Engineering Synthetic TAL Effectors with Orthogonal Target Sites

7584–7595 Nucleic Acids Research, 2012, Vol. 40, No. 15 Published online 11 May 2012
doi:10.1093/nar/gks404

Engineering synthetic TAL effectors with orthogonal target sites

Abhishek Garg[1], Jason J. Lohmueller[1], Pamela A. Silver[1,2,*] and Thomas Z. Armel[1]

[1]Department of Systems Biology, Harvard Medical School and [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

Received March 22, 2012; Revised April 17, 2012; Accepted April 18, 2012

ABSTRACT

The ability to engineer biological circuits that process and respond to complex cellular signals has the potential to impact many areas of biology and medicine. Transcriptional activator-like effectors (TALEs) have emerged as an attractive component for engineering these circuits, as TALEs can be designed *de novo* to target a given DNA sequence. Currently, however, the use of TALEs is limited by degeneracy in the site-specific manner by which they recognize DNA. Here, we propose an algorithm to computationally address this problem. We apply our algorithm to design 180 TALEs targeting 20 bp cognate binding sites that are at least 3 nt mismatches away from all 20 bp sequences in putative 2 kb human promoter regions. We generated eight of these synthetic TALE activators and showed that each is able to activate transcription from a targeted reporter. Importantly, we show that these proteins do not activate synthetic reporters containing mismatches similar to those present in the genome nor a set of endogenous genes predicted to be the most likely targets *in vivo*. Finally, we generated and characterized TALE repressors comprised of our orthogonal DNA binding domains and further combined them with shRNAs to accomplish near complete repression of target gene expression.

INTRODUCTION

A central goal of synthetic biology is the creation of gene regulatory circuits that specifically and robustly control gene expression in response to cell state and environmental cues (1–4). While much progress has been made toward developing genetic systems that detect biological signals, the ability to integrate these signals has been limited by the lack of modular and mutually orthogonal genetic elements available for use. Additionally, the functionality of these systems can be hampered by unwanted interference with the host cell machinery. The generation of high-fidelity gene circuits would thus benefit from a set of mutually orthogonal synthetic regulatory components that have minimal effects on endogenous cell machinery. In the case of transcriptional systems, it would be ideal to have a set of transcriptional regulators that would only target DNA sequences that exist within the artificial circuit. Such regulators would have minimal affinity for DNA sequences present in the endogenous promoter regions of the host cell, thus minimizing unwanted effects on host gene expression (Figure 1). Transcription factors with programmable DNA binding domains offer one potential approach toward this goal. Transcriptional activator-like effectors (TALE) proteins have been recently demonstrated to have modular and predictable DNA binding domains, thereby allowing for the *de novo* creation of synthetic transcription factors that bind any DNA sequence of interest (5–10).

Originally discovered in phytopathogenic bacteria of the genus *Xanthomonas,* TALE proteins are made up of three distinct regions: (i) an N-terminal region housing the protein secretion and translocation signals, (ii) a central repeat domain composed of a series of tandem repeats containing repeat variable di-residues (RVDs) that specifically recognize and bind DNA and (iii) a C-terminal domain containing two nuclear localization signals (NLSs) and a transcriptional activation domain (Figure 2A) (11–14). The central DNA-binding domain is composed of a variable number of 33–35 amino acid repeats such that each binding domain recognizes a different DNA base pair (bp) and can be recombined to recognize any given DNA sequence (12,15). Recent studies have deciphered the code by which the repeat elements bind to DNA, showing that the residues at amino acid positions 12 and 13 in each repeat determine which nucleotide is

*To whom correspondence should be addressed. Tel: +1 617 432 6401; Fax: +1 617 432 5012; Email: pamela_silver@hms.harvard.edu
Present address:
Pamela A. Silver, Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

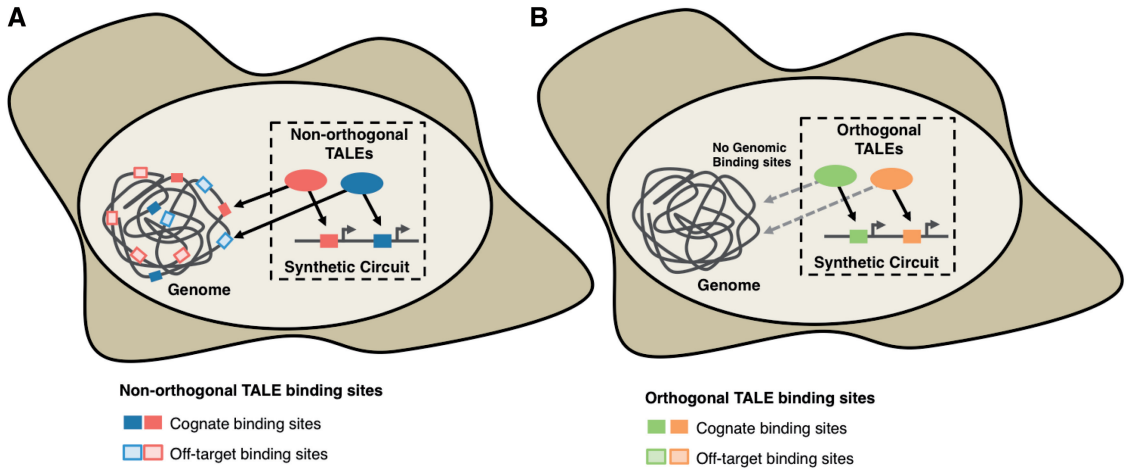
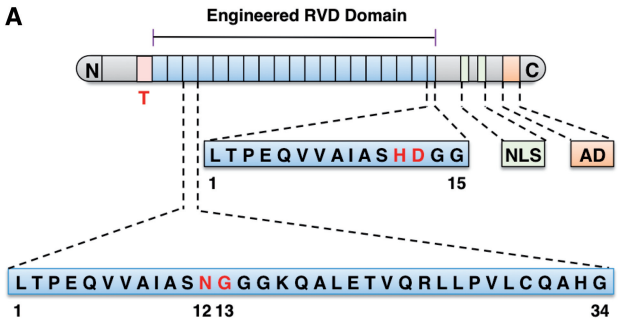


Figure 1. Orthogonal TALEs as ideal regulatory components for insulated synthetic gene circuits. (**A**) Non-orthogonal TALEs designed to bind and regulate gene expression of a synthetic gene circuit may also bind to cognate and off-target (containing mismatches) binding sequences in the endogenous promoter regions in the genomic DNA. (**B**) Orthogonal TALEs bind and regulate gene expression of a synthetic gene circuit and have no predicted binding sites in the endogenous promoter regions.

preferentially bound (Figure 2B) (16,17). The modularity of these repeat elements has enabled TALEs to become a powerful tool, allowing for the creation of synthetic transcriptional activators that can target a specific DNA sequence and activate a desired gene (6,7). Furthermore, because of the proteins' modular nature, TALEs are amenable to hierarchical ligation-based construction strategies, enabling the development of large libraries of proteins (5,7,18–20).

At present, however, drawbacks to the use of TALEs as targeted transcription factors exist. Most notably, each TALE repeat does not bind to a given DNA base pair with perfect complementarity (Figure 2B) (21,22). While it has been shown that in some cases including a single mismatch in the binding site of a given TALE can significantly inhibit its off-target activity, there are known instances where designed TALEs have been demonstrated to bind to unintended off-target DNA sequences that differ from their cognate target sequence by up to 3 bp (as defined by the TALE binding code) (16,17). These observations indicate that while a synthetic TALE can be designed to efficiently target a given DNA sequence, unintended off-target effects can frequently occur and may limit the utility of TALEs for specifically controlling the expression of a targeted gene (Supplementary Methods). This limitation also restricts the application of TALEs as components of synthetic circuits where orthogonality to the host cell's genome is an important constraint.

We have developed an algorithm that allows one to computationally design TALEs with cognate binding sites that are at least a given number of mismatches away from a set of DNA sequences. We apply our algorithm to design TALEs with 20 bp cognate binding sites that are at least 3 nt mismatches away from all 2000 bp putative human promoter sequences and at least 4 nt mismatches from 500 bp putative human promoter sequences. These TALEs represent a potentially powerful set of insulated transcriptional regulators for the construction of synthetic gene circuits. We generated



B

RVD	Amino acid sequence	Preferred Target Nucleotide
NI	LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG	A
HD	LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG	C
NG	LTPQQVVAIASNGGGKQALETVQRLLPVLCQAHG	T
NK	LTPDQVVAIANNKGGKQALETLQRLLPVLCQAHG	G

Figure 2. TALE protein architecture and DNA binding specificities. (**A**) Schematic of a representative TALE protein with 18.5 repeat variable di-residue (RVD) domains. Each RVD domain is composed of 34 amino acids and differs only in the variable amino acids highlighted in red. The C-terminal RVD domain is a 15 amino acid half repeat domain. The two endogenous NLS domains and the endogenous activation domain (AD) present in naturally occurring TALEs were replaced with SV40 NLSs and the VP64 activation domain, respectively. (**B**) The amino acid sequences and the preferred target nucleotides of RVD domains NI, HD, NG from AvrBs3 and RVD domain NK from pthA2.

DNA constructs encoding eight of these TALEs as transcriptional activators and assessed their function in human cells. We demonstrate that each TALE effectively activates transcription from its targeted binding site and that the TALE activators are mutually orthogonal in their activities. We also show that the TALEs do not activate transcription from artificial promoters containing binding sites comparable to potential off-target sites in human

promoter regions and provide additional evidence that the TALEs do not activate their closest off-target endogenous genes. Finally, we use two of the TALE DNA binding domains to generate TALE repressors and demonstrate strong TALE-mediated repression of a reporter gene. We further combine these TALE repressors with synthetic shRNAs targeting the same reporter to obtain even stronger, near complete gene repression. Our methodologies and TALE transcription factors address a major gap in synthetic biology and provide a new set of tools toward the design of robust genetic circuits that function orthogonally to the cells in which they are utilized (23–26).

MATERIAL AND METHODS

Human genome DNA sequences

The sequences corresponding to the 2000 bp regions upstream of all annotated transcription start sites (TSSs) in human RefSeq genes with annotated 5′ untranslated regions (UTRs) were downloaded from the UCSC Genome browser website (http://genome.ucsc.edu/). If multiple upstream regions per RefSeq gene were found due to multiple annotates TSSs, then all upstream regions were used for computing orthogonal 20-mers. Downloaded sequence files correspond to the February 2009 assembly of the human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37).

Recombinant DNA constructs of TALEs and reporters

Amino acid sequences encoding all TALE constructs were derived from the AvrBs3 amino acid sequence (GenBank locus id. CAA34257.1), including sequences encoding the sub-modules corresponding to the constant 5′ region, variable repeats regions (for di-residues HD, NI and NG) and the constant 3′ region. Within these sequences, the naturally existing NLS regions and activation domains in AvrBs3 were identified in the 3′ constant region and replaced with mammalian SV40 NLS and VP64 activation domains. For TALE repressors, the VP64 activation domain was replaced with the KRAB transcriptional repression domain. DNA sequences encoding these components were codon-optimized for expression in human cells and synthesized by Integrated DNA Technology (Coralville, IA, USA). The exact positions and sequences used are listed in Supplementary Methods. These components were combined to generate TALE expression constructs using a hierarchal cloning scheme outlined in Supplementary Methods. t2A and mCherry were combined to full length TALE activator coding regions and t2A and DsRed-shRNA constructs were combined to full-length TALE repressor coding regions using BioBrick cloning. These complete coding regions were cloned into the NheI and NotI sites of pCDNA5insVector for expression from the CMV promoter (27–29).

Reporter constructs for activators and repressors were cloned using BioBrick assembly, cut with SpeI and NotI, and cloned between the SpeI and NotI sites of pCDNA5/FRT/TO for mammalian expression (Invitrogen, Carlsbad, CA, USA). Finally, to create combined TALE repressor and shRNA reporter constructs, shRNA target sites FF4′ and FF6′ were cloned into the NotI site of the CFP reporter constructs of the TALE repressors.

Cell culture

The human osteosarcoma-derived epithelial cell line U-2OS (American Type Culture Collection, Manassas, VA, USA) was maintained at 37°C, 5% CO_2 in growth medium (McCoy's 5A medium supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin). The human embryonic kidney cell line HEK293 (American Type Culture Collection, Manassas, VA, USA) was maintained at 37°C, 5% CO_2 in growth medium (Dulbecco's Modified Eagle Medium supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin). All transfections were performed in 12-well plates seeded with ~175 000 cells using 3 μl Lipofectamine LTX transfection reagent and 1 μl PLUS reagent (Invitrogen, Carlsbad, CA, USA). All TALE activator transfections were performed in U-2OS cells and used 25 ng of TALE expression plasmid with 975 ng of reporter plasmid in 1 ml of growth medium. TALE repressor experiments were performed in HEK293 cells and used 100 ng of TALE expression plasmid with 10 ng of reporter plasmid and 890 ng of empty pCDNA5insVector in 1 ml of growth medium.

Microscopy

All microscopy was performed on live cells in glass-bottomed wells (MatTek, Ashland, MA, USA) in phenol red-free growth medium 24 h post-transfection. Cells were imaged using a Nikon TE-2000 microscope with a 20× PlanFluor NA = 0.5, DIC M/N2 objective and collected with an ORCA-ER charge-coupled device camera. Data collection and processing were performed with Metamorph 7.0 software (Molecular Devices, Sunnyvale, CA, USA). All images for a given experimental set and the corresponding controls were collected with the same exposure times, averaged over three frames and underwent identical processing.

Flow cytometry

Approximately 30 000 cells from each transfected well were analyzed using an LSRII cell analyzer (BD Biosciences, San Jose, CA, USA) in three biological replicates. Cells were trypsinized with 0.1 ml of 0.25% trypsin-EDTA, pelleted and resuspended in 100 μl of Dulbecco's phosphate buffered saline containing 0.1% FBS. For activator experiments, output was assayed 24 h post-transfection. The total AmCyan fluorescent protein (CFP) signal of mCH+ cells was calculated by multiplying the frequency of CFP+ cells in the mCH+ population by the mean CFP signal of these double positive cells. The fold change of AmCyan reporter fluorescence was then calculated as the ratio of total AmCyan fluorescence intensity of cells transfected with on-target TALE expression plasmids to cells transfected with reporter plus an off-target input. For repressor experiments, outputs was assayed 48 h post-transfection and fold change of AmCyan reporter fluorescence was calculated as the

ratio of total AmCyan fluorescence intensity of DsRED+ cells transfected with on-target TALE-shRNA expression plasmids to DsRED+ cells transfected with a reporter plus an off-target input. To isolate the effects of TALEs and shRNAs, expression constructs with different combinations of TALE5R, TALE8R, FF4 and FF6 represented different on- and off-target combinations depending on the co-transfected reporter (Supplementary Table S9).

Quantitative PCR

For mRNA quantification, mCherry positive U-2OS cells were sorted and collected 48 h post-transfection. RNA was extracted from mCherry positive cells using the RNeasy mini kit (Qiagen, Valencia, CA, USA), and mRNA levels were quantified using the SYBR Green Assay (Applied Biosystems, Foster City, CA, USA). The mRNA to cDNA conversion was performed using the SuperScript III RT kit (Invitrogen, Carlsbad, CA, USA). Three biological replicates per sample and three technical replicates per assay were analyzed for absolute quantification of mRNA levels in transfected cells. Two biological replicates were analyzed for the mRNA levels quantification of OSGIN2 and ZC3H10 in cells transfected with TALE5 and TALE8, respectively. Relative transcript levels were assessed using the $2^{-\Delta\Delta Ct}$method (45) with GAPDH as a reference gene. Statistical comparison between groups was made by the pair-wise fixed reallocation randomization test using the publicly available Relative Expression Software Tool (REST) (30).The off-target and on-target DNA sequences of TALEs are detailed in Supplementary Table S7. Primer sequences used for qPCR are detailed in Supplementary Table S7.

Algorithm implementation

The algorithm was implemented in C++ and the software binaries are made available for download at http://silver.med.harvard.edu/tale.html. Further details about the algorithm are provided in Supplementary Methods. All the results presented here were obtained by running our software on the Harvard Medical School shared research cluster of computation nodes.

RESULTS

Design and implementation of an algorithm for finding orthogonal TALE binding sites

We first sought to computationally design a set of TALEs that bind to 20 bp nucleotide sequences (20-mers) and are orthogonal to human promoter regions. A TALE is defined to be orthogonal to a set of sequences if it is not predicted to bind to any sequence in that set. In this context, the number of base pair mismatches between a TALE's target sequence and a potential off-target sequence (also referred to as the *hamming distance* between the two sequences) is the main determinant of the orthogonality of the TALE. Thus, a large hamming distance between the TALE target site and a potential off-target sequence corresponds to a lower chance of the TALE binding to that off-target sequence.

To design synthetic TALEs that function orthogonally to a set of non-intended target sites, we developed an algorithm based on the *farthest string problem*. Given a set, S, of n-mers defined over an alphabet, Γ, (e.g. Γ = {A,C,G,T}), the objective of the farthest string problem is to find an *n*-mer (over the alphabet Γ) that has the largest minimum hamming distance to *n*-mers in set S. The farthest string problem belongs to a class of NP-hard problems for which no polynomial time solution is known to exist (31). Therefore, it may take an exponential amount of time to enumerate all possible 4^{20} nucleotide sequences and test each to find a 20-mer at a maximum hamming distance from the set of genomic 20-mers. At present, no algorithm exists to efficiently compute a set of such *n*-mers. However, by designing careful heuristics, our algorithm can efficiently find a list of 20-mers that are orthogonal to human genome promoter regions by a hamming distance of 3 bp or more.

The steps followed by our algorithm are outlined in Figure 3. We began by using a sliding window approach to enumerate all possible 20-mers present across both DNA strands in the promoter region of all genes in the human genome. We define promoter regions as the 2000 bp regions upstream of the TSS of each gene. Because the presence of a 5′ T has been demonstrated to be a necessary condition for efficient TALE binding, 20-mers that do not begin with T were not considered, yielding a total of 17×10^6 20-mers that are potential TALE binding sites (16). To further reduce the number of 20-mers, the parental set of 17×10^6 20-mers was divided into subsets, such that each subset could be represented by a single 20-mer within a 7 bp hamming distance from all sequences in that subset (Supplementary Figure S1). Due to the reverse triangle inequality property of hamming distances, all 20-mers that

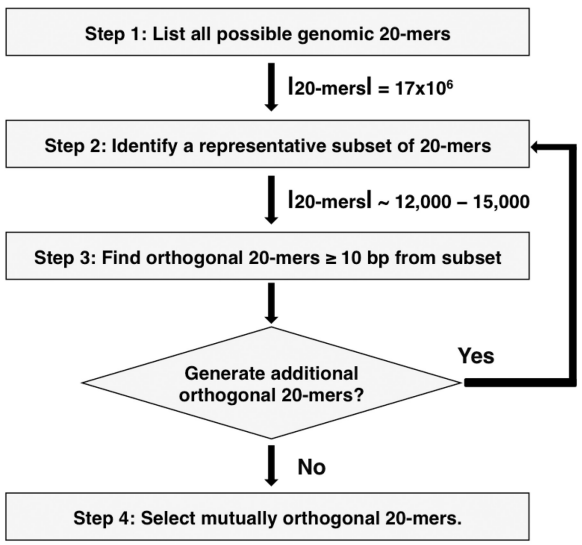


Figure 3. Flowchart enumerating the steps used in our algorithm to compute orthogonal 20-mers. Steps 1 and 2 describe the process used to reduce the set of genomic 20-mers. Steps 2 and 3 describe the process of obtaining 20-mers orthogonal to the genomic set. Steps 2 and 3 of the algorithm can be iterated until the desired number of orthogonal sequences has been computed. Finally, the resulting sets of TALEs are checked for mutual orthogonality to avoid cross-interference within the synthetic circuits.

are at a minimum 10 bp hamming distance from these representative sequences will also be at a minimum hamming distance of 3 bp from the parental set of 17×10^6 genomic 20-mers (Supplementary Figure S1). Our algorithm uses symbolic modeling techniques and Boolean algebra to find all possible 20-mers at a minimum hamming distance of 10 bp from representative sequences of each subset (Supplementary Methods). Multiple solutions to finding such subsets exist and each solution is typically comprised of 12 000–15 000 subsets, each having a representative 20-mers. By generating multiple sets of representative 20-mers and applying our algorithm iteratively, we identified over 180 potential binding sites for synthetic TALEs at a minimum hamming distance of 3 bp from any 20-mer in the promoter regions of the human genome (Supplementary Table S2).

We chose to generate and characterize 8 of these 180 TALEs predicted to be orthogonal to human promoter regions (Table 1). Chosen TALEs had a hamming distance of 3 bp from all 2000 bp genomic promoter regions and a hamming distance of 4 bp from 500 bp genomic promoter regions. The hamming distance to the more stringent 500 bp genomic promoter regions was used as an additional criterion as native transcription factor binding sites are known to be highly concentrated within these 500 bp regions proximal to the TSS (32–35). From our set of 150 synthetic TALEs, 100 proteins possessed a minimum hamming distance of 4 bp from 500 bp proximal promoter regions, while the remaining 50 proteins had a hamming distance of 3 bp. To minimize potential cross-activation between the selected TALEs, we also ensured that the eight selected TALEs were predicted to be mutually orthogonal.

In vivo characterization demonstrates activity and mutual orthogonality of synthetic TALE activators

To generate each of our eight computationally designed TALEs for assaying *in vivo*, a library of subparts was synthesized containing both individual di-residue repeats and each pair-wise combination of repeats, codon-optimized for expression in mammalian cells. Individual TALEs were created using a hierarchical, modular cloning strategy that leverages type IIS restriction enzymes to readily combine members of a library of subparts into any desired TALE (Supplementary Figure S2). The modular cloning scheme we use is similar to the techniques reported in the recent literature (18,19,20,36,37). For each protein, both native NLSs were replaced with eukaryotic versions, and the native activation domain was replaced

Table 1. Constituent RVDs and cognate binding sites of the 8 TALEs that were constructed and functionally characterized

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
TALE 1		HD	NI	NI	NG	NI	HD	NG	NG	NI	HD	NI	NI	NI	HD	NG	HD	HD	NG	NG
	T	C	A	A	T	A	C	T	T	A	C	A	A	A	C	T	C	C	T	T
TALE 2		HD	HD	NI	HD	HD	NI	NI	NI	NG	NG	HD	NI	NI	HD	NI	HD	NG	NG	NG
	T	C	C	A	C	C	A	A	A	T	T	C	A	A	C	A	C	T	T	T
TALE 3		HD	NI	NG	HD	NG	NI	HD	NI	NI	HD	NI	HD	NG	NI	HD	NG	NI	NG	NG
	T	C	A	T	C	T	A	C	A	A	C	A	C	T	A	C	T	A	T	T
TALE 4		HD	HD	HD	NI	NI	NG	NI	HD	NI	HD	NG	NI	NG	NI	NI	HD	NI	HD	NI
	T	C	C	C	A	A	T	A	C	A	C	T	A	T	A	A	C	A	C	A
TALE 5		NI	NI	HD	NG	NG	NI	HD	HD	NG	NG	HD	NG	HD	NI	NI	HD	NI	HD	NI
	T	A	A	C	T	T	A	C	C	T	T	C	T	C	A	A	C	A	C	A
TALE 6		NI	NG	HD	HD	NG	HD	NG	NG	NI	HD	NI	NI	NG	NI	NG	HD	HD	HD	NI
	T	A	T	C	C	T	C	T	T	A	C	A	A	T	A	T	C	C	C	A
TALE 7		NI	HD	NG	NG	NI	HD	HD	HD	NG	NI	NI	HD	HD	HD	NI	NI	NG	NG	NG
	T	A	C	T	T	A	C	C	C	T	A	A	C	C	C	A	A	T	T	T
TALE 8		NI	NG	NI	HD	NG	NI	NG	HD	HD	NI	NI	NG	HD	HD	NI	NI	HD	NG	NG
	T	A	T	A	C	T	A	T	C	C	A	A	T	C	C	A	A	C	T	T
TALE OSGIN2		HD	HD	NG	HD	HD	HD	HD	NI	HD	HD	NG	NG	NG	NI	NI	NG	NG	NG	NG
	T	C	C	T	C	C	C	C	A	C	C	T	T	T	A	A	T	T	T	T
TALE ZC3H10		NI	HD	HD	NI	NG	NI	NG	HD	HD	HD	NI	NG	HD	HD	NI	NI	HD	NG	HD
	T	A	C	C	A	T	A	T	C	C	C	A	T	C	C	A	A	C	T	C

with the VP64 mammalian transcriptional activation domain. TALEs were expressed from the cytomegalovirus (CMV) promoter and tagged with an auto-catalytically cleaved t2A peptide fused to mCherry fluorescent protein as a transfection control (Figure 4A).

The ability of our synthetic proteins to recognize a binding site and activate gene expression was tested by co-transfecting TALE expression constructs with reporter plasmids containing a 20-mer binding site driving expression of two tandem copies of the CFP fused to an NLS (Figure 4B). Experiments were performed in the U-2OS human osteosarcoma cell line and assayed by fluorescence microscopy and flow cytometry 24 h post-transfection. Each TALE was co-transfected with its corresponding binding site reporter plasmid to determine if it was capable of activating transcription from its targeted reporter, as well as with reporter plasmids containing binding sites for the seven other constructed TALEs in order to ensure that all proteins are mutually orthogonal. Results from fluorescence microscopy indicate that all TALEs were efficiently expressed, as determined by presence of mCherry positive cells (Supplementary Figure S3). Furthermore, the TALEs efficiently activated gene expression from promoters containing their cognate binding site and not those targeted by other TALEs in the set, indicating that our synthetic TALEs function in a mutually orthogonal manner (Figure 5A). Flow cytometry was performed to quantify TALE-activated CFP expression from each promoter. Activity was measured as the total CFP signal of mCherry positive cells. As a control, an off-target TALE was co-transfected with each TALE reporter plasmid and the level of activation for each synthetic TALE was calculated relative to this off-target control. These results confirmed our fluorescence microscopy findings, with synthetic TALEs demonstrating a 10- to 102-fold induction of the CFP reporter with no significant signal observed for off-target binding sites (Figure 5B and Supplementary Table S1).

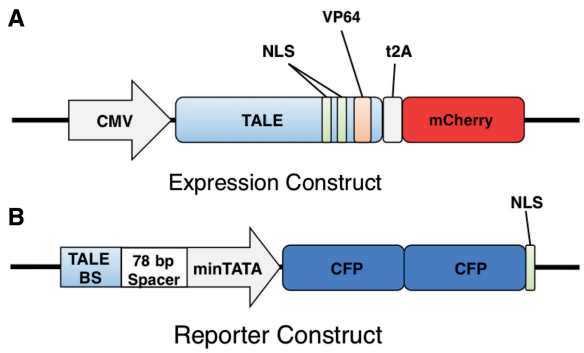


Figure 4. (**A**) Schematic of TALE expression constructs. Each TALE coding region was cloned into a mammalian expression vector downstream of the CMV promoter. All synthetic TALEs were also tagged with an self-cleaving t2A:mCherry fluorescent protein as a transfection control. (**B**) Schematic of TALE reporter constructs. Reporter constructs were generated by cloning a 20 bp TALE target sequence upstream of a minimal TATA box separated by a 78 bp spacer region. Binding of a TALE activator to the 20 bp target sequence drives expression of two tandem copies of NLS-tagged CFP cloned downstream of the TALE-responsive promoter as an output for TALE functionality.

Synthetic TALEs do not activate transcription of a set of off-target endogenous genes

To investigate the orthogonality of our TALEs to potential genomic promoter binding sites, we began by assessing the effect of target site mismatches on the ability of TALEs to bind a given 20-mer. It has previously been shown that TALE activity generally decreases with the number of mutations in its target site (18,22,37,38,39). However, as positional and contextual effects of these mutations have also been reported, it is important to analyze the specific effect of mutations in the context of our TALEs that have a different protein architecture and bind to longer DNA sequences (20 bp) than those previously studied. TALE8 was chosen as a representative protein and reporter constructs were generated with 20-mers at a hamming distance of 1–7 from TALE8's on-target binding site. To avoid potential position-specific bias, mismatches were distributed evenly throughout the binding sites (Figure 6A). TALE8 was co-transfected with each reporter construct, and reporter expression was assayed by fluorescence microscopy and flow cytometry with TALE5 serving as an off-target control. Expression from reporter constructs was observed to decrease with the hamming distance and 20-mers at a hamming distance of 3 bp from the on-target site exhibited output signal that was one-tenth of the full signal, and 20-mers at a hamming distance of 4 bp or more from the on-target site exhibited an output signal comparable to background (Figure 6B).

We next sought to ascertain the influence of mismatch position on protein function. Three additional reporter plasmids were generated for TALE8 with a hamming distance of 3 bp. The positions of these mismatches were localized to either the 5′-end, the 3′-end or the center of the target site (Figure 6A). Our results illustrate that mismatches in the 5′-end and center of the target site abolish TALE activated expression, while mismatches in the 3′ end appear to have less of an impact, more closely resembling mismatches uniformly distributed throughout the binding site (Figure 6C). These results indicate that the location of mismatches should be considered when designing orthogonal TALEs. Within the 2 kb promoter regions, the longest matching endogenous sequences to our designed eight TALEs were at most 14 bp long and these off-target sequences had four or more mismatches in positions 14–20. Thus, our constructed TALEs satisfy the combined constraints set by number and position of mismatches in Figure 6 (Supplementary Table S3).

To more directly characterize the orthogonality of our synthetic TALEs to endogenous promoter regions *in vivo*, we measured mRNA expression levels from the most likely predicted target genes following transfection with two representative TALEs, TALE5 and TALE8. The nearest predicted off-target sequence for TALE5 was in the promoter of oxidative stress induced growth inhibitor family member 2 (OSGIN2), and for TALE8 the nearest predicted off-target sequence was in the promoter of

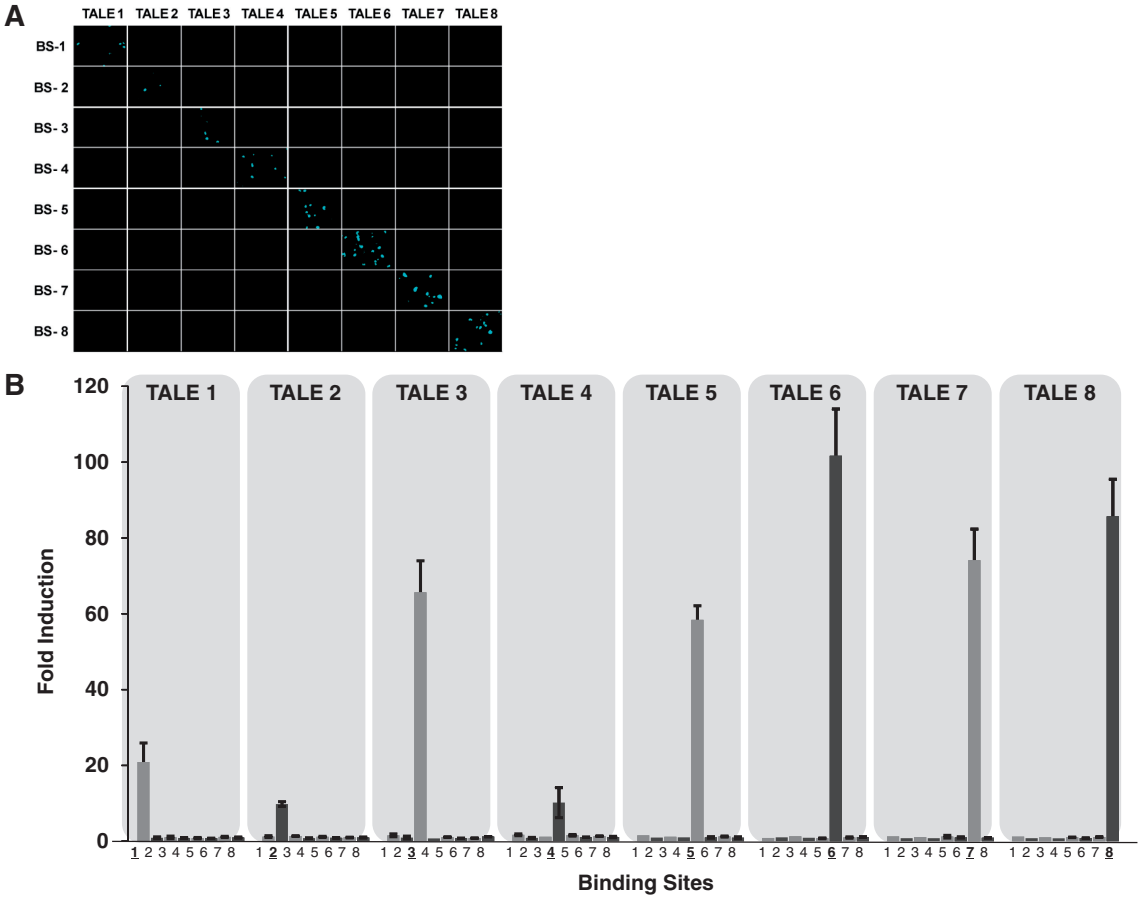


Figure 5. Functional characterization of TALE activators. (**A**) Fluorescence microscopy images of TALE-induced CFP reporter expression. Each column of the 8 × 8 matrix represents U2-OS cells co-transfected with a synthetic TALE and reporter constructs for each 20-mer binding site (BS). The CFP signal is only visible along the diagonal of the matrix, indicating that the TALEs described here function in a mutually orthogonal manner. (**B**) Bar graphs representing mutually orthogonal TALE activity as determined by flow cytometry. The fold induction of CFP expression, as calculated relative to an off-target control TALE, displays values ranging from approximately 10-fold to 100-fold for cognate target sites and demonstrates the functionality and mutual orthogonality of these TALEs.

A

Binding Site	Target Sequence
TALE 8	TATACTATCCAATCCAACTT
TALE 8 + HD-1	TATACTATCCCATCCAACTT
TALE 8 + HD-2	TATACTCTCCAATACAACTT
TALE 8 + HD-3	TATAATATACAATCCCACTT
TALE 8 + HD-4	TATCATATCCAATCCCCCTT
TALE 8 + HD-5	TACCATATCCCATCCAACTC
TALE 8 + HD-7	TCTCATATAAAATAAAACTT
TALE 8 + HD-3 (5`)	TCCCCTATCCAATCCAACTT
TALE 8 + HD-3 (M)	TATACTATCCACCACAACTT
TALE 8 + HD-3 (3`)	TATACTATCCAATCCAAACC

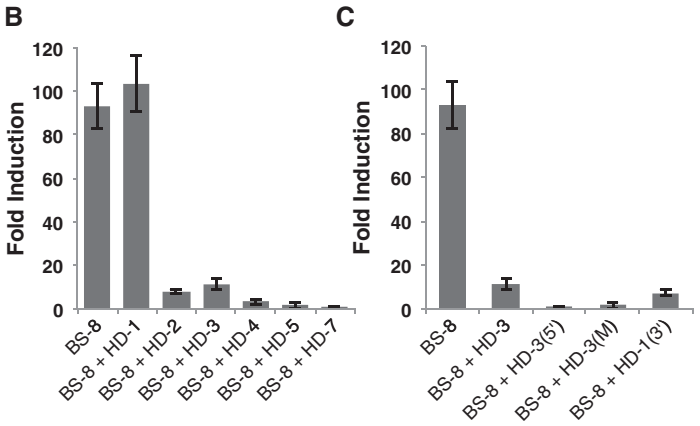


Figure 6. Effect of binding site mutations on TALE-mediated transcriptional activation. (**A**) TALE8 activity in the presence of an increasing number of uniformly distributed binding site mismatches. BS-8 is the corresponding binding site for TALE8 with additional binding sites tested at a hamming distance (HD) of 1–7 bp from BS-8 (HD-1–HD-7). The ability of TALE8 to activate CFP expression from each binding site reporter was measured by flow cytometry relative to TALE5 as an off-target control. The presence of two or more mismatches in the binding site significantly decreases the ability of TALE8 to activate gene expression, with binding sites at a hamming distance of more than 3 bp displaying no reporter activity. (**B**) Effect of binding site mismatch position on TALE activation. The ability of TALE8 to activate gene expression from binding sites with a hamming distance of 3 bp was tested with the position of the mismatches either uniformly distributed, HD-3, localized to the 5′-end of the binding site, HD-3(5′), to the middle of the binding site, HD-3(M) or to the end of the binding site, HD-3(3′). (**C**) Tested DNA binding sequences. Underlined nucleotides represent mismatches with respect to BS-8.

zinc-finger CCCH-type containing 10 (ZC3H10). The targets of each TALE chosen for analysis were determined based on the presence of the closest off-target binding site, having a minimum number of mismatches, in the 500 bp region upstream of the TSS. As a positive control, we designed two TALEs, TALE-OSGIN2 and TALE-ZC3H10, that are predicted to effectively bind in the 500 bp upstream promoter regions of OSGIN2 and ZC3H10, respectively. Off-target sequences for TALE5 and TALE8 and target sequences for TALE-OSGIN2 and TALE-ZC3H10 are listed in Supplementary Table S7. All TALEs were transfected in U-2OS cells and the fold change in mRNA level relative to a mock-transfected control was measured at 48 h post-transfection by qPCR (Figure 7).

Results from qPCR demonstrate that while our positive control, TALE-OSGIN2, is capable of inducing OSGIN2 mRNA expression by 4.8-fold, no significant induction is observed following transfection with TALE5 (Figure 7A). Similarly, transfection with TALE-ZC3H10 leads to a significant induction of targeted ZC3H10 mRNA, while no significant induction is observed following transfection with TALE8 (Figure 7B). In order to ensure that an adequate amount of TALE transcription factor was expressed in cells, we analyzed the fold induction in mRNA expression of off-target genes from TALE5 and TALE8 with 10× higher amount of TALE expression plasmid (Figure 7). We observed no significant induction of off-target genes even in the presence of the higher concentration of TALE expression plasmid. To further investigate the orthogonality of our synthetic TALEs, we assayed mRNA expression of the next four nearest predicted target genes of TALE5 and the next three nearest predicted target genes of TALE8 (Figure 7). In all cases, no significant induction of potential target genes was seen relative to mock-transfected controls, providing further evidence for the orthogonality of these TALEs relative to human promoter regions.

Construction and characterization of TALE repressors

Next, we designed and tested TALE repressor proteins composed of our orthogonal TALE DNA binding domains. We generated TALE repressors TALE5R and TALE8R, by replacing the VP64 activation domain with the KRAB transcriptional repression domain in TALE5 and TALE8 constructs, respectively (Figure 8A). The ability of these TALEs to repress transcription was tested by co-transfecting them with CMV-driven CFP expression vectors containing the cognate TALE binding site located on the transcriptional start site of the CMV promoter. TALE repressors efficiently repressed CFP expression from 36- to 97-fold compared to off-target TALE controls (Figure 8).

Finally, we demonstrated the ability to tightly repress gene expression to near background levels by combining the TALE repressors with shRNAs targeting the same transcripts. We designed expression constructs that co-express a TALE repressor, an shRNA and the DsRed fluorescent protein from the same promoter. The shRNAs were generated in the miR30 context and embedded within the SV40 intron in the DsRED red fluorescence protein gene (40,41). We used the shRNAs, 'FF4' and 'FF6', previously designed to target the Firefly Luciferase gene as they are commonly used as off-target negative control shRNAs and are reported to be orthogonal to endogenous transcripts (41). When co-expressed, the TALE and shRNA combination repressed CFP expression from 740- to 4853-fold. Of note, the level repression mediated by the TALE repressors alone was at least 5-fold higher than that of the shRNAs expressed alone.

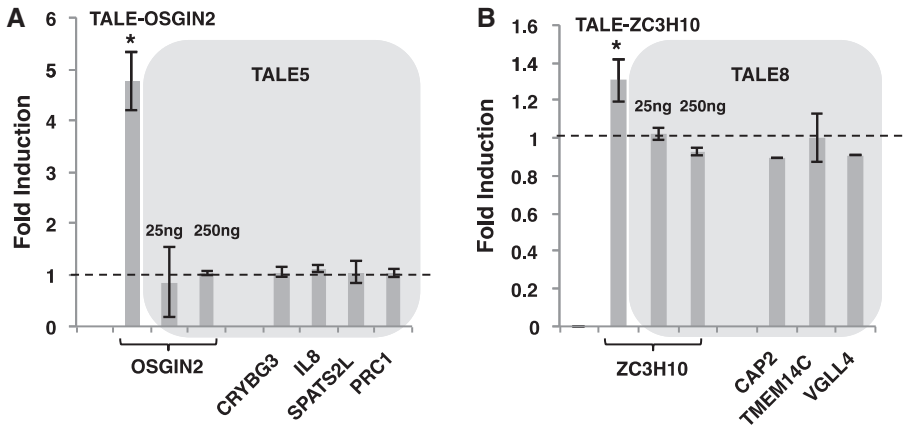


Figure 7. Characterization of TALE-mediated off-target endogenous gene activation *in vivo*. Fold change in mRNA levels of potential target genes following TALE expression. mRNA levels of the most likely target genes of TALE5 and TALE8 were measured by qPCR 48 h post-transfection with the corresponding TALE construct and plotted as fold change over mock-transfected cells. TALE-OSGIN2 and TALE-ZC3H10 are the positive control TALEs predicted to activate the two closest off-target genes of TALE5 and TALE8 respectively. (**A**) A 4.8-fold induction of nearest target gene OSGIN2 by the positive control TALE-OSGIN2, and no significant change in mRNA levels of OSGIN2 and the other four nearest target genes of TALE5 is observed in response to TALE5. The 10× higher concentration (250 ng) of TALE5 also shows no significant induction in mRNA levels of its off-target gene OSGIN2. (**B**) The positive control TALE-ZC3H10 leads to a modest but significant induction of nearest target gene (ZC3H10) of TALE8. There is no significant change in mRNA levels of the four nearest target genes of TALE8 in response to TALE8 expression. Asterisk indicates $P < 0.03$.

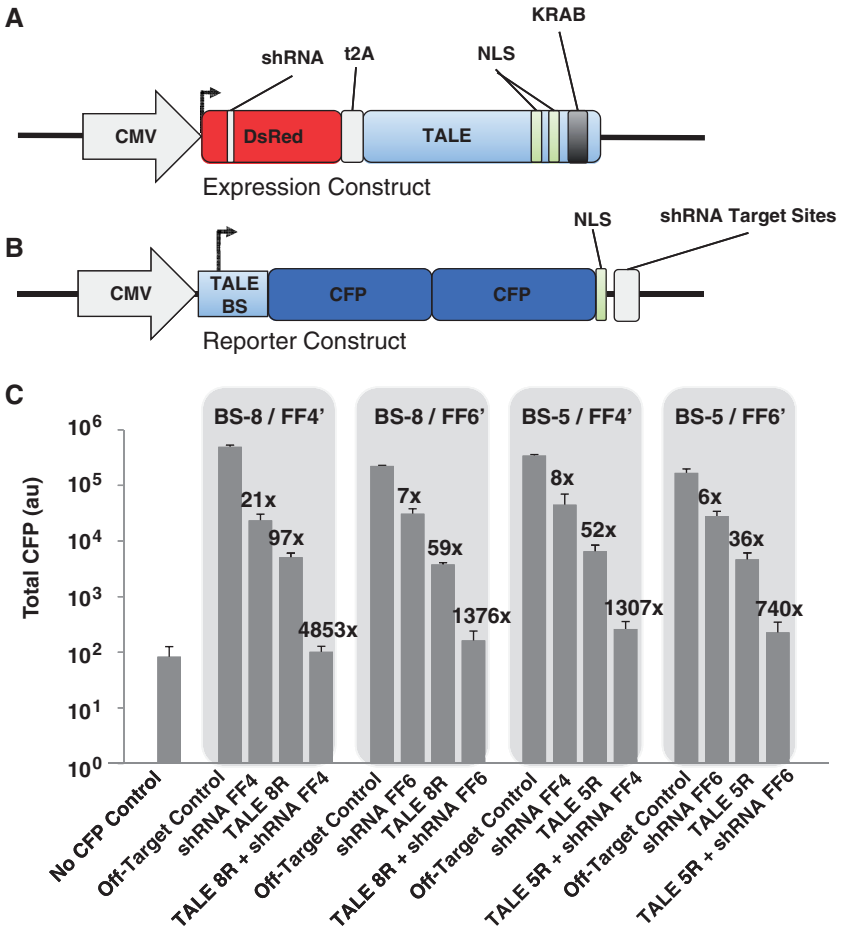


Figure 8. Schematics and characterization of TALE repressor-shRNA constructs. (**A**) The VP64 activation domain of the TALE activators was replaced with the KRAB repression domain and the resulting TALE repressor coding region was cloned into a mammalian expression vector together with a self-cleaving DsRed:t2A fluorescent protein. Synthetic shRNAs were expressed from an intron in the DsRed gene. (**B**) Reporter constructs were generated by cloning a 20 bp TALE target sequence into the TSS of the CMV promoter. On binding its recognition site in the promoter, the TALE represses the constitutive expression of the downstream CFP protein. The reporter construct also contains a single copy of the cognate shRNA recognition sequence in the 3′-UTR, which when recognized by the target shRNA leads to degradation of the CFP transcript. (**C**) TALE repressors, TALE5R and TALE8R were combined with shRNAs, FF4 and FF6, to repress CFP expression from reporter constructs carrying cognate TALE and shRNA recognition sites. Repressions ranged 6× in the case of shRNA alone to ~4800× in the case of shRNA + TALE repressor combination.

DISCUSSION

Robust synthetic networks would enable the ability to sense a wide variety of cellular cues and respond in a desired fashion to modulate cell behavior, but so far efforts to design these networks have been limited by the reliance on a small set of commonly used gene regulatory components. A large set of mutually orthogonal and modular regulatory components would be a useful tool for generating such networks. Additionally, using components for which interference with the host cell's machinery is minimized would help to reduce the chance of unwanted cellular behaviors and system failures.

TALE transcription factors present a powerful tool with many potential applications including use as a set of reliable gene regulatory components for synthetic gene circuits. However, their utility is limited by degenerate binding and the strong potential for off-target effects (6–8,16,19). While recent work has demonstrated the ability of designer TALE activators to turn on expression of desired genes, they have not been optimized to minimize off-target effects and likely activate the expression of genes other than those intended (Supplementary Methods) (37). Here, we present a novel and general method to design TALE DNA binding domains with cognate binding sites orthogonal to a given set of sequences. We create a set of synthetic TALE activators and repressors that specifically recognize and act upon 20 bp binding sites that are at least 3 nt mismatches away from 20 bp sequences contained in all putative human promoter regions.

Applying our algorithmic approach to find TALEs that are specific to a given endogenous gene promoter should be relatively less computationally intensive as the search space for such TALEs is very small compared to the exponentially large search space for TALEs orthogonal to

every human promoter. Starting from the set of all possible TALEs that can bind on a given promoter region, the heuristics presented here based on reverse triangle inequality property of hamming distance can be applied to efficiently screen for TALEs that are orthogonal by a given number of base pairs to the rest of the promoters in the genome.

Our synthetic TALE activators displayed high activation of on-target reporters with levels of activation ranging from 10-fold up to 102-fold and are mutually orthogonal. These activation levels are similar to other recently reported TALEs designed to function in mammalian cells, although we employ a different promoter architecture for TALE expression (37). We further characterized the effects of binding site mismatches on TALE orthogonality by selecting a single TALE and generating synthetic target sites containing between 1 and 7 evenly distributed mismatches. We found that the activation dropped off quickly with an increase in hamming distance—indicating the minimum hamming distance for orthogonality of our TALEs recognizing 20 bp falls in the range of 3–4 bp.

We also found that the distribution of mismatches in the binding site affects TALE protein activity. Testing 20 bp TALE binding sites with sets of three mismatches located at either the 5′ end of the binding site, the 3′ end of the binding site, the middle of the binding site or distributed uniformly throughout the 20-mer, we observed that 3 bp mismatches are able to abolish TALE activation when these mutations are introduced at either the 5′ end or in the middle of TALE binding site (Figure 6C). Three consecutive mutations introduced at the 3′ end of binding site show low off-target activity, about one-tenth of the full factor, as did the three mutations distributed throughout the binding site. These results suggest that for binding sites with a 3 bp hamming distance the position of the mutations should be considered.

With these results in mind, we compared the set of 180 computationally derived orthogonal TALE binding sites to all possible 20-mers in 2000 bp upstream promoter regions of the human genome. We found that for genomic sites predicted to be the most likely targets for our synthetic TALEs, the longest region with perfect complementarity from the 5′ end was <14 bp long for the majority of our synthetic target sites. Furthermore, within this small subset of target sites possessing stretches of sequence complementarity, four or more mutations are typically found between positions 13–20 bp, suggesting that likelihood of a synthetic, orthogonal TALE efficiently binding to a genomic promoter site is extremely low (Supplementary Table S3).

To provide further functional evidence for the orthogonality of our synthetic TALEs to genomic promoter regions *in vivo*, we measured mRNA expression levels from the nine most likely target genes following transfection with two representative TALEs. All potential target genes displayed no increase in mRNA expression levels relative to control, while TALEs designed to specifically target two of those same genes were able to induce mRNA expression up to 4.8-fold. While we cannot rule out the activation of other potential off-target genes by our TALEs, nor the activation of genes by TALE binding to distant enhancer regions outside of the 2 kb promoter regions, these results, combined with data detailing the effect of target site mismatches and bioinformatics approaches, provide evidence supporting the ability of our TALEs to function orthogonally to the human promoter regions.

Next, we designed TALE repressors by replacing the VP64 activation domain in the 3′ constant back region with the KRAB repressor domains. We assayed two synthetic TALE repressors made from our orthogonal TALE DNA binding domains, along with two synthetic shRNAs, and demonstrate that TALE repressors can provide strong transcriptional repression. The TALE-mediated gene repression was more potent than that accomplished by the two shRNAs tested using the same assay. Double repression of a target gene by the LacI transcriptional repressor and an shRNA was previously reported to be capable of tightly controlling transgene expression (42). We show that such regulation is also possible by combining TALE repressors and shRNAs. Combined repression reduced the expression level of target protein to near background levels. As TALEs are highly programmable compared to LacI this result allows for the generation of a set of tightly repressed gene modules and opens the possibility of tightly regulating endogenous target genes. TALE repressors have been shown to be a powerful tool for regulating the expression of genes in yeast and plants (43,44). Our results demonstrate that TALE repressors can work efficiently in mammalian cells as well.

Finally, it is worth noting that our proposed algorithm can easily accommodate additional constraints. For example, it can be readily adapted to identify orthogonal sequences of different lengths and for different sequences including the genomes of other organisms. It could also be modified to find TALEs that have larger hamming distances to especially critical promoter regions. In addition to addressing the problem of generating synthetic circuit components with minimal effects on endogenous genes, the methods that we employ to generate TALEs are general and can be applied to any system requiring specific DNA binding domains. Other potential applications of orthogonal TALE DNA binding domains include TALE nucleases, TALE recombinases or TALE-based DNA methylases, and TALE transcriptional activators and repressors that specifically target endogenous genes. The computational approach and transcription factors presented here provide important tools and methods for the precise engineering of biological systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–10, Supplementary Figures 1–7, Supplementary Methods and Supplementary References [46–48].

ACKNOWLEDGEMENTS

The authors wish to acknowledge all members of the Silver lab for helpful comments and discussion. The authors would also like to acknowledge F. Leinert, Y. P. Hung, D. B. Thompson and L. E. Carey for carefully reading the manuscript.

FUNDING

National Institute of Health [1F32 CA154195-01 to T.Z.A.]; Swiss National Science Foundation (SNSF) (fellowship for prospective researchers to A.G.); DARPA and NIH (to P.A.S.); Wyss Institute (to P.A.S. and J.J.L.). Funding for the open access charge: DARPA.

Conflict of interest statement. None declared.

REFERENCES

1. Andrianantoandro,E., Basu,S., Karig,D.K. and Weiss,R. (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, **2**, 2006.0028.
2. Khalil,A.S. and Collins,J.J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet.*, **11**, 367–379.
3. Haynes,K.A. and Silver,P.A. (2009) Eukaryotic systems broaden the scope of synthetic biology. *J. Cell Biol.*, **187**, 589–596.
4. Tabor,J.J., Salis,H.M., Simpson,Z.B., Chevalier,A.A., Levskaya,A., Marcotte,E.M., Voigt,C.A. and Ellington,A.D. (2009) A synthetic genetic edge detection program. *Cell*, **137**, 1272–1281.
5. Li,T., Huang,S., Zhao,X., Wright,D.A., Carpenter,S., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.*, **39**, 6315–6325.
6. Morbitzer,R., Romer,P., Boch,J. and Lahaye,T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl Acad. Sci. USA*, **107**, 21617–21622.
7. 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.
8. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.
9. Mak,A.N., Bradley,P., Cernadas,R.A., Bogdanove,A.J. and Stoddard,B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716–719.
10. Deng,D., Yan,C., Pan,X., Mahfouz,M., Wang,J., Zhu,J.K., Shi,Y. and Yan,N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720–723.
11. Van den Ackerveken,G., Marois,E. and Bonas,U. (1996) Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell*, **87**, 1307–1316.
12. Boch,J. and Bonas,U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.*, **48**, 419–436.
13. Zhu,W., Yang,B., Chittoor,J.M., Johnson,L.B. and White,F.F. (1998) AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol. Plant Microb. Interact.*, **11**, 824–832.
14. Herbers,K., Conrads-Strauch,J. and Bonas,U. (1992) Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature*, **356**, 172–174.
15. Scholze,H. and Boch,J. (2011) TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.*, **14**, 47–53.
16. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
17. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
18. Morbitzer,R., Elsaesser,J., Hausner,J. and Lahaye,T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.*, **39**, 5790–5799.
19. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N.V., Bogdanove,A.J. and Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
20. Weber,E., Gruetzner,R., Werner,S., Engler,C. and Marillonnet,S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, **6**, e19722.
21. Romer,P., Recht,S. and Lahaye,T. (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc. Natl Acad. Sci. USA*, **106**, 20526–20531.
22. Scholze,H. and Boch,J. (2010) TAL effector-DNA specificity. *Virulence*, **1**, 428–432.
23. An,W. and Chin,J.W. (2009) Synthesis of orthogonal transcription-translation networks. *Proc. Natl Acad. Sci. USA*, **106**, 8477–8482.
24. Barrett,O.P. and Chin,J.W. (2010) Evolved orthogonal ribosome purification for in vitro characterization. *Nucleic Acids Res.*, **38**, 2682–2691.
25. Lu,T.K., Khalil,A.S. and Collins,J.J. (2009) Next-generation synthetic gene networks. *Nat. Biotechnol.*, **27**, 1139–1150.
26. Wang,B., Kitney,R.I., Joly,N. and Buck,M. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun.*, **2**, 508.
27. Knight,T. (2003) Idempotent vector design for standard assembly of biobricks*DSpace. MIT Artificial Intelligence Laboratory;MIT Synthetic Biology Working Group.*
28. Phillips,I.S. and Pamela,A. (2006) A new biobrick assembly strategy designed for facile protein engineering*DSpace. MIT Artificial Intelligence Laboratory;MIT Synthetic Biology Working Group.*
29. Lohmueller,J.J., Armel,T.Z. and Silver,P.A. (2012) A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res.*, **40**, 5180–5187.
30. Pfaffl,M.W., Horgan,G.W. and Dempfle,L. (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.*, **30**, e36.
31. Lanctot,J.K.L., Li,M., Ma,B., Wang,S. and Zhang,L. (2003) Distinguishing string selection problems. *Inf. Comput.*, **185**, 41–55.
32. Xie,X., Lu,J., Kulbokas,E.J., Golub,T.R., Mootha,V., Lindblad-Toh,K., Lander,E.S. and Kellis,M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
33. Carninci,P., Sandelin,A., Lenhard,B., Katayama,S., Shimokawa,K., Ponjavic,J., Semple,C.A., Taylor,M.S., Engstrom,P.G., Frith,M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.*, **38**, 626–635.
34. Koudritsky,M. and Domany,E. (2008) Positional distribution of human transcription factor binding sites. *Nucleic Acids Res.*, **36**, 6795–6805.
35. MacIsaac,K.D., Lo,K.A., Gordon,W., Motola,S., Mazor,T. and Fraenkel,E. (2010) A quantitative model of transcriptional regulation reveals the influence of binding location on expression. *PLoS Comput. Biol.*, **6**, e1000773.
36. Geissler,R., Scholze,H., Hahn,S., Streubel,J., Bonas,U., Behrens,S.E. and Boch,J. (2011) Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One*, **6**, e19509.

37. 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.
38. Kay,S., Hahn,S., Marois,E., Wieduwild,R. and Bonas,U. (2009) Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Deltarep16. *Plant J.*, **59**, 859–871.
39. Romer,P., Strauss,T., Hahn,S., Scholze,H., Morbitzer,R., Grau,J., Bonas,U. and Lahaye,T. (2009) Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. *Plant Physiol.*, **150**, 1697–1712.
40. Stegmeier,F., Hu,G., Rickles,R.J., Hannon,G.J. and Elledge,S.J. (2005) A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc. Natl Acad. Sci. USA*, **102**, 13212–13217.
41. Leisner,M., Bleris,L., Lohmueller,J., Xie,Z. and Benenson,Y. (2010) Rationally designed logic integration of regulatory signals in mammalian cells. *Nat. Nanotechnol.*, **5**, 666–670.
42. Deans,T.L., Cantor,C.R. and Collins,J.J. (2007) A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell*, **130**, 363–372.
43. Mahfouz,M.M., Li,L., Piatek,M., Fang,X., Mansour,H., Bangarusamy,D.K. and Zhu,J.K. (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol. Biol.*, **78**, 311–321.
44. Blount,B.A., Weenink,T., Vasylechko,S. and Ellis,T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One*, **7**, e33279.
45. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
46. Bahar,R.I., Frohm,E.A., Gaona,C.M., Hachtel,G.D., Macii,E., Pardo,A. and Somenzi,F. (1993) Algebraic Decision diagrams and their applications. *Proceedings of the 1993 IEEE/ACM international Conference on Computer aided design*. Santa Clara, USA, pp. 188–191.
47. Bryant,R.E. (1986) Graph-Based Algorithms for Boolean Function Manipulation. *IEEE Transactions on Computers*, **35**, 677–691.
48. Somenzi,F. (2005) CUDD: CU Decision Diagram Package Release 2.4.1. University of Colorado at Boulder.