Two- and three-input TALE-based AND logic computation in embryonic stem cells

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1093/nar/gkt758</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:11878975">http://nrs.harvard.edu/urn-3:HUL.InstRepos:11878975</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Two- and three-input TALE-based AND logic computation in embryonic stem cells

Florian Lienert¹, Joseph P. Torella¹, Jan-Hung Chen¹, Michael Norsworthy¹, Ryan R. Richardson¹,² and Pamela A. Silver¹,³,*

¹Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA, ²Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA and ³Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

Received April 26, 2013; Revised July 30, 2013; Accepted July 31, 2013

ABSTRACT

Biological computing circuits can enhance our ability to control cellular functions and have potential applications in tissue engineering and medical treatments. Transcriptional activator-like effectors (TALEs) represent attractive components of synthetic gene regulatory circuits, as they can be designed de novo to target a given DNA sequence. We here demonstrate that TALEs can perform Boolean logic computation in mammalian cells. Using a split-intein protein-splicing strategy, we show that a functional TALE can be reconstituted from two inactive parts, thus generating two-input AND logic computation. We further demonstrate three-piece intein splicing in mammalian cells and use it to perform three-input AND computation. Using methods for random as well as targeted insertion of these relatively large genetic circuits, we show that TALE-based logic circuits are functional when integrated into the genome of mouse embryonic stem cells. Comparing construct variants in the same genomic context, we modulated the strength of the TALE-responsive promoter to improve the output of these circuits. Our work establishes TALEs as a tool for building logic computation with the potential of controlling expression of endogenous genes or transgenes in response to a combination of cellular signals.

INTRODUCTION

A major focus of synthetic biology is the design of genetic circuits that program cellular functions in living organisms. The development of such circuits in mammalian cells has the potential to lead to new strategies for cell-based therapies and diagnostics (1–3). Bio-molecular computing systems have been implemented using various components, including recombinases, small RNAs, riboswitches and natural and artificial transcription factors (TFs) (4). Whereas the first synthetic transcriptional networks in eukaryotes relied on well-characterized bacterial TF–promoter pairs, recent work has used zinc finger TFs (5,6). Zinc finger proteins offer the advantage that their DNA binding domain can be designed to recognize specific sites, but the predictability of this engineering process can sometimes be a challenge (7).

Transcriptional activator-like effectors (TALEs) are an alternative class of transcription activators whose DNA binding specificity is more amenable to engineering (8–10). DNA recognition by TALEs is mediated by a protein domain consisting of a variable number of linearly arranged TAL repeats that are, on average, 34 amino acids in length. Two variable amino acids within each TAL repeat determine the specificity towards a single nucleotide in the TALE DNA recognition site. When combined with transcriptional activator or repressor domains, TALEs are able to regulate expression of endogenous genes and transgenes (11–15). Furthermore, Li et al. (16) have recently developed TALE hybrids that can be regulated by either addition of exogenous ligands or by endogenous pathways, such as hypoxia signaling or microRNAs.

Given their potential as regulators of gene expression and as parts of synthetic transcriptional networks, we tested whether TALEs can be used to perform AND logic computation in mammalian cells. For this purpose, we used an intein-mediated protein splicing approach. Specifically, we made use of split inteins, which can auto-catalytically trans-splice protein fragments to which they are fused (17,18). Using this system, we implemented TALE-based computation in mammalian cells.

For practical applications of bio-computational circuits in cell therapy, it is necessary to stably maintain them in the cell type of interest. Currently, components of newly developed mammalian synthetic transcriptional networks are often expressed from separate plasmids; genomic integration of such circuits therefore necessitates the use of
multiple selection markers, and leads to differences in the copy number of individual components (19,20). To circumvent these limitations, we assembled the TALE-based logic circuits as single DNA constructs and show that they maintain their functionality. We also demonstrate that they can perform AND computation when integrated into the genome of mouse embryonic stem (ES) cells. Using a site-directed insertion approach further allowed us to optimize circuit variants in the same genomic environment. Our finding that genomically integrated TALE-based circuits are functional in pluripotent ES cells suggests that they could be used as sensors in cell therapy applications or for directing \textit{in vitro} differentiation in tissue engineering.

\textbf{MATERIALS AND METHODS}

\textbf{Recombinant DNA constructs}

From a list of 20-bp-long potential TALE binding sites that are orthogonal to gene promoters in the human genome (13), we selected one that is also orthogonal to mouse gene promoters. The corresponding TALE, TAL118, was assembled using a Golden-Gate cloning scheme (13). The split intein fragments, which have been described previously (5), were fused to split TALE fragments using polymerase chain reaction and Alw26I Type-IIS restriction enzyme methods. Expression constructs were generated by combining BioBrick subparts using Biobrick assembly (21,22). For testing circuits with each part on a separate plasmid, DNA fragments were cloned between NheI and NotI sites of vector pCDNA5\textit{ins} (5) and reporter fragments between SpeI and NotI sites of pCDNA5\textit{FRT/TO} for mammalian expression (Invitrogen). For assembly of circuits on a single DNA construct, an isothermal assembly-based hierarchal cloning scheme was used (Torella \textit{et al.}, submitted). Circuits were assembled on two modified versions of vector pETcoco-1 (Novagen); pDestRmceBAC, which includes two inverted loxP sites around the assembled circuit, and pDestPBBAC, which contains a Blasticidin resistance marker, two inverted terminal repeats and a PiggyBac transposase (from vector pHULK, DNA2.0). Sequence parts are listed in Supplementary Table S1.

\textbf{Cell culture}

The human osteosarcoma-derived epithelial cell line U-2 OS (ATCC no. HTB-96) was maintained and transfected as previously described (5). TC-1 ES cells (background 129S6/SvEvTac) containing a recombinase-mediated cassette exchange (RMCE) target site were cultured as 129S6/SvEvTac (ATCC no. HTB-96) was maintained and transfected with a mCh-tagged off-target TALE. The off-target TALE, TAL248, binds to TATACTATCCAATCC (13). Flow cytometry results for all experiments are given in Supplementary Figure S1. For testing the background activity of the reporter plasmid, it was co-transfected with a mCh-positive cells, we determined the percentage of AmCyan fluorescent protein (CFP)-positive cells (% CFP + in mCh +). An example of the raw data and gating procedure is given in Supplementary Table S3.

\textbf{RESULTS}

\textbf{Split intein-mediated splicing of a TALE enables two-input logic computation}

We used intein-mediated protein splicing to implement Boolean logic AND gates (5). Our system makes use of TALE fragments incapable of activating transcription on their own; when both are present and correctly spliced, however, they reconstitute an active TALE capable of activating transcription, thereby creating a two-input AND gate in cells (Figure 1A).

As we desired to use a synthetic TALE with minimal cross-reactivity to the endogenous genome, we first computationally designed a TALE to have no cognate binding sites or predicted off-target binding sites (up to three mismatches) in the 1000-bp promoter regions of mouse and human genomes (13). We then characterized the
Figure 1. Engineering of a split TALE-based AND circuit. (A) Schematic of the split TALE reconstitution process. TAL-N:Int-N and Int-C:TAL-C fragments are expressed from CMV expression plasmids. The two split TAL–intein fragments dimerize and undergo protein splicing to reconstitute the full TALE, which binds and activates a CFP reporter containing a binding site for TAL118 upstream of a HSV minimal promoter (m-pr). TAL118 induces expression through a transcriptional activator domain VP64 (24). The N- and C-TAL fragments contain nuclear localization signals and are tagged with co-translationally cleaved t2A:zsYellow and t2A:mCherry, respectively. (B) Characterization of the 6 split TALE–intein pairs assayed by transient transfection in U2-OS cells. Reporter activity is indicated by the percentage of CFP positive in all mCh-positive cells, as measured by flow cytometry. As a control, the reporter plasmid was transfected with TAL118 or the off-target TAL248 (with both constructs containing a mCh tag). The inlay illustrates the positions of tested split sites in TAL118, with its cognate binding site shown above (from 5′ to 3′). The structural model illustrates the positions of split sites 4 to 6 in TAL repeat 6 and 7 (25). (C) Fluorescence microscopy images and flow cytometry results (bar graph) for the TAL118_s5 fragments transfected separately or in combination illustrating the AND gate behavior of the circuit. To control for transfection efficiency, input 10 and 00 were co-transfected with a construct expressing mCh. (D) Co-transfection of TAL118_s5 AND circuit fragments with reporter plasmids containing 1–6 binding sites for TAL118. Error bars indicate standard deviation from three biological replicates.

ability of this TALE, TAL118, to activate expression of CFP from a promoter containing its cognate binding site. Using flow cytometry, we confirmed that TAL118 is able to induce reporter transcription in U2-OS osteosarcoma cells (Supplementary Figures S1 and S2).

We next determined the optimal amino acid residues at which to divide TAL118 for use in a split-intein protein splicing strategy. We chose three split sites in the protein domain that is essential for binding the 5′ thymine in the TALE binding site and three sites in TAL repeats 6 and 7 of the DNA binding domain (Figure 1B). Based on these split points, we created six pairs of TAL118 split proteins, with each pair containing an amino- (N-) and a carboxy- (C-) terminal fragment fused to the appropriate part of a split dnaB mini-intein from Rhodothermus marinus (5). We cloned these fragments into expression vectors and co-transfected them with a reporter containing a TAL118 binding site upstream of a minimal herpes simplex virus type 1 (HSV) promoter and a CFP gene (Figure 1A). We assayed for reporter activation using flow cytometry and controlled for transfection efficiency by calculating the percentage of CFP-expressing cells within the population of successfully transfected cells (based on the expression of mCh). These values, as well as the mean CFP signal within mCh-expressing cells, can be found in Supplementary Table S3. Two of six TAL118 split pairs showed a greater than 10-fold induction in the number of CFP+ cells relative to an off-target TALE activator, with both pairs originating from split sites in TAL repeat 6 (Figure 1B). The TAL118 fragment pair originating from split site 5 (TAL118_s5) led to the highest induction (21-fold) and was used throughout the remaining experiments. Importantly, transfection of the TAL118_s5 fragment pair led to a 10-fold and 9-fold higher induction in the number of CFP+ cells relative to either the N- or C-fragments alone, confirming the AND gate behavior of this circuit (Figure 1C). Furthermore, circuit output increased when additional TAL118 DNA binding sites were added to the reporter, with inclusion of six binding sites leading to 16% CFP+ cells (Figure 1D). Taken together, these results demonstrate the feasibility of building an AND gate based on a split TALE, and show that TALE-based synthetic transcriptional activator proteins can be used for bio-molecular computation in mammalian cells.
A three-piece intein enables TALE-based three-input logic AND computation

A three-piece intein, originating from a dnaB mini-intein from Synechocystis sp., has previously been shown to induce protein trans-splicing in Escherichia coli (26). We investigated whether we could use such a three-piece intein to make the transcriptional AND gate circuit dependent on expression of an additional peptide and thereby extend it to three inputs. However, as the Ssp dnaB intein is expected to show low splicing efficiency in mammalian cells (27), we chose homologous split sites in the dnaB mini-intein from R. marinus (Figure 2A). We fused the N- and C-parts (Int-Ns and Int-C) of this three-piece intein to the corresponding TAL118_s5 fragments (Figure 2B). Co-transfection of these two parts with the middle part of the three-piece intein (int-M) led to a 43-fold induction in the number of CFP+ cells relative to an off-target control (Figure 2C), and to a 5-fold induction relative to a control lacking int-M, showing that high transcriptional output depends on the presence of all three input parts. These results demonstrate that three-piece intein-mediated splicing is functional in mammalian cells and that it can be applied to build a TALE-based transcriptional circuit that performs three-input AND computation.

Expression of AND gates from single plasmids and optimization of reporter output

The performance of synthetic transcriptional circuits whose parts are expressed from separate plasmids can be hampered by noise stemming from variations in
transfection efficiency. Assembly of a circuit as a single DNA construct circumvents this problem by ensuring a 1:1 ratio of its subparts and also facilitates genomic integration, which is necessary in many practical applications. For these reasons, we cloned all components of the TAL118_s5 two-input AND circuit onto a single DNA construct. To do so, we used a modular, isothermal assembly-based cloning method that allows efficient construction of DNA parts in a bacterial artificial chromosome. Using this system, we generated construct 2i-AND consisting of a TAL118-responsive CFP reporter upstream of the two constitutively-expressed C- and N-input parts of TAL118_s5 (Figure 3A). To decrease the likelihood of transcriptional interference between each part, we separated the individual gene components by inserting HS4 insulator sequences (28). We also made a control circuit (2i-Ctrl) by replacing four TAL repeats in the N-fragment (TAL-N) of 2i-AND. 2i-Ctrl reconstitutes a TALE with a DNA binding domain that is unable to recognize the TAL118 binding site in its reporter part. As such promoters are often weaker than the viral promoters used here, we sought to optimize transcriptional output of 2i-AND by modifying its reporter part. We replaced its minimal HSV promoter with either of two minimal CMV promoter variants, CMV-53 and CMV-74, which contain more regulatory sequences than the single TATA box present in the HSV promoter (29). We found that transfection of the CMV-53-containing 2i-AND circuit led to a 10-fold induction in the number of CFP+ cells relative to its corresponding control circuit (Figure 3B), and to a 3-fold induction relative to the HSV-containing 2i-AND, demonstrating that a stronger core promoter can enhance transcriptional output. Although inclusion of CMV-74 in 2i-AND led to similar levels of CFP+ cells (4-fold induction relative to the HSV-containing circuit), this circuit showed an only 4-fold increase in CFP+ cells relative to its corresponding control circuit due to the increased background activity of the CMV-74 promoter (Figure 3B).

In practical applications of the 2i-AND circuit, promoters of endogenous genes might drive expression of its input fragments. As such promoters are often weaker than the viral promoters used here, we sought to optimize transcriptional output of 2i-AND by modifying its reporter part. We replaced its minimal HSV promoter with either of two minimal CMV promoter variants, CMV-53 and CMV-74, which contain more regulatory sequences than the single TATA box present in the HSV promoter (29). We found that transfection of the CMV-53-containing 2i-AND circuit led to a 10-fold induction in the number of CFP+ cells relative to its corresponding control circuit (Figure 3B), and to a 3-fold induction relative to the HSV-containing 2i-AND, demonstrating that a stronger core promoter can enhance transcriptional output. Although inclusion of CMV-74 in 2i-AND led to similar levels of CFP+ cells (4-fold induction relative to the HSV-containing circuit), this circuit showed an only 4-fold increase in CFP+ cells relative to its corresponding control circuit due to the increased background activity of the CMV-74 promoter (Figure 3B).
Three-piece intein splicing can be performed in mammalian cells. This finding allowed us to generate an AND circuit in mammalian cells that can integrate three positive signals. The two- and three-input TALE-based logic circuits presented here extend the toolbox for engineering synthetic genetic networks in mammalian cells.
TALE-based logic gates have both advantages and disadvantages compared with other approaches in biological computation. In comparison with RNA-based circuits, one disadvantage of TALE- and other TF-based logic computation is its relatively long response-time. On the other hand, whereas RNA-based logic is generally used to integrate combinations of small molecules (32) or miRNAs (33), transcriptional circuits make it possible to respond to the activity of a combination of endogenous promoters. This has, for example, been useful in building a TF-based AND gate circuit that integrates two cancer-cell-specific promoters (34). Our TALE-based AND gate circuit offers several advantages due to the flexibility of DNA binding sequences that TALEs provide. First, the flexibility in DNA binding makes it possible to build logic circuits that are orthogonal to each other and to the host genome. Second, by changing the sequence of the TALE, one could envision direct regulation of endogenous genes by spliced TALEs. Finally, fusing the split TALE to a repressor domain instead of a transcriptional activation domain could enable to repress specific genes in response to two or three concurrent signals, effectively performing NAND gate computation. By combining two repressors or activators, TALEs could also be used to build NOR and OR gates, respectively (5).
Currently, most genetic circuits in the field of mammalian synthetic biology are encoded on multiple plasmids and tested in transient transfections (5,33,35). As many applications may require stable maintenance in the cell type of interest, it is important to show that circuits are functional when integrated in the genome at a low copy. We here assembled relatively large genetic constructs and tested them after PiggyBac transposase-mediated low-copy genomic integration. This allowed us to demonstrate that a TALE-based logic circuit is functional in a genomic environment. Practical applications of bio-computational networks, for example in a clinical setting, may further necessitate their genomic integration into safe harbor sites (36). Along this line, we here used a method for site-directed single-copy integration in the genome of mouse ES cells (23). The system depends on negative selection and thereby circumvents the need of a resistance cassette that would increase construct size and might interfere with circuit function. In addition, this integration method makes it possible to compare the activity of different circuit variants in the same genomic environment. Taking advantage of this system, we compared variants of an AND circuit in the same genomic context and identified the promoter strength in the reporter as a critical parameter to enhance its transcriptional output. Our finding that TALE-based logic computation is functional after single-copy genomic integration in ES cells suggests that such circuits could be used as sensors in cell therapy applications or for directing in vitro differentiation in tissue engineering.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors acknowledge A. Garg and J. Lohmueller for technical assistance, for helpful discussions and for carefully reading the manuscript.

FUNDING

This work was supported by funds from a European Molecular Biology Organization Fellowship and a Human Frontier Science Program Fellowship (to F.L.); a National Science Foundation Graduate Research Fellowship awarded (to J.P.T.); a National Sciences and Engineering Research Council of Canada Postdoctoral Fellowship (to J.H.C.); a National Institutes of Health training grant [GM007598 to R.R.R.]; funds from NIH [R01GM036373]; and Defense Advanced Research Projects Agency (DARPA) [4500000572 to P.A.S.]. Funding for open access charge: NIH [R01GM036373].

Conflict of interest statement. None declared.

REFERENCES