

Positive selection of DNA-protein interactions in mammalian cells through phenotypic coupling with retrovirus production

Ulrich Tschulena¹, Kenneth R Peterson², Beatriz Gonzalez¹, Halyna Fedosyuk² & Carlos F Barbas III¹

Through the shuffling of predefined modular zinc finger domains with predictable target site recognition *in vitro*, we have generated a large repertoire of artificial transcription factors with five zinc finger domains (TF_{ZF}s). Here we report an effective strategy for the selection of ATF libraries by coupling expression of transcriptional activators of the promoter of interest to the enhanced production of retroviral vector particles transferring the TF_{ZF} encoding gene. Using this strategy, we successfully selected specific TF_{ZF}s that upregulate the expression of the γ -globin promoter. Selected transcription factors induced the expression of γ -globin when coupled to an activation domain and reduced expression when linked to a repression domain. This new retroviral approach might be used to select other TF_{ZF}s but might also be generalized for the selection of other protein and small-molecule interactions.

Artificial transcription factors (ATFs) are proteins that are designed to specifically bind to DNA and modulate gene expression. Modular zinc finger DNA-binding domains allow for the assembly of zinc finger transcription factors (TF_{ZF}) with predictable target site recognition *in vitro*. Typically, a single zinc finger domain binds a 3-base pair (bp) DNA sequence through the formation of specific contacts that occur primarily within the major groove of the DNA. Our laboratory and others have successfully generated ATFs that target unique sites within complex genomes^{1–5}. ATFs have been designed and constructed for the regulation of various genes, such as those encoding ErbB2, ErbB3, VEGF, AP3 and EPO^{2,4,6–8}. Although direct design and synthesis of TF_{ZF}s has been successful in many cases, this rational design strategy is often limited by a lack of information about both the regulatory areas of the target gene and the endogenous factors affecting transcription factor–DNA interactions, for example, endogenous transcription factors, chromatin structure and DNA accessibility, that affect gene expression^{4,8}.

To overcome this limitation, we have generated large repertoires of zinc finger proteins through combination of the available zinc finger domains. When fused to a desired effector domain, these zinc finger libraries can be used as genome-wide screening tools for selection of novel functional TF_{ZF}s. Previously, we reported construction and selection of TF_{ZF} libraries based on the modulation of cell-surface

markers^{9,10}. Although effective, these screens were time-consuming and laborious: after each of the several rounds of selection, the selected library had to be recovered and recloned into an expression vector before the next screening cycle. Additionally, these selections were confined to the identification of activators of cell-surface markers or a reporter system suitable for fluorescence-activated cell sorting (FACS) methods.

In this study, we describe a powerful strategy for the selection of potent activators and repressors of gene expression. The key feature of this strategy is the coupling of the expression of transcriptional activators of the promoter of interest to the enhanced production and release of retroviral vector particles that specifically package and transfer the gene encoding the TF_{ZF} that caused enhanced expression. We have applied this strategy to select for modulators of γ -globin expression.

The human β -globin locus, located on chromosome 11, contains five homologous globin genes. In the early stages of fetal development, the embryonic (ϵ -) globin variant is predominantly expressed. During the later stages of fetal development, after the site of hematopoiesis has switched to the fetal liver, fetal γ -globin is the predominant globin form produced. Shortly after birth, there is yet another switch in expression to the adult β -globin genes within the bone marrow. Transcriptional control of these genes is mediated by a complex interplay between *cis*- and *trans*-acting regulatory elements^{11–15}. Upstream of these globin genes is a locus control region (LCR), which is necessary for regulation of the entire locus^{16,17}.

The genetic diseases β -thalassemia and sickle cell disease are associated with defective expression of β -globin or mutations within the β -globin product. These diseases can be mitigated by the expression of the fetal globin chain γ , which forms tetrameric $\alpha_2\gamma_2$ fetal hemoglobin (HbF). The hematologic benign condition known as ‘hereditary persistence of fetal hemoglobin’ (HPFH) is characterized by an increase in the level of fetal hemoglobin. In adults, the fetal globin level is normally 1–2% of the total hemoglobin; individuals with HPFH continue to express this gene in adult life, and expression may be as high as 20%. Patients with sickle cell disease and HPFH often do not have the serious life-threatening symptoms of those who have sickle cell anemia alone¹⁸. Consequently, chemical inducers of fetal hemoglobin such as hydroxyurea, butyrate and others have been identified and used for the treatment of sickle cell disease and

¹The Skaggs Institute for Chemical Biology and the Departments of Molecular Biology and Chemistry, The Scripps Research Institute, La Jolla, California, USA. ²The Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, Kansas, USA. Correspondence should be addressed to C.F.B. (carlos@scripps.edu).

Received 1 May; accepted 17 August; published online 18 October 2009; doi:10.1038/nsmb.1677

β -thalassemia¹⁹. Treatment with these compounds, however, produces a range of serious side effects, preventing long-term treatment. A possible gene-based therapeutic intervention for these patients is the induction of fetal hemoglobin expression by an artificial transcription factor^{20–22}. In our laboratory, we designed the artificial transcription factor gg1-VP64 and demonstrated that it increased the expression of endogenous γ -globin^{21,23}.

In this study, we selected for modulators of expression of the globin promoter from a library of TF_{ZF}s by coupling viral particle production with targeted promoter activation. We studied the coupling of a viral particle production phenotype with a defined phenotype of interest to explore the efficiency of this type of selection protocol. We then analyzed the selected transcription factors for their ability to modulate transcription of reporter constructs and cell lines that mimic endogenous human γ -globin gene expression profiles. Several of the selected transcription factors bound to the γ -globin promoter directly and upregulated human γ -globin gene expression in mouse β -YAC cells.

RESULTS

Coupling gene activation and retroviral particle production

We developed a library of zinc finger transcription factors that each recognizes a 15-bp target site. The 1.7×10^7 member library of proteins containing five zinc finger domains (the 5-ZF library) library provided a repertoire consisting of an equimolar mixture of a subset of defined zinc finger DNA sequences specific for binding to GNH (H represents A, C or T) and ANN subsites at each finger position. The defined pool excludes GNG domains, which in some cases have been implicated in target site overlap interactions that decrease zinc finger domain modularity. The resulting 5-ZF DNA-binding protein coding sequence was fused to the VP16-derived transactivation domain VP64 and cloned into the retroviral transfer vector pMX-5ZF-Lib-VP64, which expressed TF_{ZF} via the viral long terminal repeat (LTR) and green fluorescent protein (GFP) from an internal CMV promoter.

To select for activators of the γ -globin gene, we constructed a plasmid (pPur- γ -gp) in which murine leukemia virus (MLV) gag-pol expression is driven by the γ -globin promoter. This plasmid encodes the γ -globin promoter fragment sequence from –206 to +47, relative to the start of transcription. Also included in pPur- γ -gp were DNase I-hypersensitive sites 2 and 3 from the β -globin LCR. This portion of the LCR contains enhancer elements that are important for high-level activation of the entire β -globin locus²⁴. MLV gag-pol genes encode the major structural and enzymatic proteins necessary for formation of virions when expressed with an envelope gene. In this case, we used the VSV envelope to pseudotype viral particles, as it enables high infectivity. The virions package a viral transfer vector, in this case pMX-5ZF-Lib-VP64, thus introducing the TF_{ZF} gene into the genome of target cells (Fig. 1).

The key feature for this selection is the coupling of the expression of transcriptional activators of the γ -globin promoter to the enhanced production and release of retroviral vector particles that specifically package and transfer the gene encoding the TF_{ZF} responsible for activation. Cells that express γ -globin promoter activator show increased expression of MLV gag-pol genes, leading to increased viral particle production and quantitatively greater transduction of target cells.

Selection of transactivators of the γ -globin promoter

To begin the selection, we transfected pPur- γ -gp, pMX-5ZF-Lib-VP64 and a VSV-envelope (pMD.G) plasmid into human embryonic kidney (HEK) 293T cells. After 2 d, we collected the supernatant and used it to transduce fresh HEK 293T cells. After expansion, we again transfected these cells with pPur- γ -gp and VSV envelope (see Online

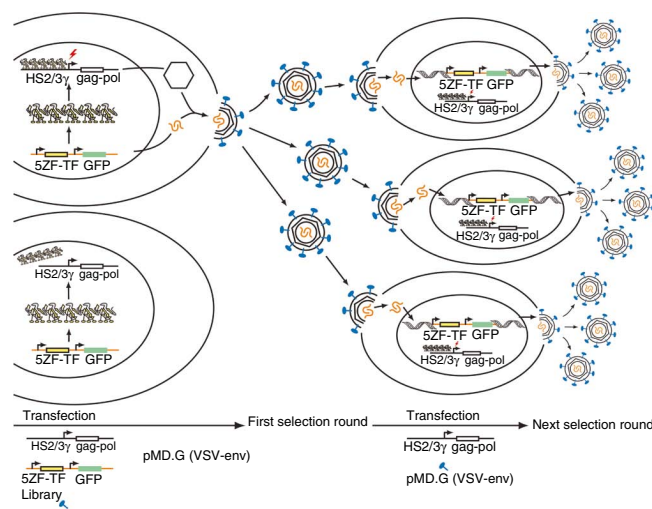


Figure 1 Schematic illustration of the selection strategy. For the selection of γ -globin promoter activators, cells were transfected with VSV-G envelope and a 5-ZF-VP64 library (5-ZF-TF) expressed by a retroviral transfer vector. Cells were co-transfected with the pPur- γ -gp plasmid encoding the MLV structural genes gag and pol under the control of γ -globin promoter and enhancer elements from the HS2 and HS3 regions of the globin promoter (HS2/3 γ). Cells that expressed an activator of the γ -globin promoter showed increased expression of MLV gag-pol genes, which lead to increased transduction efficiency of target cells. To start the next selection round, target cells were transfected directly with pPur- γ -gp and VSV-G envelope (VSV-env) for the production of new viral particles. Five rounds of selection were carried out.

Methods). For the next selection round, we used the supernatants to transduce fresh HEK 293T cells. After harvesting of the supernatant following that step, we collected the producer cells and analyzed their genomic DNA (Fig. 1). We PCR amplified the retrovirally integrated zinc finger pools and recloned them into the pMX-VP64 vector, and carried out five rounds of selection. We then transfected the resulting pMX-5ZF-VP64 library constructs into HEK 293T cells to enable characterization of the selected zinc finger variants using a transduction-efficiency test. The number of GFP-expressing cells was analyzed by flow cytometry. Included in each experiment was a positive control, pMX-gg1-VP64-CMV-GFP, which encoded a designed zinc finger protein previously shown to strongly activate γ -globin expression²¹.

With each selection round, the number of GFP-expressing cells increased. After the fifth selection round, 79% of cells expressed GFP (Fig. 2). This was a nearly nine-fold increase compared to cells transfected with the starting library (9% GFP-expressing cells) or the non-DNA-binding control zinc finger (Ala-VP64, 10% GFP-expressing cells). When transfected with the positive control, pMX-gg1-VP64, 75% of cells expressed GFP (Fig. 2).

To further analyze the ability of the selected zinc finger variants to activate the γ -globin promoter, we tested their ability to upregulate luciferase expression in the γ β -luc construct. Similar to the pPur- γ -gp vector, γ β -luc encodes the γ -globin promoter fragment from –206 to +47, relative to the start of transcription, and also includes the DNase I-hypersensitive sites 2 and 3 from the β -globin LCR²¹. This plasmid also encodes a β -globin promoter fragment consisting of the sequence from –206 to +47 relative to start of transcription, which drives expression of *Renilla* luciferase. By this method, specific activation of the γ -globin promoter is determined by dividing the firefly luciferase values with *Renilla* luciferase values. We co-transfected HeLa cells with the pMX-VP64 plasmids, expressing the selected transcriptional

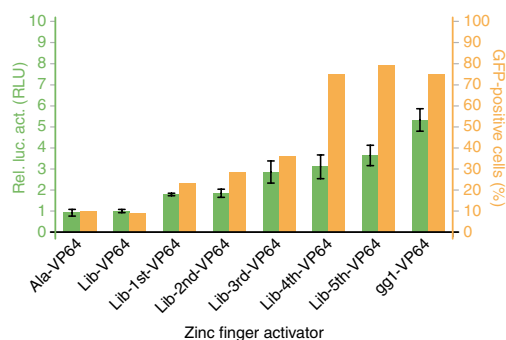


Figure 2 Selected TF_{ZF} libraries show increased GFP-positive transduced cells after each round of selection. Retroviral particles were produced in HEK 293T cells co-transfected with pPur- γ -gp and selected TF_{ZF} libraries (denoted Lib-1st to Lib-5th) were used for transduction of fresh HEK 293T target cells. Those cells were then compared to cells transduced with the non-DNA-binding control zinc finger (Ala-VP64), with unselected pMX-5ZF-library (Lib) or the positive control gg1-VP64. After 3 d of transduction, cells were analyzed by FACS to assess GFP expression. Percentage of GFP-positive cells is depicted in yellow (right scale). Selected TF_{ZF} libraries show increased transcriptional activation on the γ -globin promoter after each selection round. HeLa cells were transfected with control zinc fingers (Ala-VP64 and gg1-VP64) and TF_{ZF} libraries (denoted Lib-1st to Lib-5th) together with a luciferase reporter plasmid expressing firefly luciferase under the control of the γ -globin promoter and *Renilla* luciferase under the control of the β -globin promoter. Specific activation of the γ -globin promoter was calculated by normalizing to *Renilla* luciferase values. Normalized luciferase activity of the selected TF_{ZF} libraries was compared to that of unselected pMX-5ZF-library (Lib = 1) and the positive control gg1-VP64. The normalized luciferase activities (relative luciferase activity (Rel. luc. act.) in relative luciferase units (RLU)) are shown in green (\pm s.d.) (left scale). For Lib-VP64 versus Lib-1st-VP64, $P = 0.00015$; for Lib-VP64 versus Lib-5th-VP64, $P = 0.00073$ (Student's *t*-test).

activators, and γ β -luc. With each selection round, luciferase activity increased. The pool of the selected variants after the fifth selection round showed a 3.6-fold increase in luciferase activity compared to cells transfected with the starting library or with the non-DNA-binding control Ala-VP64 (Fig. 2). This was slightly less expression than we observed in cells transfected with the positive control pMX-gg1-VP64, which produced a 5.3-fold increase relative to untransfected cells. Thus, we were able to select for zinc fingers that strongly upregulated expression from the γ -globin promoter and enhancer region.

Characterization of selected zinc-finger transactivators

We analyzed further a total of 15 individual clones from the fifth round of selection. Transfection with seven of the individual clones resulted in more than double the number of GFP-positive HEK 293T cells when compared to transfection with the non-DNA-binding control Ala-ZF. Clones # γ A-VP64 and # γ L-VP64 increased the number of GFP-positive cells 7.3-fold and 7.8-fold, respectively, a level comparable to that of the positive control pMX-gg1-VP64 (Supplementary Table 1).

Next, we investigated whether the selected TF_{ZF}s were able to modulate expression from the γ -globin promoter in the luciferase reporter system. The strongest activators in this assay, # γ A-VP64, # γ K-VP64, # γ L-VP64 and # γ M-VP64, showed between 2.9-fold and 3.9-fold increase in luciferase activity when compared to the non-DNA-binding control Ala-ZF-VP64; the positive control pMX-gg1-VP64 increased expression 5.8-fold (Supplementary Table 1). These data demonstrate that our selection system can be used to select clones that strongly upregulate the γ -globin promoter.

Binding of selected zinc finger to γ -globin promoter target site

We analyzed the γ -globin promoter and the DNase I-hypersensitive sites 2 and 3 from the β -globin LCR as well as the plasmid backbone of the pPur- γ -gp vector using the Zinc Finger Tools software (<http://www.zincfingertools.org>) to identify putative DNA-binding sites. For six of the zinc finger proteins that strongly activated expression from the γ -globin promoter, we could find no binding site with less than four mismatches in the γ -globin promoter and the DNase I-hypersensitive sites 2 and 3 or the pPur- γ -gp vector backbone.

A target site search using the zinc finger domains of # γ A identified the site 5'-GAT GCC GTT TGA GGT-3' on the reverse strand, 301 nucleotides (nt) upstream of the site of transcription initiation; this site contains a 3-nt mismatch. To determine the binding affinity of # γ A to this target site, we performed an electrophoretic mobility shift assay (EMSA) with a purified # γ A-maltose-binding fusion protein (MBP). The calculated K_d value of # γ A with this DNA target was 3.2 nM (data not shown).

Because we found that the predicted 15-bp target site of the # γ A protein was not the optimal site for this zinc finger, we optimized the 5-ZF protein for specific binding to this particular site using rational design. We replaced three of the five zinc finger domains with domains that we determined to have better specificity for this sequence. The modified protein, # γ A-opt, was composed of the following α -helical zinc finger domains: F1-GGT (TSGHLVR), F2-TGA (QAGHLAS), F3-GTT (TSGSLVR), F4-GCC (DCRDLAR) and F5-GAT (TSGNLVR). This represents a zinc finger clone that had not been present in the starting library, which did not contain zinc finger DNA sequences specific for binding TNN subsites. The K_d of purified MBP-# γ A-opt for the target site DNA was 7.1 nM (data not shown). We had expected a slight reduction in binding affinity because the incorporated domains were chosen on the basis of specificity rather than affinity. In transduction efficiency and luciferase activity assays, # γ A-opt-VP64 induced an increase in activity similar to that induced by the parental # γ A-VP64 construct (Supplementary Table 1).

Downregulation of γ -globin promoter

To analyze the variants' ability to also repress expression of the γ -globin promoter, we subcloned the four selected zinc finger clones (# γ A, # γ L, # γ K and # γ M) showing the strongest activation in the reporter assays, as well as gg1, Ala and # γ A-opt, into the pMX retroviral vector as fusions to the transcriptional repressor Krüppel-associated box (KRAB). We tested the resulting constructs (# γ A-KRAB, # γ L-KRAB, # γ A-KRAB, # γ A-opt-KRAB, # γ L-KRAB, # γ K-KRAB and # γ M-KRAB gg1-KRAB) for repression in luciferase assays. # γ L-KRAB, # γ K-KRAB and # γ M-KRAB did not repress expression from the γ -globin promoter when compared to the non-DNA-binding Ala-ZF-KRAB. In contrast, # γ A, which may bind directly to a sequence in the γ -globin promoter, showed a 55% reduction in luciferase activity, similar to that of the positive control pMX-gg1-KRAB (Fig. 3). Notably, the optimized TF_{ZF} # γ A-opt-KRAB showed an even stronger reduction, with 64% decreased luciferase activity when compared to Ala-ZF-KRAB (Fig. 3). Thus, these data show that exchange of the transactivator domain VP64 with the transcriptional repressor domain KRAB produced repressors of the γ -globin promoter, indicating direct regulation through binding of the target sequence.

Analysis of human γ -globin gene expression in mouse β -YAC cells

Encouraged by the high level of promoter activation in the reporter assays, we investigated the regulation of the endogenous γ -globin gene by the selected ATFs. We fused full-length cDNAs of # γ A, # γ L, # γ K and # γ M and control zinc fingers to the activation domain VP64 under

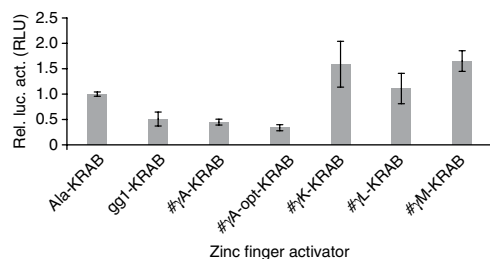


Figure 3 Selected TF_{ZF}s shows repression of γ -globin promoter expression when fused to the repressor domain KRAB. HeLa cells were transfected with TF_{ZF}s fused to KRAB (# γ A-KRAB, # γ A-opt-KRAB, # γ L-KRAB, # γ K-KRAB and # γ M-KRAB gg1-KRAB), a luciferase reporter plasmid expressing firefly luciferase under the control of the γ -globin promoter and *Renilla* luciferase under the control of the β -globin promoter. Specific activation of the γ -globin promoter was calculated by normalizing to *Renilla* luciferase values (relative luciferase activity (Rel. luc. act.) in relative luciferase units (RLU)). Error bars represent s.d. Normalized luciferase activity of the selected TF_{ZF}s was compared to that of cells transfected with the non-DNA-binding negative control Ala-KRAB and the positive control gg1-KRAB. For Ala-KRAB versus # γ A-KRAB, $P = 0.00018$; for Ala-KRAB versus # γ A-opt-KRAB, $P = 9.6 \times 10^{-5}$ (Student's *t*-test).

the control of a CMV promoter (in a pcDNA3.1-hygro context). We transfected these constructs into established immortalized cells that were derived from the bone marrow of β -YAC transgenic mice expressing exclusively β -globin from the YAC; YAC-derived γ -globin is not detectable either in these mice or in the derivative cells²³. In these cells, human γ -globin expression can be reactivated by various treatments, and the endogenous mouse globin genes maintain their native expression profiles.

We generated pools of β -YAC cells stably expressing the TF_{ZF}s and used quantitative RT-PCR to assess γ -globin expression. Levels of TF_{ZF} transcription ranged from approximately 0.4 to 1.2 relative to GAPDH gene expression, as measured by RT-PCR (data not shown). We observed γ -globin expression in each of these cell lines. Human β -globin expression was not increased in cells expressing any of the TF_{ZF}s (data not shown); in fact, in some instances it decreased reciprocally with the increase in γ -globin, as expected when γ -globin and β -globin genes compete for interaction with the LCR. The total β -like globin chain balance was maintained in these cells, and mouse α -globin levels also were unchanged (data not shown). In cells stably expressing # γ M-VP64, γ -globin was expressed at a level similar to that in cells expressing the positive control gg1-VP64 (Fig. 4). These data indicate that all 4 of the 15 TF_{ZF}s analyzed reactivated transcription of the endogenous γ -globin gene.

DISCUSSION

We report a new strategy for selection of zinc finger transcription factor regulators of the promoter and enhancer region of γ -globin. This system couples activation of a promoter to the increased production of retroviral particles that encode the regulatory zinc finger protein. After five rounds of selection, we successfully selected specific modulators of γ -globin expression from a library of 1.7×10^7 TF_{ZF}s. The TF_{ZF}s that were selected after the fifth round of selection showed activation of the γ -globin promoter comparable to that of the rationally designed activator of the γ -globin promoter gg1-VP64, as measured by the transduction-efficiency test. Thus, the number of selection rounds performed seemed sufficient for the selection carried out in this study. However, more rounds of selection might lead to an increase in selection efficiency. Many of the selected clones from the 5-ZF library induced substantially more luciferase than controls. Some of the selected clones, when fused to the activation domain VP64, showed levels of activation similar to that of gg1-VP64. Additionally,

the selected TF_{ZF}s were able to substantially induce human γ -globin expression in adult bone marrow cells derived from β -YAC transgenic mice. These experiments suggest that selected transcription factors, should they ultimately prove specific, could be used to treat patients suffering from sickle cell or thalassemic diseases. The efficacy of zinc finger-based therapeutics agents has been demonstrated in mouse models, and human studies are ongoing^{25,26}.

Previously, we selected TF_{ZF} libraries for transcriptional activators of different cell-surface markers^{9,10}. The selection method we used was time-consuming and laborious, as after each round of selection the selected library had to be recovered from genomic DNA and then recloned into an expression vector; each selection round took 2–3 weeks. The selection system presented here eliminates the need for a laborious cloning process after each selection round, reducing the time needed for each round to less than 1 week. As transduced cells are used directly as producer cells for the next selection round, cloning and PCR steps that introduce mutations are avoided. Moreover, this method is not limited to selection of activators of cell-surface markers and, thus, should be generally useful. The promoter region of any transcript of interest, including promoters of noncoding RNAs, such as microRNAs, might in principle be used in the selection. This method differs significantly from the OPEN bacterial selection approach²⁷, which has been shown to work for the selection of three-finger zinc finger proteins that bind defined 9-bp target sites, as well as our phage-based approach²⁸, which can be used to select proteins against larger DNA fragments. In this new approach, selections are performed in a mammalian cell background, taking into account chromatin structure and the potential to select for indirect activators that act elsewhere in the mammalian genome.

One of the strongest activators of the γ -globin promoter, the five-finger zinc finger # γ A, interacted directly with a fragment of the γ -globin promoter encompassing positions 301–315 relative to transcriptional start site. Despite a predicted 3-bp mismatch, the selected zinc finger bound to the promoter-derived oligonucleotide with an affinity of 3.1 nM, slightly weaker than that of the rationally designed six-finger TF_{ZF} gg1, which bound with a K_d of 0.7 nM to its target site in the γ -globin promoter²¹. The five-finger library used here excluded

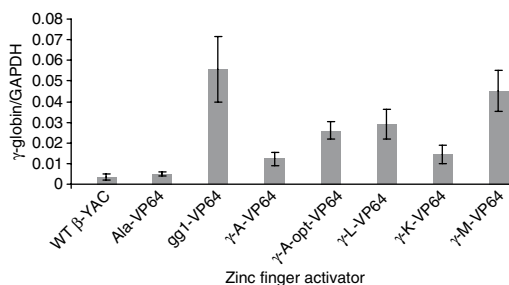


Figure 4 Activation of human γ -globin in β -YAC bone marrow cells (BMCs) containing selected zinc fingers, as measured by RT-PCR. β -YAC BMCs were transfected with TF_{ZF}s (Ala-VP64, # γ A-VP64, # γ A-opt-VP64, # γ L-VP64, # γ K-VP64 and # γ M-VP64 gg1-VP64) expressed in pcDNA-hygro. TF_{ZF}-expressing cells were selected for hygromycin resistance. RNA was isolated and γ -globin expression was measured by RT-PCR normalized to GAPDH expression. Averages of duplicates or triplicates from at least two independent experiments for each sample are shown. Error bars represent s.d. For Ala-VP64 versus # γ A-opt-VP64, $P = 0.0012$ (Student's *t*-test).

CNN domains, preventing reselection of gg1 target site-binding TF_{ZF}s. When the #yA zinc finger was linked to the repressor domain KRAB, luciferase expression from the γ -globin was repressed. Thus, the method described here can also be used for the selection of activators or repressors of a promoter of interest. In addition to identifying TF_{ZF}s that modulate gene expression, this approach might be used to identify zinc finger proteins that target the activities of zinc finger enzymes such as methylases, nucleases and recombinases^{23,29–34}.

For many of the selected clones we were unable to define a definitive binding site on the γ -globin promoter fragment used in the selection. It is possible that these clones recognize degenerate sequences in the promoter region or regulate a gene elsewhere in the genome of the cell that ultimately affects γ -globin expression. Using reliable parameters for prediction of the zinc finger target sites, the strongest activator in luciferase reporter assays, #yL, was not predicted to bind to the γ -globin promoter. This TF_{ZF} might regulate a gene involved in γ -globin expression or regulate a gene involved in post-transcriptional modification, and therefore this approach might be applicable to the identification of regulatory genes of a defined promoter. We identified TF_{ZF}s that were direct regulators of the γ -globin promoter and showed that they induce expression of the endogenous γ -globin gene in immortalized cells derived from the bone marrow of β -YAC transgenic mice, which otherwise exclusively express β -globin. Thus, these TF_{ZF}s are potential candidates for the development of TF_{ZF} therapy of sickle cell patients.

In conclusion, we have developed a new and efficient method to identify functional zinc finger proteins that regulate a targeted promoter by screening large combinatorial libraries of TF_{ZF}s. The method couples the increased expression from a promoter of interest to the enhanced production and release of retroviral vector particles that specifically package and transfer the gene responsible for upregulation. This method should be amenable to adaptation to other types of selections. For example, selection of receptor activators that lead to induction of a specific promoter could be coupled to virus production and could therefore be used for the identification of proteins other than transcription factors that function within the context of particular mammalian cell type.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

Thanks to F.C. Costa and R. Neades for technical assistance with the BMC transfections and expression assays. This study was supported by the Skaggs Institute for Chemical Biology and in part by US National Institutes of Health Grants RO1 DK61803 and RO1GM065059 (C.F.B.) and RO1 DK061804 (K.R.P.).

AUTHOR CONTRIBUTIONS

U.T., K.R.P. and C.F.B. designed the research; U.T., K.R.P., B.G. and H.F. performed the experiments; U.T., K.R.P. and C.F.B., wrote the manuscript, which all authors commented on.

Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Beerli, R.R. & Barbas, C.F. III. Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* **20**, 135–141 (2002).
- Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. & Barbas, C.F. III. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **276**, 29466–29478 (2001).
- Greisman, H.A. & Pabo, C.O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657–661 (1997).
- Liu, P.Q. *et al.* Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J. Biol. Chem.* **276**, 11323–11334 (2001).
- Segal, D.J., Dreier, B., Beerli, R.R. & Barbas, C.F. III. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA* **96**, 2758–2763 (1999).
- Beerli, R.R., Dreier, B. & Barbas, C.F. III. Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl. Acad. Sci. USA* **97**, 1495–1500 (2000).
- Stege, J.T., Guan, X., Ho, T., Beachy, R.N. & Barbas, C.F. III. Controlling gene expression in plants using synthetic zinc finger transcription factors. *Plant J.* **32**, 1077–1086 (2002).
- Zhang, L. *et al.* Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *J. Biol. Chem.* **275**, 33850–33860 (2000).
- Blancafort, P., Magnenat, L. & Barbas, C.F. III. Scanning the human genome with combinatorial transcription factor libraries. *Nat. Biotechnol.* **21**, 269–274 (2003).
- Magnenat, L., Blancafort, P. & Barbas, C.F. III. *In vivo* selection of combinatorial libraries and designed affinity maturation of polydactyl zinc finger transcription factors for ICAM-1 provides new insights into gene regulation. *J. Mol. Biol.* **341**, 635–649 (2004).
- Chada, K., Magram, J. & Costantini, F. An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature* **319**, 685–689 (1986).
- Chada, K. *et al.* Specific expression of a foreign γ -globin gene in erythroid cells of transgenic mice. *Nature* **314**, 377–380 (1985).
- Kollias, G., Wrighton, N., Hurst, J. & Grosfeld, F. Regulated expression of human γ , β , and hybrid $\gamma\beta$ -globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell* **46**, 89–94 (1986).
- Peterson, K.R. Hemoglobin switching: new insights. *Curr. Opin. Hematol.* **10**, 123–129 (2003).
- Rutherford, T. & Nienhuis, A.W. Human globin gene promoter sequences are sufficient for specific expression of a hybrid gene transfected into tissue culture cells. *Mol. Cell. Biol.* **7**, 398–402 (1987).
- Li, Q., Peterson, K.R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077–3086 (2002).
- Wijgerde, M., Grosfeld, F. & Fraser, P. Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377**, 209–213 (1995).
- Murray, N., Serjeant, B.E. & Serjeant, G.R. Sickle cell-hereditary persistence of fetal haemoglobin and its differentiation from other sickle cell syndromes. *Br. J. Haematol.* **69**, 89–92 (1988).
- Swank, R.A. & Stamatoyannopoulos, G. Fetal gene reactivation. *Curr. Opin. Genet. Dev.* **8**, 366–370 (1998).
- Blouin, M.J. *et al.* Genetic correction of sickle cell disease: insights using transgenic mouse models. *Nat. Med.* **6**, 177–182 (2000).
- Gräslund, T., Li, X., Magnenat, L., Popkov, M. & Barbas, C.F. III. Exploring strategies for the design of artificial transcription factors: targeting sites proximal to known regulatory regions for the induction of γ -globin expression and the treatment of sickle cell disease. *J. Biol. Chem.* **280**, 3707–3714 (2005).
- May, C. *et al.* Therapeutic haemoglobin synthesis in β -thalassaemic mice expressing lentivirus-encoded human β -globin. *Nature* **406**, 82–86 (2000).
- Blau, C.A. *et al.* γ -Globin gene expression in chemical inducer of dimerization (CID)-dependent multipotential cells established from human β -globin locus yeast artificial chromosome (β -YAC) transgenic mice. *J. Biol. Chem.* **280**, 36642–36647 (2005).
- Collis, P., Antoniou, M. & Grosfeld, F. Definition of the minimal requirements within the human β -globin gene and the dominant control region for high level expression. *EMBO J.* **9**, 233–240 (1990).
- Powars, D.R., Chan, L. & Schroeder, W.A. The influence of fetal hemoglobin on the clinical expression of sickle cell anemia. *Ann. NY Acad. Sci.* **565**, 262–278 (1989).
- Rebar, E.J. *et al.* Induction of angiogenesis in a mouse model using engineered transcription factors. *Nat. Med.* **8**, 1427–1432 (2002).
- Maeder, M.L. *et al.* Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31**, 294–301 (2008).
- Lund, C.V., Blancafort, P., Popkov, M. & Barbas, C.F. III. Promoter-targeted phage display selections with preassembled synthetic zinc finger libraries for endogenous gene regulation. *J. Mol. Biol.* **340**, 599–613 (2004).
- Gordley, R.M., Smith, J.D., Gräslund, T. & Barbas, C.F. III. Evolution of programmable zinc finger-recombinases with activity in human cells. *J. Mol. Biol.* **367**, 802–813 (2007).
- Miller, J.C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
- Nomura, W. & Barbas, C.F. III. *In vivo* site-specific DNA methylation with a designed sequence-enabled DNA methylase. *J. Am. Chem. Soc.* **129**, 8676–8677 (2007).
- Szczepiek, M. *et al.* Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786–793 (2007).
- Urnov, F.D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
- Xu, G.L. & Bestor, T.H. Cytosine methylation targeted to pre-determined sequences. *Nat. Genet.* **17**, 376–378 (1997).

ONLINE METHODS

Plasmids. The $\gamma\beta$ -luc plasmid used for luciferase reporter assays has been described²¹. Each member of the 5-ZF library of artificial transcription factors recognizes a 15-bp DNA site. We constructed the 5-ZF library by combinatorial assembly of five different zinc finger repertoires, connected by the canonical TGEKP linker. Each repertoire consisted of an equimolar mixture of a subset of defined zinc finger DNA sequences that were previously selected *in vitro* by phage display for specific binding to each of the possible GNH/ANN subsites^{2,5}. We constructed the library as described (B.G., L.J. Schwimmer, R.P. Fuller, Y. Ye, L. Asawapornmongkol and C.F.B., unpublished data). We fused the resulting 5-ZF DNA-binding protein-coding sequences to the VP16-derived transactivation domain VP64 and cloned the 5-ZF-VP64 library into the retroviral vector pMX-5ZF-Lib-VP64 by SfiI digest, as described^{9,10}. The pMX-5ZF-Lib-VP64 plasmid expressed TF_{ZF} via the viral LTR and GFP from an internal CMV promoter. To fuse the 5-ZF DNA-binding domains to the KRAB repressor domain, we excised the zinc fingers from pMX-5ZF-VP64 by SfiI digest and inserted them into a corresponding pMX-KRAB^{9,10}.

The construction of pPur- γ -gp is described in **Supplementary Methods**. Briefly, this plasmid expresses MLV gag and pol genes under the control of the γ -globin promoter and enhancer region.

The construction of the modified protein # γ A-opt is described in detail in **Supplementary Methods**. The # γ A-opt zinc finger was composed of the following α -helical zinc finger domains: F1-GGT (TSGHLVR), F2-TGA (QAGHLAS), F3-GTT (TSGSLVR), F4-GCC (DCRDLAR) and F5-GAT (TSGNLVR).

Cell lines. The HEK 293T cells (CRL-11268) and HeLa cells (ATCC: CCL-2) were obtained from ATCC. CID-dependent β -YAC BMCs were maintained and transfected with various plasmid DNAs as described²³.

Selection of activators of γ -globin regulatory elements. We co-transfected HEK 293T cells (3.5×10^6) with pMD.G plasmid encoding the vesicular stomatitis virus-G envelope protein, pMX-5ZF-VP64 library DNA and pPur- γ -gp plasmid. We then incubated cell-free and concentrated virus with HEK 293T target cells in presence of $8 \mu\text{g ml}^{-1}$ polybrene. Cells were expanded and the next selection cycle was initiated by co-transfecting them with pMD.G and pPur- γ -gp plasmid, as described above. After transduction of the target cells, the DNA encoding the pool of TF_{ZF}s was recovered by PCR from the producer cells using the primers pMXf2 (forward) 5'-TCAAAGTAGACGGCATCG-3' and VP64AscB (reverse) 5'-TCGTCCAGCGCGCTCGGCGCG-3' and cloned into the pMX vector. Selections were repeated for five rounds.

Luciferase assays. For each sample, we co-transfected pMX-5ZF-VP64 or pMX-5ZF-KRAB with reporter plasmid $\gamma\beta$ -luc. We analyzed the cell lysates for firefly and *Renilla* luciferase expression using the dual luciferase reporter assay system (Promega). We normalized the firefly luciferase values to the *Renilla* luciferase activity to determine the specific activation of the γ -globin promoter. Data shown represent the average of three experiments. Ala-ZF (which does not bind to DNA) was used as a negative control and gg1-ZF (which has a target site in the γ -globin promoter, 5'-GTC AAG GCA AGG CTG GCC-3') was used as a positive control.

Transduction-efficiency test. The transduction efficiency test was similar to the first round of selection. We co-transfected HEK 293T cells with pMD.G plasmid, pMX-5ZF-VP64 DNA and pPur- γ -gp plasmid. We collected the

product retroviral particles and incubated them with HEK 293T target cells in presence of $8 \mu\text{g ml}^{-1}$ Polybrene (Sigma). After 3 d of transduction, we analyzed the cells for GFP expression by flow cytometry.

Cloning, expression, purification and characterization of DNA binding of the selected zinc fingers. We digested all of the assembled five-finger coding regions with the restriction endonuclease SfiI and cloned them into pMal-CSS, a derivative of the bacterial expression vector pMal-C2 (New England Biolabs). We purified the fusion proteins to >90% homogeneity by using the Protein Fusion and Purification System (New England Biolabs) as per the manufacturer's instructions, except that we used zinc buffer A (ZBA: 10 mM Tris-HCl, pH 7.5, 90 mM KCl, 1 mM MgCl₂, 90 μM ZnCl₂, 5 mM DTT) as the column buffer. Target oligonucleotides were labeled at their 5' or 3' ends with [³²P]ATP and gel purified. We incubated 11 three-fold serial dilutions of protein (starting with 1.6 μM) in 20 μl of binding reaction solution (ZBA with 1% w/w BSA), 10% (v/v) glycerol, 1 pM target oligonucleotide) for 3 h at 23 °C, then resolved them on a 5% polyacrylamide gel in 0.5 \times TBE buffer (90 mM Tris, 64.6 mM boric acid, 2.5 mM EDTA, pH 8.3). Quantitation of dried gels was performed using a PhosphorImager and IMAGEQUANT software (Molecular Dynamics) and the K_d values were determined by Scatchard analysis.

Theoretical analysis of selected zinc fingers. Theoretical binding analysis of the selected zinc fingers to the γ -globin locus was performed using the Zinc Finger Tools software (<http://www.zincfingertools.org>)³⁵.

Analysis of endogenous γ -globin gene expression. We isolated total RNA from CID-dependent β -YAC BMCs, either stably transfected with linear fragments purified from plasmid constructs or not, using the Promega RNagents Kit according to the manufacturer's instructions. For reverse transcription, we used 1 μg total RNA to synthesize cDNA with SuperScript II Reverse Transcriptase (Invitrogen). We analyzed the samples by quantitative real-time PCR with iQ SYBR Green Supermix (BioRad). PCR conditions for both γ -globin and GAPDH primers were as follows: 95 °C for 3 min, followed by 40 cycles of 95 °C, 10 s and 60 °C, 25 s. The melting curve was obtained from measurements taken every 0.5 s from 60–90 °C. A series of dilutions of samples from –117 Greek HPHF β -YAC BMCs was used to generate a standard curve. Untransfected wild-type β -YAC BMC samples were included as controls in all runs. Primer sequences were Mouse GAPDH forward, 5'-AGGTTGTCTCCTGCGACTT CA-3'; GAPDH reverse, 5'-CCAGGAAATGAGCTTGACAAAG-3'; human γ -globin forward, 5'-GTATTGCTTGAGATAAAGCC-3'; and γ -globin reverse, 5'-ACCGTTTTGGCAATCCATTC-3'. The data shown are the averages of duplicates or triplicates from at least two independent experiments for each sample. Error bars represent s.d. from the mean. TF_{ZF} transcription was measured by RT-PCR and compared to GAPDH transcription within the same cell pools using a pair of primers that universally amplified TF_{ZF} transcripts, regardless of the zinc finger sequence present. Primer sequences were TF_{ZF} forward, 5'-ATGGCCCAGGCGCCCTCG-3'; and TF_{ZF} reverse, 5'-AAAGTCATCGAGGGCATCAG-3'.

35. Mandell, J.G. & Barbas, C.F. III. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.* **34**, W516–W523 (2006).