Published online 1 April 2019

Nucleic Acids Research, 2019, Vol. 47, No. 12 ***e67***
doi: 10.1093/nar/gkz210

Multiplexed and tunable transcriptional activation by promoter insertion using nuclease-assisted vector integration

Alexander Brown[1,†], Jackson Winter[1,†], Michael Gapinske[1,†], Nathan Tague[1], Wendy S. Woods[1] and Pablo Perez-Pinera[1,2,3,4,*]

[1]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [3]Carle Illinois College of Medicine, Champaign, IL 61820, USA and [4]Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received January 03, 2019; Revised February 20, 2019; Editorial Decision March 15, 2019; Accepted March 18, 2019

ABSTRACT

The ability to selectively regulate expression of any target gene within a genome provides a means to address a variety of diseases and disorders. While artificial transcription factors are emerging as powerful tools for gene activation within a natural chromosomal context, current generations often exhibit relatively weak, variable, or unpredictable activity across targets. To address these limitations, we developed a novel system for gene activation, which bypasses native promoters to achieve unprecedented levels of transcriptional upregulation by integrating synthetic promoters at target sites. This gene activation system is multiplexable and easily tuned for precise control of expression levels. Importantly, since promoter vector integration requires just one variable sgRNA to target each gene of interest, this procedure can be implemented with minimal cloning. Collectively, these results demonstrate a novel system for gene activation with wide adaptability for studies of transcriptional regulation and cell line engineering.

INTRODUCTION

The activation of endogenous genes with artificial transcription factors (ATFs) is an enticing technology, not only for developing gene therapies or disease models (1,2), but also for interrogating gene function through genome-wide screenings (3,4). ATFs consist of a programmable DNA binding domain that can be customized to target a transcriptional activation domain to the appropriate locus for upregulation of gene expression. While zinc finger proteins (5) and Transcriptional Activator-Like Effectors (TALEs) (6–9) have been used for gene activation, the RNA guided nuclease (RGN) platform (1,10–16) is arguably the most popular since the DNA binding specificity can be engineered rapidly and at low cost (17–22). CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) activation or CRISPRa, requires a single-guide RNA (sgRNA) and catalytically dead Cas9 (dCas9) coupled with a transcriptional activator. First generation transcriptional activators, which typically used VP64 or VP16 activation domains, required multiple ATFs acting in synergy near the transcriptional start site (TSS) of the gene of interest for optimal gene activation (6,7,13,15). This important limitation is lessened when using second-generation transcriptional activators, including VP160 (10), SAM (23), VPR (24), SunTag (25), VP64-dCas9-BFP-VP64 (26), Scaffold (27) and P300 (28), which are capable of activating expression of some target genes when used individually.

Unfortunately, it is becoming evident that even second generation CRISPRa technologies are often limited by their need for multiple sgRNAs to achieve adequate activation of many genes (29) and the lack of established parameters to best position ATFs within endogenous promoters for effective upregulation of gene expression. Importantly, CRISPRa systems often fail at activating genes whose expression is tightly regulated. These constraints limit widespread adoption of CRISPRa for applications in synthetic biology, tissue engineering or gene therapy.

In these studies, we sought to develop an alternative architecture that bypasses the limitations of current platforms for activating native genes. We reasoned that, since genomic context at the promoter greatly impacts output expression when using ATFs, it might be possible to circumvent this problem through insertion of a synthetic promoter near the TSS of target genes. This system would not only

*To whom correspondence should be addressed. Tel: +1 217 265 5147; Email: pablo@illinois.edu
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.
Present address: Alexander Brown, Horae Gene Therapy Center, University of Massachusetts School of Medicine, Worcester, MA 01605, USA.

override negative regulatory elements, but would also be highly customizable, given the existing assortment of well-characterized synthetic promoters capable of both constitutive and chemically inducible gene expression.

In this manuscript, we utilized a universal vector integration platform (30,31) to engineer a synthetic system for activating endogenous genes. To this end, we describe how this platform enables rapid, robust and inducible activation of both individual and multiplexed gene transcripts.

MATERIALS AND METHODS

Cell culture and transfection

293T, HCT116 and Neuro-2A cells were maintained in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C with 5% CO_2. SF7996 primary glioblastoma cells were a kind gift from Joseph Costello (32) and were cultured in DMEM/Ham's F-12 1:1 media, 10% FBS, 1% Penicillin/Streptomycin. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Transfection efficiencies were routinely higher than 80% for 293T and Neuro-2A cells, ∼50% for HCT116 cells and ∼25% for SF7996 cells as determined by fluorescent microscopy following delivery of a control GFP expression plasmid. Selection of transfected cells was performed by culturing in complete medium containing puromycin for 72 h. Concentrations of puromycin were 2 μg/ml for 293T, 0.5 μg/ml for HCT116, 1 μg/ml for SF7996 and 3 μg/ml for Neuro-2A. Induction of gene expression, unless otherwise noted, was carried out with 200 ng/ml doxycycline in DMEM prepared with 10% tetracycline-free FBS for 4 days. Growth rate comparison was performed by seeding 20 000 cells per well, counting the number of cells using a hemocytometer after 6 days and seeding 20 000 cells back to analyze at the next time point. Population Doublings were calculated using the equation $PDs = 3.32(\log(Y) - \log(X))$, where Y is the final cell count and X is the initial cell count (33).

Plasmids and oligonucleotides

The plasmids encoding SpCas9 (Plasmid #41815), sgRNA (#47108) and SpdCas9-VPR (#63798) were obtained from Addgene. The backbone for the targeting vectors was synthesized by IDT as gene blocks and cloned into a pCDNA3.1 plasmid. The oligonucleotides used to create the guide sequences were obtained from IDT, hybridized, phosphorylated and cloned in the sgRNA vector using BbsI (15,34). The target sequences are provided in Supplementary Table S1.

PCR

Seventy-two hours after transfection, genomic DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen). PCRs were performed using KAPA2G Robust PCR kits (KAPA Biosystems). A typical 25 μl reaction used 20–100 ng of genomic DNA, Buffer A (5 μl), Enhancer (5 μl), dNTPs (0.5 μl), 10μM forward primer (1.25 μl), 10 μM reverse primer (1.25 μl), KAPA2G Robust DNA Polymerase (0.5 U) and water (up to 25 μl). The DNA sequence of the primers for each target are provided in Supplementary Table S2. The PCR products were visualized in 2% agarose gels and images were captured using a ChemiDoc-It^2 (UVP).

qPCR

Cells were harvested and flash-frozen in liquid nitrogen prior to RNA-extraction using the RNeasy Plus RNA isolation kit (Qiagen) according to manufacturer's instructions. cDNA synthesis was carried out using the qScript cDNA Synthesis Kit (Quanta Biosciences) from 1 μg of RNA and reactions were performed as directed by the supplier. For RT-qPCR, SsoFast EvaGreen Supermix (Bio-Rad) was added to cDNA and primers targeting the gene of interest and *GAPDH* (Supplementary Table S3). Following 30 s at 95°C, PCR amplification (5 s at 95°C, 20 s at 55°C, 40 total cycles) preceded melt-curve analysis of the product by the CFX Connect Real-Time System (Bio-Rad). Ct values were used to calculate changes in expression level, relative to *GAPDH* and control samples by the $2^{-\Delta\Delta Ct}$ method. qPCR standard curves were prepared for each target (Supplementary Figure S1). RNA integrity numbers (RINs) for representative RNA samples prepared using the methods described were calculated by the Functional Genomics Unit of the Roy J. Carver Biotechnology Center using an Agilent Bioanalyzer RNA Nano chip according to manufacturer's instructions (Supplementary Table S4).

Western blot

Cell pellets were resuspended in 1X NuPAGE LDS Sample Buffer (Life Technologies) with 2.5% β-mercaptoethanol, heated for 5 min at 95°C, and sonicated. Lysates were loaded into a 4–12% NuPAGE Bis-Tris Protein Gel alongside Precision Plus Protein Dual Color Standard ladder (BioRad), electrophoresed, and transferred onto a 0.45 μm nitrocellulose membrane (BioRad). Membranes were blocked for 2 hours with 5% nonfat dried milk in TBS containing 0.1% Tween-20 (TBST). Membranes were then rinsed three times in TBST and incubated with anti-human GAPDH rabbit antibody (1:1000, Cell Signaling Technology, 14C10) or anti-human NEUROD1 rabbit antibody (1:1000, Cell Signaling Technology, D35G2) in 5% BSA in TBST overnight with gentle rocking at 4°C. Following primary antibody incubation, membranes were rinsed three times in TBST and incubated for 40 min with HRP-conjugated anti-rabbit IgG (1:3000, Cell Signaling Technology, #7074). Membranes were then rinsed with TBST three times and incubated for five minutes in Clarity Western ECL Substrate (BioRad). Membranes were then imaged using 60 min exposure time for NEUROD1 and 10 min exposure time for GAPDH with a ChemiDoc-It^2 (UVP).

Western blot densitometry analysis

NEUROD1 protein expression levels were compared using densitometry analysis of the western blot membrane images using ImageJ software. After subtracting background noise, NEUROD1 band intensities were normalized to GAPDH band intensities, where band intensity is the sum of each pixel grayscale value within the selected area of the band.

Statistics

Statistical analysis was performed by two-way ANOVA with alpha equal to 0.05 or with *t* tests in Prism 7.

RESULTS

We chose nuclease-assisted vector integration (NAVI) (31) for insertion of promoters at target sites. NAVI can be rapidly adapted to integrate heterologous DNA at virtually any locus via two simultaneous DSBs: first in the genome, guided by a primary sgRNA, and second within the targeting vector (TV), guided by a universal secondary sgRNA (31). The TV is then integrated into the genomic locus through Non-Homologous End Joining (NHEJ). This platform is universal since vector integration at any target site can be simply accomplished by customizing the primary sgRNA. To develop a universal system of NAVI-based gene activation (NAVIa), we designed two vectors for constitutive expression and one vector for inducible expression.

The two constitutive vectors contain either one CMV promoter followed by a target site for a universal secondary sgRNA (constitutive single promoter targeting vector, cspTV) or two opposing constitutive promoters separated by the secondary sgRNA target site (constitutive dual promoter targeting vector, cdpTV), each containing a cassette for expression of the puromycin *N*-acetyl-transferase gene (Figure 1A). The targeting vector for inducible expression (inducible dual promoter targeting vector, idpTV) includes two identical promoters in opposite orientations, each consisting of seven TetO repeats and a minimal CMV promoter (mCMV). The idpTV also carries a puromycin N-acetyl-transferase gene linked with a reverse tetracycline transactivator (rtTA) via a T2A peptide. As in the cdpTV, the opposing promoters of the idpTV flank a universal secondary sgRNA target sequence. A DSB introduced in either idpTV or cdpTV by Cas9 generates a linear fragment of DNA with diametric promoters oriented towards the free ends of the vector (Figure 1A). The architecture of the dual promoter TV ensures that there is always a promoter correctly positioned regardless of integration orientation, thereby addressing NAVI's lack of directionality.

In order to evaluate this gene activation architecture in the context of the human genome, we first selected three target genes whose reported levels of activation utilizing CRISPRa are either high (*ASCL1,* $\sim 10^3$-fold), medium (*NEUROD1,* $\sim 10^2$-fold), or low (*POU5F1,* ∼10-fold) (23,24). The primary sgRNAs targeting the genome were co-transfected into 293T cells with three plasmids containing (a) an expression cassette for active Cas9, (b) our customized cspTV, cdpTV or idpTV, and (c) a universal secondary sgRNA. Following transfection, cells with integration of the TV were selected using puromycin and, in cells transfected with the idpTV, gene expression was induced with doxycycline. Isolated clones were screened by PCR to verify integration of the idpTV at the locus of interest (Supplementary Figure S2, Supplementary Table S5). In parallel, one sgRNA or a mixture of 4 sgRNAs (previously validated for use with CRISPRa) was co-transfected into 293Ts with dCas9-VPR for comparison of our system with CRISPRa (23,24,29). Gene expression using an individual sgRNA directing dCas9-VPR to target promoters was increased ∼10-fold for all targets tested but the results were not statistically significant. Utilization of four sgRNAs simultaneously activated gene expression more effectively than 1 sgRNA (*ASCL1:* ∼1800-fold, *NEUROD1:* ∼2900-fold, *POU5F1:* ∼90-fold). With NAVIa, the degree of gene activation using the cspTV (*ASCL1:* ∼730-fold, *NEUROD1:* ∼600-fold, *POU5F1:* ∼200-fold) or cdpTV (*ASCL1:* ∼8500-fold, *NEUROD1:* ∼3000-fold, *POU5F1:* ∼1000-fold) was superior to CRISPRa using 1 sgRNA but lower or not statistically different from activation obtained using CRISPRa with 4 sgRNAs for two of the three targets. However, the idpTV (*ASCL1:* ∼7200-fold, *NEUROD1:* ∼76000-fold, *POU5F1:* ∼5370-fold) surpassed activation obtained using dCas9-VPR using four sgRNAs (Figure 1B). Interestingly, the improvement of NAVIa over dCas9-VPR was higher for targets branded as difficult to regulate with CRISPRa (*POU5F1:* ∼60-fold improvement, *NEUROD1:* ∼26-fold improvement) than for a target considered easy to activate (*ASCL1:* ∼4-fold improvement).

When using CRISPRa it is difficult to predict optimal sgRNA target sites for efficient gene activation. While it is generally accepted that proximity to the TSS of the target gene is important, other parameters such as presence of enhancers or local chromatin structure are also critical and, perhaps, more difficult to predict (23–25,29). We investigated a potential correlation between gene activation using NAVIa and distance between integration site and TSS by measuring gene expression induced with sgRNAs that target DNA sequences between positions –1010 and +1995, relative to the TSS of three different genes (Figure 1C). Plotting these data for all three genes showed that NAVIa can activate gene expression efficiently from any integration site on this range, with the most activity being derived from sgRNAs between –500 and +200 bp relative to the TSS.

Since maximal gene activation may not be desirable in all experimental settings, CRISPRa has been adapted for tunable gene expression through combinatorial delivery of multiple sgRNAs (10–16). However, such efforts to modulate gene expression have proven unpredictable. Alternatively, NAVIa enables facile integration of any TV, which can facilitate gene activation by a wide variety of regulatory mechanisms provided by existing artificial promoters. As the idpTV used in these experiments introduces a doxycycline-inducible promoter, we anticipated a precise temporal control of gene expression that could be tuned by the concentration of doxycycline in the growth medium. Induction of gene expression for 96 h with concentrations of doxycycline ranging from 2 ng/ml to 2 μg/ml led to a dose-dependent increase in gene expression ranging between ∼337-fold and ∼26 015-fold (Figure 1D). Considering this result, we chose to use 200 ng/ml doxycycline for a time course that demonstrated that induction of *NEUROD1* is detectable 12 h after treatment (∼4000-fold) and continues to increase at 24 h (∼5000-fold), 48 h (∼10 000-fold) and 96 h (∼15 000-fold) (Figure 1E). Western blots of a clonal population of 293Ts with heterozygous idpTV integration at the *NEUROD1* locus confirmed an increase in *NEUROD1* protein expression levels in the induced samples compared to the WT controls and uninduced samples (Figure 1F). Furthermore, upon removal of doxycycline, the RNA expression of *NEUROD1*

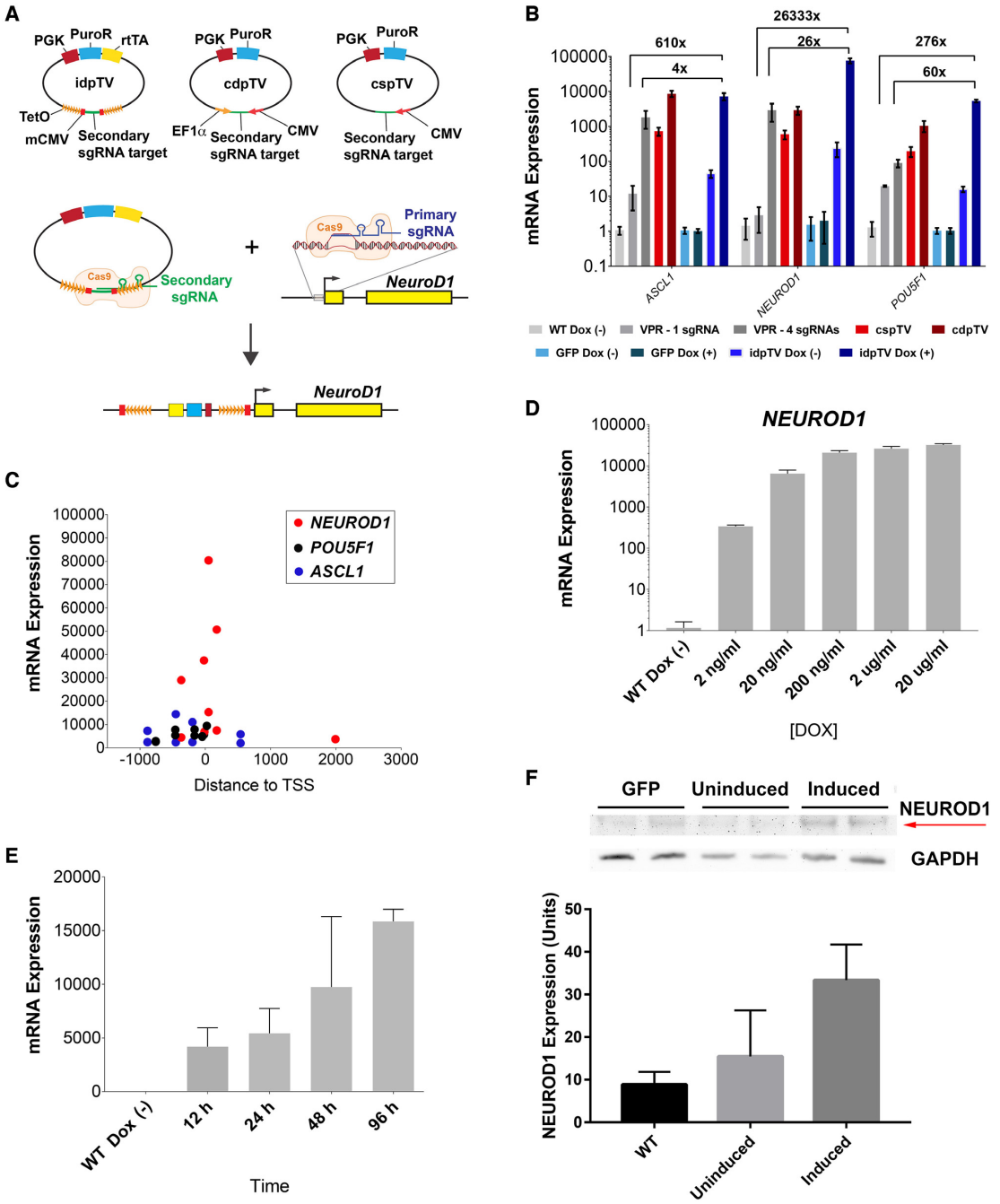

Figure 1. NAVIa Activation of Native Gene Expression is Tunable and Surpasses CRISPRa. (**A**) The architecture of the NAVIa system includes a plasmid containing a human codon-optimized expression cassette for active Cas9, which is co-transfected with two separate sgRNA plasmids and a targeting vector (idpTV, cdpTV or cspTV). The primary sgRNA, shown in dark blue, is designed to bind and target Cas9 to the 5′ region of the gene of interest, while the secondary sgRNA target site (green) is at the 3′ end of the cspTV promoter, or between the diametric promoters of the cdpTV and idpTV. After Cas9 cuts the TV, the resulting linearized vector is integrated at the target site in genomic DNA, presumably via NHEJ repair of the double-stranded breaks. (**B**) The ability of NAVIa to upregulate the expression of target transcript within pooled, selected 293T cells was evaluated using qPCR across a panel of three genes: *ASCL1*, *NEUROD1*, and *POUF51*. Each sgRNA employed within NAVIa was also used for gene activation with CRISPRa (dCas9-VPR) either alone or in conjunction with three additional sgRNAs, previously reported to activate expression of the target mRNA. Data shown as the mean ± S.E.M. ($n = 3$ independent experiments). *P* values were determined by *t* test: idpTV versus four sgRNAs: $P \leq 0.05$ for all targets, cdpTV versus 4 sgRNA: $P \leq 0.05$ for *ASCL1*, idpTV, cspTV or cdpTV versus 1 sgRNA: $P \leq 0.05$ for all targets. (**C**) Representation of levels of activation relative to distance between sgRNA targeting and the canonical TSS. (**D**) Expression of *NEUROD1* was induced using NAVIa for a period of 4 days at concentrations of doxycycline ranging from 2 ng/ml to 2 μg/ml and measured using qPCR. (**E**) Expression of *NEUROD1* was measured by qPCR upon induction with 200 ng/ml doxycycline for 12, 24, 48 and 96 h in 293T cells in which *NEUROD1* was edited using NAVIa. Data in B, D and E are shown as the mean ± S.E.M. ($n = 3$ independent experiments). (**F**) Western blot analysis of NEUROD1 protein expression was performed using cell lysates prepared from wild type 293T cells and a 293T clonal population with idpTV integration at the NEUROD1 locus without induction or after 4 days of culture in Tet-Free DMEM containing 200 ng/ml doxycycline. Densitometry analysis demonstrated an increase in NEUROD1 protein expression in the induced samples compared to the wild type controls and the uninduced samples. Error bars represent the S.D. ($n = 2$).

returns to near basal levels within 96 hours (Supplementary Figure S3).

Tetracycline-inducible systems have been designed for high responsiveness to doxycycline, yet background expression in the absence of inducer, while low, continues to be a problem that hinders applications requiring precise control over gene activation (35). While inducibility is a significant advantage of NAVIa over CRISPRa, tetracycline-inducible promoters are typically used to modulate expression cassettes within a vector and not in a genomic context where the surrounding transcriptional regulatory elements may contribute to undesired expression at steady state. Analysis of *NEUROD1* activation within samples not induced with doxycycline revealed significant background expression (∼432-fold over basal expression, Figure 2A). While we failed to identify a correlation between background and distance from the integration to ATG codons (Supplementary Figure S4) or between background expression and basal expression (Supplementary Figure S5), we reasoned that expression of rtTA from unintegrated plasmids still transiently present from the transfection might be partly responsible for high background levels of expression. Indeed, background expression in clones with heterozygous or homozygous integrations was significantly lower than in pooled populations, while gene induction in heterozygous clones was similar to that observed in pooled populations but significantly lower than activation in homozygous clones. The ratio of gene expression between samples with and without doxycycline treatment was improved from ∼22-fold induction in pooled cells to ∼126-fold and ∼1486-fold in heterozygous and homozygous clones respectively (Figure 2A). The levels of activation observed in heterozygous clones, homozygous clones, and pooled populations were similar in the presence of doxycycline, which indicates that the high levels of expression seen in the pooled population are unlikely due to a minority of cells expressing large amounts of transcript.

One important feature of CRISPRa architectures is multiplexability. Different genes can be activated simultaneously by delivering sgRNAs targeting each promoter (10,23,24,36). Two benefits of NAVI over other integration platforms, such as those utilizing HR, are the universal adaptability of the system to target different genomic loci by simply providing additional primary sgRNAs and the facile clone screening and isolation upon selection. Since activation of different genes using NAVIa can be accomplished using a set of vectors in which the only variable element is the primary sgRNA, this flexible architecture is also compatible with multiplexing. To demonstrate these capabilities, we first identified sgRNAs for targeting additional genes with NAVIa including *IL1B*, *IL1R2*, *LIN28A* and *ZFP42* (Supplementary Figure S6). To facilitate multiplexing, we utilized a custom Golden Gate cloning plasmid to prepare two multi-sgRNA (mgRNA) vectors capable of delivering a total of seven individual sgRNAs targeting genes and one sgRNA for linearizing the idpTV, each under independent promoter control (36). Co-transfection of these plasmids alongside the idpTV and Cas9 vectors into 293T cells was followed by induction of gene expression with doxycycline for two days. Analysis of mRNA expression across all targeted genes demonstrates that multiplexed gene activation with NAVIa surpasses CRISPRa for all targets tested (ranging from ∼15-fold to ∼400-fold) (Figure 2B). Together, these results emphasize the multiplexing capabilities of NAVIa, as well as a clear advantage over CRISPRa when only one sgRNA is employed.

To further validate the trends we observed in 293T cells, we targeted *NEUROD1* using our cdpTV in other cell lines. NAVIa effectively activated expression of *NEUROD1* in the human colorectal carcinoma cell line HCT116, the primary human fibroblast cell line MRC-5, and the mouse neuroblastoma cell line Neuro-2A (Supplementary Figure S7).

Finally, we chose *TERT* as a target to demonstrate the applicability of NAVIa for activating genes that are difficult to regulate with CRISPRa (15,23) in a primary cell line that is difficult to transfect as well as to demonstrate a physiological effect. Following transfection and selection of SF7996 cells (primary glioblastoma cells, which depend on *TERT* for survival and proliferation (32,37)), we derived a clonal population in which expression of *TERT* is controlled by the idpTV and can be induced in a dose-dependent manner with doxycycline. It is noteworthy that following idpTV integration, *TERT* expression could no longer be detected without induction (Figure 2C), but the addition of doxycycline upregulated gene expression and enabled ∼40-fold activation over control untreated cells (Figure 2C). Since SF7996 cells depend on *TERT* expression for survival over multiple cell divisions, they were maintained in regular growth media supplemented with doxycycline. When the growth media was switched to tetracycline-free medium, the proliferation rate decreased relative to that of the doxycycline-induced cells (Figure 2D), demonstrating a method to immortalize cell lines by regulating expression of native *TERT*.

DISCUSSION

In this manuscript we describe a novel platform to activate native gene expression based on integration of heterologous promoters, which provides some advantages over CRISPRa. For example, NAVIa enables robust activation across target genes following a single transfection and with minimal cloning and facile isolation of isogenic cell lines expressing the selected gene. Furthermore, NAVIa can be adapted for gene regulation using any constitutive or inducible promoters of interest and achieves consistent activation from a wide range of positions near the TSS, which minimizes the screening needed to identify optimal sgRNAs.

Promoter integration is accomplished by NAVI (31), which utilizes NHEJ, which provides significant advantages over DNA integration platforms that rely on Homologous Recombination (HR). For example, NHEJ is more effective than HR in non-dividing cells and has been exploited to integrate therapeutic transgenes in post-mitotic cells (38). Although NAVI is subject to some shortcomings associated with its specific gene editing mechanism, such as the error-prone nature of NHEJ and chromosomal translocations (39), we have observed only minor indels at target sites both here (Supplementary Figure S8) and in previous studies (31). In a recent study, Niu *et al.* (40) identified chromosomal abnormalities in cell lines that had been modified

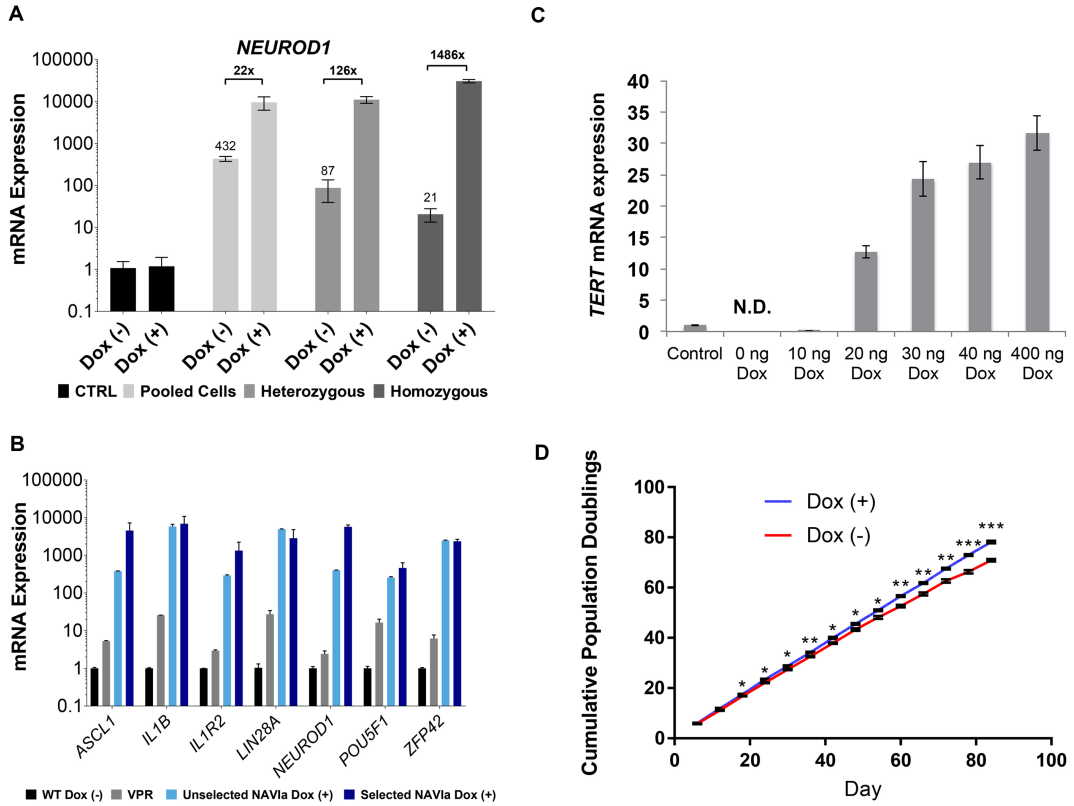

Figure 2. Multiplexed Gene Activation Using NAVIa. (**A**) Comparison of background and induced expression of *NEUROD1* targeted using NAVIa between pooled HCT116 cells (diploid) and clones that were positive for idpTV integration at either one or both alleles ($n = 3$ independent experiments). Untreated pooled cells versus heterozygous, $P \leq 0.003$. Untreated heterozygous versus homozygous, $P \leq 0.07$. Untreated pooled cells versus homozygous, $P \leq 0.0005$. Doxycycline treated heterozygous versus homozygous, $P \leq 0.001$. Doxycycline treated pooled cells versus homozygous, $P \leq 0.001$. (**B**) 293T cells were transfected with CRISPRa or NAVIa targeting simultaneously the genes *ASCL1*, *NEUROD1*, *POUF51, IL1B, IL1R2, LIN28A* and *ZFP42*. Expression of the target genes without selection was measured at day 3 using qPCR ($n = 2$ independent experiments). Data is shown as mean ± S.E.M. *P* values were determined by *t* test (NAVIa versus VPR, $P \leq 0.001$ *ASCL1*, $P \leq 0.02$ *IL1B* (Ct value of control sample was not detected and assumed to be 40), $P \leq 0.004$ *IL1R2*, $P \leq 0.001$ *LIN28A*, $P \leq 0.001$ *NEUROD1*, $P \leq 0.007$ *POUF51*, $P \leq 0.001$ *ZFP42*). (**C**) The idpTV was integrated at the *TERT* locus in SF7996 primary glioblastoma cells and expression of *TERT* was increased in a dose-dependent manner by addition of doxycycline compared with untreated control cells ($n = 4$, $P < 0.005$). N.D.: not detected. (**D**) The proliferation rates between cells cultured in doxycycline-free medium and cells cultured in 400 ng/ml doxycycline was compared by tracking cumulative population doublings over 84 days ($n = 3$, * represents $P \leq 0.05$, ** represents $P \leq 0.01$, *** represents $P \leq 0.001$). Data in A, B, C and D are shown as the mean ± S.E.M.

via multiplexed DSBs, however they concluded that chromosomally normal clones could be obtained through genomic screening. While this is a risk inherent to any editing strategy that relies on DSBs, ATFs are not necessarily safer in this regard as permanent regulation of gene expression through ATFs can only be obtained through stable integration using viral vectors, gene editing, or transposases, which entail similar risks to NAVIa. Another potential problem resulting from integration via NHEJ is that it is also possible to have multiple copies of the transfer vector integrated at the site of the genomic DSB, however these events can be easily screened for by PCR. It should be noted that we were unable to detect any instances of multiple integrations in a pooled population of 293T cells containing idpTV integrations at the *NEUROD1* locus, as well as three clonal populations. Furthermore, the dual promoter architecture ensures that there is always a synthetic promoter activating gene expression should multiple linearized transfer vectors be integrated.

In the majority of the experiments shown in this manuscript we demonstrate very high, supraphysiological levels of gene activation, which may not be necessary and/or relevant in all experimental settings. However, it is important to note that these high levels of expression can be modulated by controlling the dose of inducer to achieve any desired outcome. We have observed that in the absence of an inducer the levels of expression of the target genes increased over background due to the inherent leakiness of the inducible promoter used in this particular study. However, since this gene activation system is universal, we anticipate that any other tightly-controlled promoter can be used to minimize this potential problem. Furthermore, we also demonstrated that the levels of background expression can be decreased by controlling the number of modified alleles (Figure 2A). Despite the drastic increase in rates of *NEUROD1* transcription, only ∼2-fold increase was observed at the protein level, indicating that rates of translation remain a bottleneck in overexpression studies. Additionally, while we achieved very high levels of expression in model cell lines,

such as 293Ts, activation in primary cells, in which native gene expression is more difficult to regulate, more closely resembled physiological conditions. Importantly, in experiments involving primary cells and/or genes that are tightly regulated, such as *TERT* (Figure 2C, D), CRISPRa is often ineffective (23,24).

Conceptually, gene activation by NAVIa resembles transgene expression from a heterologous promoter. However, NAVIa provides some advantages over these systems. For instance, expression of large transgenes is challenging using viral vectors such as AAV or lentiviruses, however, NAVIa can be easily applied for activation of any gene. Additionally, lentiviral systems rely on random integration of transgenes, which often results in different copy number across cells leading to variable and unpredictable levels of expression, which are difficult to control precisely. Another disadvantage of transgenes is that they cannot recapitulate the various protein isoforms that are expressed from a given mammalian gene. NAVIa is expected to maintain natural splicing patterns, as we have previously observed that the overall patterns of gene expression and splicing patterns are maintained despite target gene activation of several orders of magnitude (15,41). Additionally, NAVIa allows multiple genes to be upregulated simultaneously by using additional sgRNAs while delivery of multiple heterologous genes using viral vectors can be challenging.

One important concern about the NAVIa system is that it is prone to Cas9 off-target nuclease activity (42). Such activity may lead to off-target vector integration and the inadvertent upregulation of additional genes. It should be noted that we did not detect off-target integrations in any of the clones that we screened (Supplementary Figure S9, Supplementary Tables S6 and S7). Furthermore, the risk of off-target integration can be mitigated by using truncated sgRNAs (43) or enhanced versions of Cas9 that have increased specificity (44,45). While CRISPRa is also susceptible to off-target activation (23), one fundamental difference between both systems is that, for sustained gene activation, CRISPRa necessitates the stable expression, or repeated introduction, of heterologous system components, which may have obvious negative implications on their own. In contrast, NAVIa only necessitates transient nuclease activity to integrate a single synthetic element and is easily amenable to customization to reduce or completely eliminate off-target effects. Additionally, NAVIa can easily be adapted for reversible activation by adding LoxP sites to the integration vector, which would allow for removal of the promoter system using Cre recombinase. While this strategy could be used to remove the synthetic promoter, it should be noted that Cre-Lox recombination leaves a genomic scar that may affect expression levels. Furthermore, this technique should be avoided if a multiplexed approach is being used, as the presence of multiple LoxP sites within the genome may result in chromosomal rearrangements or other aberrations.

One significant advantage of NAVIa over existing CRISPRa methods is the rapid and facile generation and screening of stable cell lines with tunable or programmable properties and a highly predictable pattern of integration. Inducible CRISPRa methods have been developed by integrating a tetracycline-inducible Cas9-based transcriptional activator at random genomic loci (24,25). Induction of target gene expression with these systems requires persistent expression of the sgRNA while expression of the ATF, and ultimately target gene activation, is controlled by treatment with doxycycline. Although these systems are tunable, they also exhibit significant background expression in the absence of doxycycline (25). In contrast, NAVIa replaces native promoters via targeted integration of a tetracycline-inducible promoter to achieve a rapid response to the inducer while avoiding unpredictable lentiviral integration patterns.

Another potential limitation of NAVIa in these experiments was the integration of two promoters in different orientations. While this approach ensures that one promoter is always positioned in the correct orientation for overexpression of the target gene, it is possible that the other promoter can modify expression in the opposite orientation. While this shortcoming also occurs with bidirectional gene activation induced by CRISPRa, it can be overcome by simply using a TV with a single promoter and isolating clones with only integrations in the desired orientation. Though this alternative strategy requires screening, it effectively prevents potential aberrant activation at the opposite end of the vector. In general, for experiments simply focused on activating the gene of interest and where a stable cell line is not required, the use of the bidirectional promoter will ensure that upregulation of the gene of interest is achieved in most cells within a pooled population. However, for creation of stable cell lines it might be desirable to use the single promoter, which ensures that neighboring genes will not be affected.

In summary, the robust levels of activation, multiplexing capabilities, and adaptability make NAVIa an attractive new platform for a variety of synthetic biology applications including metabolic engineering, drug screening, and signal transduction pathway analysis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Charles Gersbach for providing the Golden Gate plasmids for cloning sgRNAs. We thank the Office of Undergraduate Research at the University of Illinois for a Summer Undergraduate Research Fellowship to Nathan Tague.

FUNDING

ZJU-Illinois Institute Research Program, American Heart Association Scientist Development Grant [17SDG33650087], and National Institutes of Health [R01GM127497] to P.P.; National Science Foundation Graduate Research Fellowship Program [DGE-1746047] to M.G.

Conflict of interest statement. A.B., W.W., and P.P. have filed patent applications related to genome editing and gene activation.

REFERENCES

1. Gersbach,C.A. and Perez-Pinera,P. (2014) Activating human genes with zinc finger proteins, transcription activator-like effectors and

CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert Opin. Ther. Targets*, **18**, 835–839.
2. Didovyk,A., Borek,B., Tsimring,L. and Hasty,J. (2016) Transcriptional regulation with CRISPR-Cas9: principles, advances, and applications. *Curr. Opin. Biotechnol.*, **40**, 177–184.
3. Joung,J., Konermann,S., Gootenberg,J.S., Abudayyeh,O.O., Platt,R.J., Brigham,M.D., Sanjana,N.E. and Zhang,F. (2017) Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.*, **12**, 828–863.
4. Fellmann,C., Gowen,B.G., Lin,P.C., Doudna,J.A. and Corn,J.E. (2017) Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.*, **16**, 89–100.
5. Beerli,R.R., Dreier,B. and Barbas,C.F. 3rd. (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 1495–1500.
6. Maeder,M.L., Linder,S.J., Reyon,D., Angstman,J.F., Fu,Y., Sander,J.D. and Joung,J.K. (2013) Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods*, **10**, 243–245.
7. Perez-Pinera,P., Ousterout,D.G., Brunger,J.M., Farin,A.M., Glass,K.A., Guilak,F., Crawford,G.E., Hartemink,A.J. and Gersbach,C.A. (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods*, **10**, 239–242.
8. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
9. Zhang,F., Cong,L., Lodato,S., Kosuri,S., Church,G.M. and Arlotta,P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.
10. Cheng,A.W., Wang,H., Yang,H., Shi,L., Katz,Y., Theunissen,T.W., Rangarajan,S., Shivalila,C.S., Dadon,D.B. and Jaenisch,R. (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.*, **23**, 1163–1171.
11. Farzadfard,F., Perli,S.D. and Lu,T.K. (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.*, **2**, 604–613.
12. Gilbert,L.A., Larson,M.H., Morsut,L., Liu,Z., Brar,G.A., Torres,S.E., Stern-Ginossar,N., Brandman,O., Whitehead,E.H., Doudna,J.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–451.
13. Maeder,M.L., Linder,S.J., Cascio,V.M., Fu,Y., Ho,Q.H. and Joung,J.K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*, **10**, 977–979.
14. Mali,P., Aach,J., Stranges,P.B., Esvelt,K.M., Moosburner,M., Kosuri,S., Yang,L. and Church,G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol*, **31**, 833–838.
15. Perez-Pinera,P., Kocak,D.D., Vockley,C.M., Adler,A.F., Kabadi,A.M., Polstein,L.R., Thakore,P.I., Glass,K.A., Ousterout,D.G., Leong,K.W. *et al.* (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods*, **10**, 973–976.
16. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
17. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
18. Jinek,M., East,A., Cheng,A., Lin,S., Ma,E. and Doudna,J. (2013) RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.
19. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
20. Doudna,J.A. and Charpentier,E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.
21. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
22. Wright,A.V., Nunez,J.K. and Doudna,J.A. (2016) Biology and applications of CRISPR Systems: harnessing Nature's toolbox for genome engineering. *Cell*, **164**, 29–44.
23. Konermann,S., Brigham,M.D., Trevino,A.E., Joung,J., Abudayyeh,O.O., Barcena,C., Hsu,P.D., Habib,N., Gootenberg,J.S., Nishimasu,H. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, **517**, 583–588.
24. Chavez,A., Scheiman,J., Vora,S., Pruitt,B.W., Tuttle,M., E,P.R.I., Lin,S., Kiani,S., Guzman,C.D., Wiegand,D.J. *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*, **12**, 326–328.
25. Gilbert,L.A., Horlbeck,M.A., Adamson,B., Villalta,J.E., Chen,Y., Whitehead,E.H., Guimaraes,C., Panning,B., Ploegh,H.L., Bassik,M.C. *et al.* (2014) Genome-Scale CRISPR-Mediated control of gene repression and activation. *Cell*, **159**, 647–661.
26. Chakraborty,S., Ji,H., Kabadi,A.M., Gersbach,C.A., Christoforou,N. and Leong,K.W. (2014) A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Rep.*, **3**, 940–947.
27. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.
28. Hilton,I.B., D'Ippolito,A.M., Vockley,C.M., Thakore,P.I., Crawford,G.E., Reddy,T.E. and Gersbach,C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.*, **33**, 510–517.
29. Chavez,A., Tuttle,M., Pruitt,B.W., Ewen-Campen,B., Chari,R., Ter-Ovanesyan,D., Haque,S.J., Cecchi,R.J., Kowal,E.J., Buchthal,J. *et al.* (2016) Comparison of Cas9 activators in multiple species. *Nat. Methods*, **13**, 563–567.
30. Gapinske,M., Tague,N., Winter,J., Underhill,G.H. and Perez-Pinera,P. (2018) Targeted gene knock out using nuclease-assisted vector Integration: hemi- and homozygous deletion of JAG1. *Methods Mol. Biol.*, **1772**, 233–248.
31. Brown,A., Woods,W.S. and Perez-Pinera,P. (2016) Multiplexed targeted genome engineering using a universal nuclease-assisted vector integration system. *ACS Synthetic Biology*, **5**, 582–588.
32. Mancini,A., Xavier-Magalhaes,A., Woods,W.S., Nguyen,K.T., Amen,A.M., Hayes,J.L., Fellmann,C., Gapinske,M., McKinney,A.M., Hong,C. *et al.* (2018) Disruption of the beta1L isoform of GABP reverses glioblastoma replicative immortality in a TERT promoter mutation-dependent manner. *Cancer Cell*, **34**, 513–528.
33. Hayflick,L. (1973) *Tissue Culture*. pp. 220–223.
34. Brown,A., Woods,W.S. and Perez-Pinera,P. (2017) Targeted gene activation using RNA-Guided nucleases. *Methods Mol. Biol.*, **1468**, 235–250.
35. Das,A.T., Tenenbaum,L. and Berkhout,B. (2016) Tet-on systems for doxycycline-inducible gene expression. *Curr. Gene Ther.*, **16**, 156–167.
36. Kabadi,A.M., Ousterout,D.G., Hilton,I.B. and Gersbach,C.A. (2014) Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.*, **42**, e147.
37. Bell,R.J., Rube,H.T., Kreig,A., Mancini,A., Fouse,S.D., Nagarajan,R.P., Choi,S., Hong,C., He,D., Pekmezci,M. *et al.* (2015) Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*, **348**, 1036–1039.
38. Suzuki,K., Tsunekawa,Y., Hernandez-Benitez,R., Wu,J., Zhu,J., Kim,E.J., Hatanaka,F., Yamamoto,M., Araoka,T., Li,Z. *et al.* (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144–149.
39. Kosicki,M., Tomberg,K. and Bradley,A. (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**, 65–771.
40. Niu,D., Wei,H.J., Lin,L., George,H., Wang,T., Lee,I.H., Zhao,H.Y., Wang,Y., Kan,Y., Shrock,E. *et al.* (2017) Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science*, **357**, 1303–1307.
41. Polstein,L.R., Perez-Pinera,P., Kocak,D.D., Vockley,C.M., Bledsoe,P., Song,L., Safi,A., Crawford,G.E., Reddy,T.E. and Gersbach,C.A. (2015) Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res.*, **25**, 1158–1169.

42. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
43. Fu,Y., Sander,J.D., Reyon,D., Cascio,V.M. and Joung,J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.*, **32**, 279–284.
44. Kleinstiver,B.P., Pattanayak,V., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Zheng,Z. and Joung,J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.
45. Slaymaker,I.M., Gao,L., Zetsche,B., Scott,D.A., Yan,W.X. and Zhang,F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–88.