supplementary Fig. 13b). This seemingly paradoxical result suggested possible competition for transcriptional resources among components within the hybrid system, for example, overlapping TADs encoded in *SAM* and *dCas9-VPR* genes, including VP64 and p65<sup>42,43</sup>. It remains to be tested whether alternative CRISPR-based gene activation strategies, such as the CRISPR-assisted *trans* enhancer, would be compatible for combinatorial use with our system for auxiliary augmentation of transcription<sup>44</sup>.

On the other hand, introns, which enhance gene expression in mammalian cells when applied as a positive regulatory element<sup>45</sup>, have been proposed to control post-transcriptional RNA processing or transport<sup>35</sup>. Here, the incorporation of an SI immediately downstream of the operator in the 5' UTR of the target gene<sup>46</sup> increased gene expression at least 200% with nearly all tested constructs, whether episomal or chromosomal, suggesting a high degree of compatibility between the transcriptional control by the crisprTF promoters and the intron-mediated enhancement by SI<sup>47</sup>. RT-qPCR data also suggest that an SI in the 5' UTR directly contributes to gene transcription in our crisprTF promoter system, consistent with gene expression enhancement by 5' UTR sequences<sup>48</sup>. However, overly strong target gene expression may impact cell proliferation (Fig. 5e). Thus, it is essential to balance target gene expression with cellular homeostasis when employing strong crisprTF promoters. Moreover, the large size of the crisprTF promoter constructs, especially with additional genetic regulatory elements, limits their use with LP-engineered cells for long-term expression. Broader applications in biomedicine may require minimizing construct size.

CHO cells, approved by regulatory agencies, are widely used by the pharmaceutical industry for the biomanufacturing of recombinant therapeutic proteins with human-like glycosylation profiles<sup>49,50</sup>. CMVp and Simian virus 40 early promoter are frequently used in CHO cells to produce high titers of therapeutic proteins<sup>51</sup>. Yet, protein-producing cell lines often exhibit a wide range of gene expression profiles, even at the clonal level, and the mechanisms that drive viral promoters in eukaryotic cells are ill-defined<sup>52,53</sup>. Natural promoters have often evolved synchronously with functionalities that depend on specific genetic contexts, so promoter activity may not transfer across various genes, conditions, or species<sup>39,53</sup>. The growing number of biotherapeutic proteins in development has created an increasing demand for efficient long-term protein expression systems for biomanufacturing, ideally with built-in programmable control for improved consistency and predictable yields.

Coupled with the LP technology, we demonstrated the long-term stability of precisely tuned human mAb production with the crisprTF promoter system. Antibody yields obtained with the strongest crisprTF circuit expressing a single mAb LC/HC cassette were similar to those obtained with two or six mAb cassettes driven by CMV or EF1 $\alpha$  promoters, respectively, in LP-engineered CHO cells<sup>30</sup>. The high production efficiency of our strong crisprTF circuits, along with their built-in modularity and scalability, suggests that they can be adapted to achieve industrially relevant mAb titers. Our platform may also be applied, more generally, to CHO strains optimized for industrial bioreactor conditions to: (1) increase target protein production in a predictable and controlled manner; (2) produce multiple target proteins simultaneously with specific ratios; and (3) strike a balance between cell behaviors and protein production in long-term cultures.

To exogenously modulate target gene expression by controlling gRNA transcription levels, we incorporated two types of small molecule-inducible switches into the RNA polymerase III (Pol III) promoter (Supplementary Fig. 14a). The results in the episomal context indicate that gRNA transcription could be tuned by adjusting the concentrations of the inducers doxycycline (Dox) and isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) for Tet and Lac operons, respectively (Supplementary Fig. 14b, c), following gene expression kinetics at different time points (Supplementary Fig. 14d). Thus, an additional layer of target gene tunability could be achieved in our crisprTF promoter system by using inducible Pol III promoters.

In conclusion, we have built a tunable, scalable, and sustainable gene expression system of crisprTF-based regulatory elements. This platform should enable programmable, multiplexed gene modulation for broad applications, such as mammalian synthetic biology, biomanufacturing, and precision medicine.

## Methods

### Molecular cloning and genetic circuit construction

All genetic circuits in this study were constructed using a modular Gateway-Gibson assembly method<sup>33,54</sup>. Briefly, gRNAs and related sequences were commercially synthesized (Integrated DNA Technologies IDT and GenScript) and then cloned into corresponding entry vectors using In-Fusion cloning (Takara Bio, San Jose, CA, USA) or a one-step Gibson reaction with the in-house master mix. Other genetic parts were cloned into corresponding entry vectors using the same approach. Multi-site LR reactions were performed between a promoter entry vector flanked by *attL4* and *attR1* recombination sites, a gene entry vector flanked by *attL1* and *attL2* recombination sites, and pZDonor\_Seq(*n*)-GTW-Seq(*n*+1)*R4**R2* destination vectors containing a Gateway cassette (chloramphenicol resistance and *ccdB* genes flanked by *attR4* and *attR2* recombination sites) to generate positional expression vectors [i.e., independent transcriptional units (TUs) in different position vectors used for the subsequent Gibson assembly]<sup>33</sup>. Gibson reactions were performed with the Gibson Assembly Ultra Kit (SGI-DNA, La Jolla, CA, USA) at 50 °C for 60 min, using equimolar concentrations (~40–60 fmol per 10  $\mu$ L reaction) of column-purified positional expression vectors (cleaved with I-SceI), a matching adaptor vector (cleaved with XbaI and XhoI), and a carrier vector with BxB1-*attB* or BxB1-GA-*attB* integration sites (cleaved with FseI). Gibson-assembled constructs were diluted at 1:4 and used to transform *E. coli* 10 G electrocompetent cells (60080-2, Lucigen, Middleton, WI, US). Cells were selected with appropriate antibiotics on solid and liquid culture; plasmids were column-purified with QIAprep Spin Miniprep Kit (27106, Qiagen, Hilden, Germany). After verification by restriction mapping analysis and Sanger sequencing of payload TUs, correctly assembled constructs were expanded in 25 mL liquid culture with Stb3 chemically competent cells (C737303, Invitrogen) and column-purified with QIAGEN Plasmid Plus Midi Kit (12945, Qiagen). A schematic diagram of the gene circuit construction methodology is shown in Fig. 1a. Human elongation factor-1 alpha (EF1 $\alpha$ ) promoter and mouse CMVp were used as constitutive promoter controls. Sp-dCas9-VPR was a gift from George Church (Addgene plasmid 63798)<sup>22</sup>. pRRM.CMVenh.gp91.Syn.Intron.eGFP was a gift from Didier Trono (Addgene plasmid 30470), which was used for cloning the 230-bp synthetic intron sequence<sup>55</sup>. MS2-P65-HSF1\_GFP was a gift from Feng Zhang (Addgene plasmid 61423), which was used for cloning the *MS2-P65-HSF1* gene sequences<sup>26</sup>. Genetic sequences and construct maps were viewed and edited with Geneious 9.0 (Dotmatrix, Boston, MA, USA). The sequences of genetic parts and the list of plasmids are provided in Supplementary Tables 1 and 2, respectively.

### LP vector construction

LP donor vectors for stable CRISPR-Cas9-mediated integration into the Chinese hamster ovary (CHO) cell genomes were constructed as previously reported<sup>20</sup>. Briefly, homology arm sequences for LP1-2, LP2, LP8, LP15, and LP20 integration sites (each arm approximately 0.5–1 kb long) were synthesized as a single gBlock (IDT) containing a PmeI restriction site between the left and right homology arms and unique BsaI cleavage sites in the 5' and 3' termini for Golden Gate cloning<sup>30</sup>. Each gBlock was cloned into a pIDTsmart vector modified to contain compatible BsaI cloning sites. LP cassettes containing hEF1 $\alpha$ -*attP*-BxB1-*EYFP*-P2A-*Hygro* (cassette 1), hEF1 $\alpha$ -*attP*-BxB1-*EBFP*-P2A-*Bla* (cassette 2), or hEF1 $\alpha$ -*attP*-BxB1-GA-*EYFP*-P2A-*Hygro* (cassette 3) were constructed using modular Gateway-Gibson cloning as previously described<sup>33,54</sup>. LP cassettes were cloned into PmeI-linearized pIDTsmart backbones between the left and right homology arms using In-Fusion cloning (Takara Bio USA). Using this approach, we generated the following LP donor vectors: LP1-2-cassette 1, LP2-cassette 1, LP2-cassette 2, LP8-cassette 2, LP15-cassette 2, LP20-cassette 2, and LP20-cassette 3<sup>4,30</sup>.

For human embryonic kidney (HEK)–293 cells, the LP donor vector was integrated into the *AAVSI* locus using engineered zinc finger nucleases (ZFN) as previously described<sup>54</sup>. Briefly, 800 bp homology arm sequences for the *AAVSI* locus, on the 5' and 3'-ends of the ZFN cut site, were cloned into flanking position vectors. A center position vector was assembled to encode a double *chs4* core insulator, the CAG promoter, *attP* BxB1, *EYFP-2A-Hygromycin*, a rabbit *beta-globin polyadenylation* signal, and another double *chs4* core insulator. These three-position vectors were verified and then assembled into the LP Shuttle Vector. A DNA fragment encoding the two DNA-binding ZF domains for the *AAVSI* locus, separated by a 2A self-cleavage peptide sequence, was synthesized (GeneArt, Regensburg, Germany) and cloned into an expression vector with CAG promoter via Gateway assembly to create the ZFN-expressing vector.

### Cell culture and plasmid transfection

Adherent wild-type CHO-K1 cells (CCL-61, American Type Culture Collection (ATCC), Manassas, VA, USA) and engineered CHO cells were maintained in complete Ham's F12K (cHams-F12K) medium (30-2004, ATCC) containing 10% fetal bovine serum (FBS) (F2442, Sigma-Aldrich, Saint Louis, MO, USA), 1% HyClone non-essential amino acids (SH30238.01, GE Healthcare Life Sciences, Piscataway, NJ, USA), and 1% penicillin/streptomycin (P/S) (15140-122, Gibco by Life Sciences). Cells were grown in a humidified 37 °C incubator with 5% CO<sub>2</sub> and passaged every 2–3 days. Transfections were carried out with the Neon electroporation system (MPK5000, Invitrogen) with 10 µl Neon tips. Briefly, 1 × 10<sup>5</sup> cells were suspended in R buffer and mixed well with 250 ng of individual experimental and transfection marker plasmids for 10–15 min at room temperature (RT). To ensure that all samples had the same total amount of plasmids, we supplemented the negative or positive control samples with a dummy plasmid composed of an identical expression vector backbone and a non-functional insert sequence. Cells were then electroporated with the setting of 1560 V/5 ms/10 pulses. Transfected cells were immediately transferred to a 24-well plate containing 1 mL complete culture medium without any antimicrobial reagents. Fluorescence-activated cell sorting (FACS) analysis was performed at 48 hours post-transfection. Experiments were performed with independently transfected biological triplicates.

HEK293T cells (CRL-3216), mouse C2C12 myoblasts (CRL-1772), and rat H9c2 cardiac myoblasts (CRL-1446) were purchased from ATCC. All three cell types were thawed and expanded according to the manufacturer's instructions and subsequently maintained in Dulbecco's Modified Eagle medium (DMEM; 10569-044, Thermo Fisher Scientific, Waltham, MA, USA) containing 10% FBS and 1% P/S. Transfections were carried out with ViaFect transfection reagent (E4982, Promega, Madison, WI, USA). Briefly, cells were washed once with PBS, trypsinized, and re-suspended with complete culture medium at 1 × 10<sup>5</sup> cells per 100 µL. Experimental and transfection marker plasmids (250 ng each) were added to 100 µL plain Opti-MEM medium (31985-070, Thermo Fisher Scientific) and mixed well with ViaFect at a 1:3 ratio (total DNA by weight: ViaFect by volume), according to the manufacturer's recommendation. After incubating at RT for 10 min, 1 × 10<sup>5</sup> cells were added to the plasmid/ViaFect mixture and mixed well. Transfected cells were immediately transferred to a 24-well plate containing 0.8 mL complete culture medium without any antimicrobial reagents. FACS analysis was performed at 48 hours post-transfection.

The PGP1 hiPSCs were a gift from George Church (Harvard University, MA, USA) and cultivated in compliance with the MIT IRB protocol (BRR 836)<sup>56</sup>. Briefly, culture plates were coated with Matrigel Matrix (354277, Corning Inc., Corning, NY, USA), following the manufacturer's protocol, and incubated for >1 hour at 37 °C prior to plating. hiPSCs were cultured in StemFlex medium (A3349401), split with a five-minute treatment of StemPro Accutase (A1110501, both from Thermo Fisher Scientific) at 37 °C every 3–4 days, and plated on Matrigel-

coated plates at an appropriate density with ROCK inhibitor (1254, Tocris Bioscience, UK) at a final concentration of 10 µM for 24 hours to improve viability. hiPSCs were reverse transfected with plasmids using Lipofectamine Stem transfection reagent (STEM00001) and Opti-MEM (31985088, both from Thermo Fisher Scientific) in 12-well plates in triplicate according to the manufacturer's recommended protocol. Briefly, cells were cultured until they reached 40–50% confluence at transfection. Experimental and transfection marker plasmids (250 ng each) were added to 25 µL plain Opti-MEM I medium and mixed with 25 µL diluted Lipofectamine Stem transfection reagent (1:1 ratio), following the manufacturer's protocol. After incubating at RT for 10 min, 50 µL DNA-lipid complexes were applied to each well. At 48 hours post-transfection with 1 × 10<sup>5</sup> cells per well, hiPSCs were harvested with StemPro Accutase, washed three times with PBS (10010-023, Thermo Fisher Scientific), subjected to filtration with a 35-micron cell strainer, and then run on a flow cytometer for FACS analysis.

### Engineering LP cells

Single- and multi-LP CHO cell lines with engineered LP1-2, LP2, LP8, LP15, and LP20 loci were constructed in adherent CHO-K1 cells by homologous recombination with CRISPR/Cas9 as described<sup>30</sup>. Briefly, targeted integrations were performed by co-transfecting 500 ng of circular LP donor vector with 40 ng of px330-U6-chimeric\_BB-CBh-hSpCas9 vector, a gift from Feng Zhang (Addgene plasmid 42230)<sup>57</sup>, and 150 ng of U6-gRNA GeneArt DNA String (Thermo Fisher Scientific). Roughly 10<sup>5</sup> cells were transfected in triplicate with the Neon electroporation system with 10 µl Neon tips and seeded in 24-well plate. Cells were then transferred to a six-well plate at 3 days post-transfection and subjected to antibiotic selection with either hygromycin (600 µg/ml; ant-hg-1, InvivoGen, San Diego, CA, USA) or blasticidin (10 µg/ml; ant-bl-1, InvivoGen) for two weeks, followed by clonal cell sorting with FACS. Clonal cells were verified with diagnostic PCR using locus-specific and LP-specific primers (for on-target integration) and backbone-specific primers (for off-target integration). Single-cell clones exhibiting locus-specific and backbone-free integration, as well as stable and homogenous LP expression, were expanded, banked, and subsequently used for gene circuit integration.

The single-LP HEK293 cell line was constructed with an engineered *AAVSI* locus by homologous recombination with ZFN in wild-type adherent HEK293 cells as described<sup>54</sup>. Briefly, HEK293 cells were co-transfected with equimolar amounts of the ZFN-expressing vector and the LP Shuttle Vector, allowed to recover for 72 hours, and then selected with 200 µg/ml hygromycin for 2 weeks. Single-cell clones exhibiting locus-specific integration and stable LP expression were generated by serial dilutions of the surviving population and subsequently expanded for gene circuit integration.

### Chromosomal integration of gene circuits

For chromosomal integration of gene circuits in adherent, engineered CHO and HEK293 cells, cells were independently transfected in triplicate with 500 ng of the BxB1 integrase-expressing plasmid (pEXPR-CAG-BxB1) and 500 ng DNA of the payload plasmid, using the Neon electroporation system with the same setting as described above. Transfected cells were immediately transferred to a 24-well plate containing 1 mL complete culture medium per well without any antimicrobial reagents. At 3 days post-transfection, cells were transferred to a 6-well plate with 3 mL complete culture medium per well, followed by one of the following one-time cell selection strategies (depending on the individual payload construct designs described in the Results section and figures): (1) for single integrants in single-LP (sLP), cells were selected with puromycin only (8 µg/ml for CHO and 1 µg/ml for HEK293; A11138-03, Thermo Fisher Scientific) for gene circuits with a single selection marker, or with both puromycin (8 µg/ml) and blasticidin (10 µg/ml) for gene circuits with both selection markers; (2) for single integrants in double LP (dLP), cells were selected with either puromycin

(8  $\mu\text{g}/\text{mL}$ ) alone or with hygromycin (250  $\mu\text{g}/\text{mL}$ ) and G418 (250  $\mu\text{g}/\text{mL}$ ) for the payload integrated in the LP2 locus, and with blasticidin (10  $\mu\text{g}/\text{mL}$ ) for the non-integrated, empty LP15 locus; (3) for double integrants in dLP, cells were selected with puromycin (8  $\mu\text{g}/\text{mL}$ ), blasticidin (10  $\mu\text{g}/\text{mL}$ ), hygromycin (250  $\mu\text{g}/\text{mL}$ ), and G418 (250  $\mu\text{g}/\text{mL}$ ; anti-gn-1, InvivoGen). Following 10 days of selection, correct integration was confirmed by FACS, indicated by the complete disappearance of native LP fluorescence: either enhanced yellow fluorescent protein (EYFP), or enhanced blue fluorescent protein (EBFP), or both. Cells with chromosomally integrated gene circuits were maintained as pooled populations after selection. Cell viability and density were monitored with a Vi-CELL automated cell viability analyzer (Beckman-Coulter).

### Fluorescent imaging

All fluorescent images were taken with the EVOS FL Auto cell imaging system (AMAFD1000, Thermo Fisher Scientific) equipped with multiple LED light cubes, including TagBFP, GFP, YFP, RFP, and Texas Red. Fluorescent images of the same color were taken with the same exposure settings. Scale bars were directly printed in the images.

### Flow cytometry and cell sorting

Cells were analyzed with an LSRFortessa flow cytometer, equipped with 405, 488, and 561 nm lasers (BD Biosciences, San Jose, CA, USA). Thirty thousand events per sample were collected for analysis of the median signal intensity of the transfected population, using a 488 nm laser and 530/30 nm bandpass filter for EYFP and a 405 nm laser, 450/50 filter for EBFP. Sphero rainbow calibration particles, 8 peaks (RCP-30-5A, Spherotech, Lake Forest, IL, USA) were used for instrument normalization and MEFL calculation. Median fluorescent intensity of the entire transfected population (i.e., all cells positive for both experimental and transfection marker signals) or the entire chromosomally integrated cell pool was measured in histograms (for single color) or dot-plots (for two or more colors). The same gating was used in all experiments (cutoff value set at 200 A.U. on the  $x$  axis of the histogram or on both  $x$ - and  $y$  axis of the dot-plot). Data were analyzed with FACSDiva software version 8.0.1 (BD Biosciences), FlowJo (FlowJo LLC, Ashland, OR, USA), and FCS Express 6 (De Novo Software, Pasadena, CA, USA). Flow cytometry data were largely normalized to the EF1 $\alpha$  control and presented as relative median signal intensity (%). Some fluorescent intensity data shown as A.U. (not % of the EF1 $\alpha$  control) represent the original readout of the fluorescent intensity from the flow cytometer without normalizing to the control group. Cell sorting was performed on a FACS Aria cell sorter (BD Biosciences) in the Swanson Biotechnology Center Flow Cytometry Facility at the Koch Institute (Cambridge, MA, USA). Untransfected CHO-K1 and HEK293 cells and unintegrated single- and multi-LP cells were used to set the gating. Different selection and sorting schemes were applied to target payload integration into specific LP sites, including: (1) single-positive/single negative (EYFP+/EBFP- or EYFP-/EBFP+), (2) double-negative (EYFP-/EBFP-), or (3) double-negative/mKate-positive (EYFP-/EBFP-/mKate+) to select multi-LP cells with the payload(s) integrated into each designated LP. For pooled cell sorting, all sorted cells were initially seeded at 5000–10,000 cells/ $\text{cm}^2$  in 24-well plates with 1 mL complete culture medium containing no antibiotics. For clonal cell sorting, single cells were initially sorted into flat-bottom 96-well plates with 100  $\mu\text{L}$  complete culture medium containing no antibiotics for clonal expansion and subsequently selected and expanded in 24-well and six-well plates and T75 flasks.

### RNA extraction and RT-qPCR analysis

Chromosomally integrated CHO cells ( $3 \times 10^6$ ) were collected for RNA extraction. Total RNA was extracted using the RNeasy Plus kit (74034, Qiagen) according to the manufacturer's instructions. cDNA synthesis was performed using an iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). Quantitative real-time PCR (RT-qPCR) was carried out in a

LightCycler 96 System (Roche, Basel, Switzerland) using a KAPA SYBR FAST qPCR kit (KAPA Biosystems, Wilmington, MA, USA) according to the manufacturer's instructions. The primer sequences of mKate and GAPDH for RT-qPCR are shown in Supplementary Table 1.

### Human antibody production and detection

A genetic circuit expressing one copy of dCas9-VPR was first integrated into dLP-CHO cell lines. Cells were selected once with hygromycin (250  $\mu\text{g}/\text{mL}$ ) and G418 (250  $\mu\text{g}/\text{mL}$ ) for 10 days, followed by FACS sorting into a pool of single-positive cells using the LP-fluorescent reporters. A second, crisprTF-driven gene circuit expressing a human mAb, either JUG444 (Pfizer, Andover, MA, USA) or an anti-hPD1 (5C4)<sup>10</sup>, was then integrated into the sorted cells. Doubly integrated cells were selected once by incubating them for 10 days with puromycin (8  $\mu\text{g}/\text{mL}$ ), blasticidin (10  $\mu\text{g}/\text{mL}$ ), hygromycin (250  $\mu\text{g}/\text{mL}$ ), and G418 (250  $\mu\text{g}/\text{mL}$ ), followed by FACS sorting into a pool of double-negative (EYFP-/EBFP-) cells using the LP-fluorescent reporters. All experiments were performed with three independently transfected biological replicates. For mAb expression analysis,  $1.5 \times 10^5$  cells were seeded in 24-well plates with 0.5 ml cHams-F12K medium per well without any antibiotics and maintained at 37 °C for 4 days. Conditioned media were collected on day 4 for measurements. For long-term mAb expression analysis, master cultures were maintained and passaged accordingly for up to 5 weeks with measurements performed on the same day weekly. The amount of secreted JUG444 or anti-hPD1 was measured in duplicate with the Octet RED96 system (ForteBio, Fremont, CA, USA) using Protein A biosensors (18-5010, ForteBio). Purified JUG444 or anti-hPD1 was used to generate a standard curve, from which the mAb titers were derived. Doubling time of each mAb-expressing cell population was estimated by the proliferation of  $5 \times 10^4$  cells after 2 days of culture and subsequently compared with that of singly-occupied dLP-CHO cells without any mAb gene circuit.

### CHO-tumor-T-cell co-culture model and anti-tumor activity analysis

dLP-CHO cells with an integrated gene circuit expressing anti-hPD1 (5C4) were seeded into a 12-well plate in triplicate at a density of  $1 \times 10^5$  cells/well with 1 mL CHO cell culture medium. Cells were incubated at 37 °C for 48 hours. Pre-activated human T cells and human ovarian cancer cells (OVCAR8) expressing a surface T-cell engager<sup>10</sup> were mixed at an effector-to-target (E:T) ratio of 20:1 ( $1 \times 10^6:0.5 \times 10^5$  cells) in 1 mL T-cell culture medium (RPMI-1640 with 10% FBS, 1% P/S, 10 mM HEPES, 0.1 mM non-essential amino acids, 1 mM sodium pyruvate, and 50  $\mu\text{M}$  2-Mercaptoethanol) and then added to each well containing attached dLP-CHO cells. OVCAR8 was a gift from Sangeeta N. Bhatia (Massachusetts Institute of Technology, Cambridge, MA). Co-cultures of three cell types (CHO, OVCAR8, and T cells) were incubated at 37 °C. Cell-free supernatants were collected and processed at 24 hours and stored at -20 °C. As an indicator of the T-cell response to tumor cells in the presence of actively secreted anti-hPD1 from dLP-CHO cells, interferon (IFN)- $\gamma$  produced by T cells was quantified. IFN- $\gamma$  concentrations in cell-free supernatants were determined by Human IFN- $\gamma$  DuoSet enzyme-linked immunosorbent assay (DY285; R&D systems, Minneapolis, MN, USA). EYFP-/EBFP+ dLP-CHO cells with no anti-hPD1 payload circuit were used as the control cells.

### Statistical analysis

All quantitative data are presented as mean  $\pm$  standard deviation (SD). Statistical differences between groups were analyzed by two-tailed Student's  $t$  test (for two groups), one-way ANOVA (for multiple groups), or two-way ANOVA (for long-term stability measurements) with 95% confidence interval. Statistical significance was set at  $p \leq 0.05$ . A Dunnett or a Tukey test was performed to correct for multiple comparisons in ANOVA posthoc analysis. All statistical analyses, including correlation analysis and linear regression, were performed

and graphed with GraphPad Prism 7.0 (GraphPad Software, La Jolla, CA, USA) statistics software.

### Reporting summary

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

### Data availability

The authors declare that all relevant data supporting the findings of this study are available within the paper and its supplementary Information. Source data are provided with this paper.

### References

1. Consortium, G. T. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204 (2017).
2. Shukla, A. A. & Thömmes, J. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends Biotechnol.* **28**, 253–261 (2010).
3. Matharu, N. & Ahituv, N. Modulating gene regulation to treat genetic disorders. *Nat. Rev. Drug Discov.* **19**, 757–775 (2020).
4. Chang, M. M. et al. Small-molecule control of antibody N-glycosylation in engineered mammalian cells. *Nat. Chem. Biol.* **15**, 730–736 (2019).
5. Courbet, A., Renard, E. & Molina, F. Bringing next-generation diagnostics to the clinic through synthetic biology. *EMBO Mol. Med.* **8**, 987–991 (2016).
6. Higashikuni, Y., Chen, W. C. W. & Lu, T. K. Advancing therapeutic applications of synthetic gene circuits. *Curr. Opin. Biotechnol.* **47**, 133–141 (2017).
7. Lillacci, G., Benenson, Y. & Khammash, M. Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic Acids Res.* **46**, 9855–9863 (2018).
8. Matsuura, S. et al. Synthetic RNA-based logic computation in mammalian cells. *Nat. Commun.* **9**, 4847–4847 (2018).
9. Mansouri, M., Strittmatter, T. & Fussenegger, M. Light-controlled mammalian cells and their therapeutic applications in synthetic biology. *Adv. Sci.* **6**, 1800952–1800952 (2018).
10. Nissim, L. et al. Synthetic RNA-based immunomodulatory gene circuits for cancer immunotherapy. *Cell* **171**, 1138–1150.e15 (2017).
11. Chen, C. et al. Promoter-operating targeted expression of gene therapy in cancer: current stage and prospect. *Mol. Ther. Nucleic Acids* **11**, 508–514 (2018).
12. Xie, X. et al. Targeted expression of BikDD eliminates breast cancer with virtually no toxicity in noninvasive imaging models. *Mol. Cancer Therapeutics* **11**, 1915–1924 (2012).
13. Qin, J. Y. et al. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* **5**, e10611 (2010).
14. Gray, S. J. et al. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Hum. Gene Ther.* **22**, 1143–1153 (2011).
15. Rathnam, C., Chueng, S.-T. D., Yang, L. & Lee, K.-B. Advanced gene manipulation methods for stem cell therapeutics. *Theranostics* **7**, 2775–2793 (2017).
16. Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* **79**, 213–231 (2010).
17. Moore, R., Chandrabhas, A. & Bleris, L. Transcription activator-like effectors: a toolkit for synthetic biology. *ACS Synth. Biol.* **3**, 708–716 (2014).
18. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
19. Kim, T. & Lu, T. K. CRISPR/Cas-based devices for mammalian synthetic biology. *Curr. Opin. Chem. Biol.* **52**, 23–30 (2019).
20. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563–564 (1988).
21. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci.* **95**, 14628–14633 (1998).
22. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Meth.* **12**, 326–328 (2015).
23. Pandelakis, M., Delgado, E. & Ebrahimkhani, M. R. CRISPR-based synthetic transcription factors in vivo: the future of therapeutic cellular programming. *Cell Syst.* **10**, 1–14 (2020).
24. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
25. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
26. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
27. Chavez, A. et al. Comparison of Cas9 activators in multiple species. *Nat. Meth.* **13**, 563–567 (2016).
28. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.* **2**, 604–613 (2013).
29. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
30. Gaidukov, L. et al. A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res.* **46**, 4072–4086 (2018).
31. Jusiak, B. et al. Comparison of integrases identifies Bxb1-GA mutant as the most efficient site-specific integrase system in mammalian cells. *ACS Synth. Biol.* **8**, 16–24 (2019).
32. Fitzgerald, M. M., Livingston, C., Gibbs, T. L., Deans, Rosa26 docking sites for investigating genetic circuit silencing in stem cells, *Synth. Biol.* **5**, ysaa014 (2020).
33. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* **41**, e156 (2013).
34. Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2006).
35. Rose, A. B. Introns as gene regulators: a brick on the accelerator. *Front. Genet.* **9**, 672 (2019).
36. Kim, J. H. et al. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, Zebrafish and mice. *PLoS One* **6**, e18556 (2011).
37. Inniss, M. C. et al. A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO cells. *Biotechnol. Bioeng.* **114**, 1837–1846 (2017).
38. Jost, M. et al. Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* **38**, 355–364 (2020).
39. Leavitt, J. M. & Alper, H. S. Advances and current limitations in transcript-level control of gene expression. *Curr. Opin. Biotechnol.* **34**, 98–104 (2015).
40. Brown, A. J., Gibson, S. J., Hatton, D. & James, D. C. In silico design of context-responsive mammalian promoters with user-defined functionality. *Nucleic Acids Res.* **45**, 10906–10919 (2017).
41. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
42. Qian, Y., Huang, H.-H., Jiménez, J. I. & Del Vecchio, D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* **6**, 1263–1272 (2017).

43. Wilkinson, A. C., Nakauchi, H. & Göttgens, B. Mammalian transcription factor networks: recent advances in interrogating biological complexity. *Cell Syst.* **5**, 319–331 (2017).
44. Xu, X. et al. Gene activation by a CRISPR-assisted trans enhancer. *eLife* **8**, e45973 (2019).
45. Mariati, Ho, S. C. L., Yap, M. G. S. & Yang, Y. Evaluating post-transcriptional regulatory elements for enhancing transient gene expression levels in CHO K1 and HEK293 cells. *Protein Expr. Purif.* **69**, 9–15 (2010).
46. Choi, T., Huang, M., Gorman, C. & Jaenisch, R. A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol.* **11**, 3070–3074 (1991).
47. Shaul, O. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* **91**, 145–155 (2017).
48. Cao, J. et al. High-throughput 5' UTR engineering for enhanced protein production in non-viral gene therapies. *Nat. Commun.* **12**, 4138 (2021).
49. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**, 1136 (2018).
50. Yang, Z. et al. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* **33**, 842 (2015).
51. Brown, A. J. & James, D. C. Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnol. Adv.* **34**, 492–503 (2016).
52. Romanova, N. & Noll, T. Engineered and natural promoters and chromatin-modifying elements for recombinant protein expression in CHO cells. *Biotechnol. J.* **13**, 1700232 (2017).
53. Lee, J. S. et al. Revealing key determinants of clonal variation in transgene expression in recombinant CHO cells using targeted genome editing. *ACS Synth. Biol.* **7**, 2867–2878 (2018).
54. Duportet, X. et al. A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42**, 13440–13451 (2014).
55. Barde, I. et al. Lineage- and stage-restricted lentiviral vectors for the gene therapy of chronic granulomatous disease. *Gene Ther.* **18**, 1087 (2011).
56. Guye, P. et al. Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6. *Nat. Commun.* **7**, 10243 (2016).
57. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

## Acknowledgements

The authors gratefully thank Nevin M. Summers and Dr. Louane Hann for their administrative support, Kristjan E. Kaseniit for his contribution in gRNA design, Dr. Kevin J. Lebo for his expert input in the system design, Dr. Samuel D. Perli for genetic materials, Dr. Lei Wang for providing the hiPSC cell line, and Dr. Wen Allen Tseng for genetic materials and his contribution to construct design. We greatly thank Selamawit Mamo, Kalpana Jagtap, Na Li, Cong Liu, and Jonathan L. Lyles for their technical support with CHO cell cultures and assays, Mailing Ding and Christina Harrison for their technical support with molecular cloning and plasmid preparation, and Karen Pepper for editing the manuscript. We thank Dr. G. Felsenfeld (National Institute of Diabetes and Digestive and Kidney Diseases, USA) for the HS4 insulator sequences. This work was supported by the Pfizer-MIT RCA Synthetic Biology Program (CHO2.0 and Precision Post-Translational Modification to T.K.L. and R.W.). This project was also supported in part by the National Science Foundation (CCF-1521925 to T.K.L.), the National Institutes of Health (NIH) (5-U01-CA2550554-02, 50000655-5500001351, and 5R01HL135886 to T.K.L.), and the University of South Dakota Sanford School of Medicine (Startup Fund to W.C.). W.C. was supported in part by the NIH Ruth L. Kirschstein NRSA postdoctoral fellowship (5T32HL007208) and the Department of Defense (W81XWH2110089 to W.C.).

## Author contributions

W.C. conceptualized the research project and the system design, designed and prepared genetic and cellular materials, designed and performed experiments, analyzed and interpreted data, and wrote the manuscript; L.G. designed and prepared genetic and cellular materials, designed and performed experiments, analyzed and interpreted data, and edited the manuscript; Y.L. designed and prepared genetic and cellular materials, designed and performed experiments, analyzed data, and wrote the manuscript; M-R.W. prepared cellular materials for co-culture, designed and performed co-culture experiments and associated assays, and analyzed and interpreted data; J.C. prepared genetic and cellular materials and performed experiments; M.J.G. prepared cellular materials and performed experiments; G.C. contributed to the system design and designed and prepared genetic materials; R.P.U. and Y-C.C. designed and prepared genetic materials; L.W., M.K., and L.Z. contributed to the system and genetic material design and edited the manuscript; R.W. contributed to the system and genetic material design, provided genetic and cellular materials, lab space and equipment, interpreted data, and supervised the study; T.K.L. conceptualized the study, provided genetic and cellular materials, lab space, equipment and funding, interpreted data, edited the manuscript, and supervised the study.

## Competing interests

T.K.L. is a co-founder of Senti Biosciences, Synlogic, Engine Biosciences, Tango Therapeutics, Corvium, BiomX, Eligo Biosciences, Open-Protein.AI, Bota.Bio. T.K.L. also holds financial interests in nest.bio, Amplphi, IndieBio, Cognito Health, Quark Biosciences, Personal Genomics, Thryve, Lexent Bio, MitoLab, Vulcan, Serotiny, and Provectus Algae. All other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33287-9>.

**Correspondence** and requests for materials should be addressed to William C. W. Chen or Timothy K. Lu.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**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

---

<sup>1</sup>Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>2</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>3</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>4</sup>Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA. <sup>5</sup>Division of Basic Biomedical Sciences, Sanford School of Medicine, University of South Dakota, Vermillion, SD 57069, USA. <sup>6</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>7</sup>Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02139, USA. <sup>8</sup>Department of Biochemistry, Wellesley College, Wellesley, MA 02481, USA. <sup>9</sup>Pfizer Inc., Cambridge, MA 02139, USA. <sup>10</sup>Pfizer Inc., Andover, MA 01810, USA. <sup>11</sup>Present address: Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA. <sup>12</sup>Present address: Laboratory of Combinatorial Genetics and Synthetic Biology, School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China. <sup>13</sup>Present address: Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan. <sup>14</sup>These authors contributed equally: William C. W. Chen, Leonid Gaidukov, Yong Lai. ✉ e-mail: [William.Chen@usd.edu](mailto:William.Chen@usd.edu); [tim@lugroup.org](mailto:tim@lugroup.org)