Cre-mediated, loxP independent sequential recombination of a tripartite transcriptional stop cassette allows for partial read-through transcription

Bapst, Andreas M; Dahl, Sophie L; Knöpfel, Thomas; Wenger, Roland H

Abstract: One of the widely used applications of the popular Cre-loxP method for targeted recombination is the permanent activation of marker genes, such as reporter genes or antibiotic resistance genes, by excision of a preceding transcriptional stop signal. The STOP cassette consists of three identical SV40-derived poly(A) signal repeats and is flanked by two loxP sites. We found that in addition to complete loxP-mediated recombination, limiting levels of the Cre recombinase also cause incomplete recombination of the STOP cassette. Partial recombination leads to the loss of only one or two of the three identical poly(A) repeats with recombination breakpoints always precisely matching the end/start of each poly(A) signal repeat without any relevant similarity to the canonical or known cryptic loxP sequences, suggesting that this type of Cre-mediated recombination is loxP-independent. Incomplete deletion of the STOP cassette results in partial read-through transcription, explaining at least some of the variability often observed in marker gene expression from an otherwise identical locus.

DOI: https://doi.org/10.1016/j.bbagrm.2020.194568

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-187376
Journal Article
Published Version

The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:
Bapst, Andreas M; Dahl, Sophie L; Knöpfel, Thomas; Wenger, Roland H (2020). Cre-mediated, loxP independent sequential recombination of a tripartite transcriptional stop cassette allows for partial read-through transcription. Biochimica et Biophysica Acta. Gene Regulatory Mechanisms, 1863(8):194568.
DOI: https://doi.org/10.1016/j.bbagrm.2020.194568
Cre-mediated, loxP independent sequential recombination of a tripartite transcriptional stop cassette allows for partial read-through transcription

Andreas M. Bapst\textsuperscript{a}, Sophie L. Dahl\textsuperscript{a,b}, Thomas Knöpfel\textsuperscript{a}, Roland H. Wenger\textsuperscript{a,b,*}

\textsuperscript{a} Institute of Physiology, University of Zürich, CH-8057 Zürich, Switzerland
\textsuperscript{b} National Centre of Competence in Research “Kidney.CH”, Switzerland

\textbf{ARTICLE INFO}

Keywords: Cre profiling DNA recombination Lineage tracing loxP Poly(A) signal

\textbf{ABSTRACT}

One of the widely used applications of the popular Cre-loxP method for targeted recombination is the permanent activation of marker genes, such as reporter genes or antibiotic resistance genes, by excision of a preceding transcriptional stop signal. The STOP cassette consists of three identical SV40-derived poly(A) signal repeats and is flanked by two loxP sites. We found that in addition to complete loxP-mediated recombination, limiting levels of the Cre recombinase also cause incomplete recombination of the STOP cassette. Partial recombination leads to the loss of only one or two of the three identical poly(A) repeats with recombination breakpoints always precisely matching the end/start of each poly(A) signal repeat without any relevant similarity to the canonical or known cryptic loxP sequences, suggesting that this type of Cre-mediated recombination is loxP-independent. Incomplete deletion of the STOP cassette results in partial read-through transcription, explaining at least some of the variability often observed in marker gene expression from an otherwise identical locus.

1. Introduction

The Cre-loxP recombination system derived from bacteriophage P1 is one of the most popular site-specific recombination tools. This system serves bacteriophage P1 to properly segregate P1 plasmids to daughter cells at cell division [1–3]. When a pair of directly repeated 34 bp loxP recombination sites is located on the same piece of DNA, P1 Cre recombinase excises the intervening DNA, releasing a circular DNA containing one loxP site and leaving the second loxP site behind [4]. Since its discovery, a vast number of applications of this recombination system beyond bacteria have been developed in various species, including in mouse cell lines [5] and transgenic mice [6,7].

Transgenically modified Cre-responder mouse strains which express fluorescent reporter proteins are now widely used to label specific cell populations, for instance to localize tagged cells within specific organs or to trace them during development or disease (fate mapping) [8]. A popular variant of this approach is Cre expression profiling using Cre recombinase-mediated excision of a transcriptional STOP cassette, flanked by directly repeated loxP sites, which is placed between a strong promoter and a marker gene. While this technique has originally been developed for targeted oncogene activation [7], it can also be used to permanently activate easily detectable reporter genes. Knock-in strategies allowed for the insertion of such conditional reporter genes into the permissive Rosa26 locus [9], providing a highly versatile tool to visualize the cell-type specificity of virtually any kind of promoter/Cre combination, following crossing of a Cre mouse strain with the reporter mouse strain. Using this approach, Madisen et al. [10] developed a set of Cre reporter mice, including the Ai14 strain, that employed a previously reported STOP cassette containing three identical 254 bp poly(A) signal repeats derived from the SV40 virus [11].

We recently generated a novel BAC transgenic mouse model where over 200 kb of the mouse Epo locus have been used to drive the expression of the conditional (i.e. tamoxifen-inducible) Cre\textsuperscript{ERT2} protein [12]. These animals were crossed with Ai14 reporter mice, containing the tandem dimeric Tomato (tdT) fluorescence reporter gene preceded by the loxP-flanked STOP cassette. Following exposure to tamoxifen and oxygen-deprived (hypoxic) conditions, this mouse model allowed for the permanent fluorescent tagging of renal Epo-producing (REP) cells [12]. During the isolation and cloning of REP-derived cell lines, we noticed lower tdT fluorescence in some of the cells. This finding was unexpected since tdT is expressed from the same endogenous single-copy gene locus in all cells. We hence sought for an explanation of these results and found a novel Cre activity, repeatedly resulting in partial recombination of the STOP cassette.

*Corresponding author at: Institute of Physiology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.
E-mail address: roland.wenger@access.uzh.ch (R.H. Wenger).

https://doi.org/10.1016/j.bbagrm.2020.194568

Received 9 January 2020; Received in revised form 21 April 2020; Accepted 22 April 2020
Available online 25 April 2020

1874-9399/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).
2. Materials and methods

2.1. Animals

The generation of Epo-CreERT2 (B6D2:CS7BL6/N-Tg (EPO-Cre),Rh) mice and the crossing with Ai14 reporter (B6.Cg-Gt (ROSA)26Sor < tm14(CAG-tdTomato)Hze > J) and Terminator (B6.129-Gt(ROSA)26Sor < tm1(DTR) > ) mice has been described previously [12]. To induce Cre in vivo, tamoxifen (Sigma-Aldrich, St Louis, MO, USA) at a dose of 200 mg/kg dissolved in corn oil/10% ethanol was applied daily for 5 days by gavage followed by exposure of the mice to 0.1% inspiratory carbon monoxide (CO) in air for 4 h as described in [12]. Alternatively, as a milder stimulus roxadustat (FG-4592; Selleckchem, Houston, Texas, USA) was co-applied with tamoxifen by subsequent daily gavages for 5 days at a dose of 50 mg/kg dissolved in 0.5 M Tris-HCl (pH 9.0). All animal experiments were approved by the veterinary office of the canton Zurich (license numbers ZH233/15 and ZH085/19).

2.2. Generation of renal cell lines

The generation of fibroblastoid atypical interstitial kidney (FAIK) cell lines derived from kidneys of 2 to 3 months old female Epo-CreERT2 mice crossed with Ai14 reporter mice has been described in detail previously [12]. Using the same protocol, we generated new REP-derived (REPD) cell lines from kidneys excised from homozygous Epo-CreERT2 mice containing both Ai14 and one Terminator allele in the Rosa26 locus. AB-REPD cells were obtained from a 2 months old male mouse 3 days after tamoxifen/0.1% CO induction of Cre in vivo. TK-REPD cells were obtained from non-treated 2.5 months old female mice, kept for 1 week in culture medium containing 10 μM tamoxifen and B27 supplement (Invitrogen, Carlsbad, CA, USA) and exposed twice for 16 h to 0.2% O2 at day 1 and 7 before treatment with 100 ng/ml diphertheria toxin until only a few tdT positive viable cells remained. Clonal FAIK1-10 and polyclonal TK-REPD4 cell lines were immortalized by lentiviral transduction with SV40 large T antigen as described [12]. Clonal AB-REPD2-22 were immortalized by gammaretroviral transduction with a temperature sensitive SV40 large T (pLPCX SV40 tsA58; kind gift from Parmjit Jat, London, UK) and cultured at 33 °C for 16 h to 0.2% O2. Following immortalization and expansion, pellets of TK-REPD4 cells appeared red whereas pellets of AB-REPD2-22 remained pale tdT fluorescence intensity. This variation in tdT expression was rather unexpected because tdT was expressed from the same constitutive active single-copy Rosa26 locus (Fig. 1A) in all cell lines analyzed. Following immortalization and expansion, pellets of TK-REPD4 cells appeared red whereas pellets of AB-REPD2-22 remained pale (Fig. 1B). Quantification of tdT mRNA demonstrated that tdT fluorescence intensity correlates with tdT transcript levels (Fig. 1C). In order to control for proper CreERT2-remediated transduction of the loxP-flanked STOP cassette, genomic DNA was analyzed by PCR using the primers indicated in Fig. 1A (see Supplementary Table 1 for sequences). Surprisingly, the PCR product of 201 bp (i.e. the predicted length after loxp-mediated recombination) was only detected in the REPD cell line TK-REPD4 (tdT high) whereas AB-REPD2-22 (tdT low) displayed a longer PCR product (Fig. 1D). The previously reported FAIK1-10 cell line [12] contains a non-recombined Ai14 locus (tdT null) which resulted in the expected 1072 bp PCR product (Fig. 1D). Direct sequencing of these PCR products revealed the deletion of all three poly(A) signal repeats in TK-REPD4 cells, but one remaining poly(A) signal repeat as well as both loxP sites and linker regions (the sequences between the loxP sites and the trimeric poly(A) signal repeats) in AB-REPD2-22 cells (Supplementary Fig. 1). In total, we obtained ten partially recombined AB-REPD cell clones, derived from two independent primary cell isolations, with one remaining poly(A) signal repeat (data experimental series. For confocal images, OCT-embedded tissue sections were analyzed on a Leica SP5 Mid UV-VIS microscope (Leica Microsystems, Wetzlar, Germany) and images were acquired using LAS X software (Leica Microsystems). Images were processed and analyzed using the Fiji distribution of ImageJ [15,16]. For FACS analysis, cells were detached using 2 mM EDTA (pH 8.0) in PBS, pelleted and re-suspended in PBS. Cells were analyzed on an LSRII Fortessa FACScanner (Becton Dickinson, Franklin Lakes, NJ, USA). tdT fluorescence was excited using a 561 nm laser and detected using a 586/15 nm filter. Dead cells were excluded by 4′,6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich) nuclear staining, using a 405 nm laser and a 450/50 nm filter. Data was analyzed using FlowJo software (FlowJo, Ashland, OR, USA).

2.5. RNA analysis

RNA was extracted and quantified by reverse-transcription (RT) real-time quantitative (q) PCR as described previously [12]. In brief, RT was performed with 2 μg total RNA and AffinityScript reverse transcriptase (Agilent), and the cDNA quantified using SYBR Green qPCR reagent kit (Kapa Biosystems, London, UK) in a MX3000P light cycler (Agilent). Transcript levels were calculated by comparison with calibrated standard curves and normalized to mouse ribosomal protein L28 mRNA. Primers used for RT-qPCR are listed in Supplementary Table 1 and were purchased from Microsynth (Balgach, Switzerland).

2.6. DNA analysis

DNA was isolated as described previously [12] and 100 ng per reaction were amplified by PCR (30–35 cycles of 95 °C/10 s, 64 °C/30 s, 72 °C/30 s) using KAPA2G Fast DNA polymerase (Kapa Biosystems) and the primers listed in Supplementary Table 1. PCR products were analyzed by agarose gel electrophoresis, gel-purification and DNA sequencing (Microsynth), either directly or following sub-cloning into pBluescript vector (Agilent) using Xhol and EcoRI restriction enzymes (Thermo Fisher Scientific, Waltham, MA, USA).

3. Results

3.1. Clonal variation in Cre-activated reporter fluorescence intensities

During isolation of primary tdT-positive REP cells from the kidneys of Epo-CreERT2 × Ai14 mice [12], we repeatedly observed single cells with lower tdT fluorescence intensity. This variation in tdT fluorescence was rather unexpected because tdT was expressed from the same constitutively active single-copy Rosa26 locus (Fig. 1A) in all cell lines analyzed. Following immortalization and expansion, pellets of TK-REPD4 cells appeared red whereas pellets of AB-REPD2-22 remained pale (Fig. 1B). Quantification of tdT mRNA demonstrated that tdT fluorescence intensity correlates with tdT transcript levels (Fig. 1C). In order to control for proper CreERT2-mediated recombination of the loxP-flanked STOP cassette, genomic DNA was analyzed by PCR using the primers indicated in Fig. 1A (see Supplementary Table 1 for sequences). Surprisingly, the PCR product of 201 bp (i.e. the predicted length after loxp-mediated recombination) was only detected in the REPD cell line TK-REPD4 (tdT high) whereas AB-REPD2-22 (tdT low) displayed a longer PCR product (Fig. 1D). The previously reported FAIK1-10 cell line [12] contains a non-recombined Ai14 locus (tdT null) which resulted in the expected 1072 bp PCR product (Fig. 1D). Direct sequencing of these PCR products revealed the deletion of all three poly(A) signal repeats in TK-REPD4 cells, but one remaining poly(A) signal repeat as well as both loxP sites and linker regions (the sequences between the loxP sites and the trimeric poly(A) signal repeats) in AB-REPD2-22 cells (Supplementary Fig. 1). In total, we obtained ten partially recombined AB-REPD cell clones, derived from two independent primary cell isolations, with one remaining poly(A) signal repeat (data...
not shown).

3.2. Cre dose-dependent sequential recombination of an endogenous STOP cassette

AB-REPD2-22 cells were obtained from primary REP cells isolated from mice treated with tamoxifen and a brief hypoxia exposure in vivo three days prior isolation. In contrast, TK-REPD4 cells were obtained from primary REP cells of untreated mice and were stimulated with tamoxifen and repeated hypoxia in vitro. Stronger transcriptional Cre induction and better tamoxifen accessibility in vitro than in vivo suggest that more CreERT2 was transcriptionally induced and translocated into the nucleus during TK-REPD4 than AB-REPD2-22 generation, leading to the hypothesis that a nuclear CreERT2 dose-dependent effect is responsible for the incomplete STOP cassette recombination. To analyze whether Cre dose-dependent sequential deletion of one, two or all three poly(A) signal repeats of the STOP cassette leads to a progressive increase in tdT expression (Fig. 2A), tdT null FAIK1-10 cells were transiently transfected with increasing amounts of an improved Cre (iCre) expression vector, or constant amounts of a tamoxifen-inducible Cre (ERT2-Cre-ERT2) expression vector, followed by treatment with increasing concentrations of tamoxifen. In both cases, cells with variable tdT fluorescence intensities were obtained, demonstrating that this effect is independent of a particular Cre protein modification (Fig. 2B).

While approx. 75% of all tdT-positive red cells were tdT low (blue in Fig. 2B, lower panel) after low amount iCre transfection, only 7% remained tdT low after high iCre transfection. Similarly, 26% were tdT low in the absence of tamoxifen (owing to the strong ERT2-Cre-ERT2 overexpression) but only 4% remained tdT low in the presence of 1 μM tamoxifen. It should be noted that for over 40 passages of in vitro cultivation we never observed any tdT-positive FAIK1-10 cells, neither under hypoxic conditions nor in the presence of tamoxifen or hydroxytamoxifen (data not shown), demonstrating that in the absence of exogenous Cre no STOP cassette recombination occurs.

Incomplete STOP cassette recombination was confirmed by FACS analysis, clearly showing that, besides non-recombined (corresponding to the naive FAIK1-10 population) and fully recombined (corresponding to the TK-REPD4 population) cells, there was also a substantial number of cells with intermediate tdT fluorescence intensity, corresponding to the AB-REPD2-22 population (Fig. 2C). PCR analysis of the STOP cassette again resulted in multiple bands, consistent with incomplete recombination (Fig. 2D). These PCR products were isolated from the gel, sub-cloned and sequenced, confirming the iCre-dose-dependent sequential recombination of the three poly(A) signal repeats (Supplementary Fig. 2). To confirm that Cre-mediated stepwise STOP cassette deletion is not specific for non-recombined FAIK1-10 cells, partially recombined AB-REPD2-22 cells were transfected with an iCre expression vector. Exogenous iCre expression resulted in a clear increase in tdT fluorescence intensity (Fig. 2E). PCR amplification using primers flanking the STOP cassette indicated that a large portion of the transfected cells underwent complete recombination (Fig. 2F), resulting in an increase in tdT mRNA levels (Fig. 2G).

3.3. Cre dose-dependent sequential recombination of an exogenous STOP cassette

To exclude that this incomplete STOP cassette recombination might occur only with the endogenous Ai14 reporter in the Rosa26 locus of our genetically modified mouse strains, we repeated these experiments using CHO cells transiently co-transfected with the Ai9 construct together with various Cre expression vectors as shown in Fig. 3B. Varying degrees of tdT fluorescence intensities again suggested partial recombination also of exogenous bacterial reporter plasmids (Fig. 3C). PCR amplification using the primers indicated in Fig. 3A, subcloning and sequencing confirmed sequential recombination of the tripartite STOP cassette (Supplementary Fig. 3). Of note, the recombination breakpoints inside the STOP cassette always precisely corresponded to the end/start of each poly(A) signal repeat without any similarity to the loxP sequence, suggesting that this type of Cre-mediated recombination is loxP-independent.

3.4. Mild Cre activation in vivo leads to a high proportion of cells with decreased reporter gene expression

Because the initial observation of REP cells derived from hypoxic kidneys with incomplete STOP cassette recombination may be based on
single cells with a coincidentally low Cre-mediated recombination event in vivo, we reasoned that a milder Epo inducing and hence Cre inducing stimulus would increase the proportion of tdT low REP cells in Epo-CreERT2 × Ai14 mice. Therefore, we treated these mice with roxadustat (FG-4592), a compound that has recently been approved in China for renal anemia therapy [17]. Roxadustat is a hypoxia-inducible factor prolyl-4-hydroxylase inhibitor that was designed for a relatively mild Epo induction to correct anemia in chronic kidney disease patients [18]. Combined tamoxifen/roxadustat treatment induced tdT positive REP cells in the mouse kidney (Fig. 4A). Automated analysis of full kidney slices demonstrated that many REP cells showed a weaker tdT fluorescence than others (Fig. 4B). To confirm that differences in fluorescence intensities between different cells are not due to varying z-planes, sequential images across the full 10 μm section were taken every
0.15 μm by confocal microscopy and reduced to a maximum intensity z-projection (Fig. 4C). Applying the same lookup table as used in Figs. 2 and 3, a clear tdT fluorescence intensity difference across the entire cell body could be demonstrated in tdT high compared to tdT low cells (Fig. 4D). From the same mice, primary kidney cells were isolated and expanded in vitro for several days. Microscopic analysis revealed different REP cell populations with diverse tdT fluorescence intensities (Fig. 4E), confirming our in vivo observations.

4. Discussion

The artificial SV40-derived tripartite STOP cassette [11] is widely used for many different applications in molecular biology. Our observations demonstrate that the Cre recombinase can cause, or is at least involved in, partial recombination of this cassette by targeting the end-start junctions of the three identical poly(A) signal repeats, leaving one or two poly(A) sequences intact. Without the presence of Cre we never observed any STOP cassette recombination, neither in vivo nor in vitro. Importantly, the poly(A) sequence shows no significant similarity to the loxP sequence and we did not find any abnormality in the two loxP sites outside the STOP cassette: while only one loxP site remained following complete (loxP-dependent) recombination, both loxP sites were still present following (loxP-independent) excision of one or two of the three poly(A) signal repeats. Overexpression of Cre in tdT null FAIK1-10 as well as in tdT low AB REPD2-22 was sufficient to completely remove the STOP cassette and fully restore tdT expression, further confirming that the Cre-loxP system was not generally corrupted in our experiments and that tdT mRNA expression is directly related to the extent to which the STOP cassette is truncated.

What could be the mechanism(s) underlying the partial Cre-mediated recombination of the STOP cassette? One possibility may be alternative loxP-like sequences within the STOP cassette's SV40 poly(A) repeats, which only partially match the canonical loxP sequence, such as loxB, loxL, and loxR sites [19] or other cryptic loxP sites [20–23]. However, when we aligned 36 different reported cryptic loxP sites to the STOP cassette, the best matching sequence was only 59% identical and covered only half of the 34 bp (cryptic) loxP site (#36 in Supplementary Fig. 4). Ignoring the spacer region and considering only the Cre recognition sequences thought to interact with Cre [23], one half of the palindromic loxP site almost perfectly matched the STOP cassette (#37 in Supplementary Fig. 5). However, there was one mismatch at a crucial position remaining and the other half of the palindrom did not match at all. In conclusion, while this highly similar cryptic recognition sequence might be involved in Cre recruitment to the loxP repeats, the absence of any match with the second half of the palindromic sequence suggests that other mechanisms than cryptic loxP sites may be involved.

Alternatively, the Cre recombinase may be recruited to the STOP cassette via the flanking loxP sites and creates DNA breaks which are then repaired. However, we sequenced 23 partially recombined STOP cassettes (Supplementary Figs. 2 and 3) and never observed any mutation, insertion or deletion at the end-start junctions of the three identical poly(A) signal repeats. Therefore, non-homologous end joining seems rather unlikely as a DNA repair mechanism. Regarding the tripartite perfect repeats, homologous recombination DNA repair proteins, possibly even recruited by the Cre recombinase, appear a more likely mechanism causing partial recombination of the STOP cassette.

Incompletely deleted tripartite poly(A) signal repeats led to partial transcription through the STOP cassette and suggests a Cre dose-dependent step-wise increase in marker/reporter expression. To our knowledge, such a Cre-mediated but loxP-independent recombination has not been reported before. A brief literature survey of this popular method immediately revealed several examples with apparently intermediary reporter or marker gene expression, in addition to cells with maximal expression [24–27]. Although these examples were randomly selected and certainly not comprehensive, they still illustrate that there...
might always be a minor cell population with incomplete inactivation of the STOP cassette. Especially for clonal experiments, it may be worthwhile to analyze the recombination of the STOP cassette, using the aforementioned method. Incomplete STOP cassette deletion should also be considered when analyzing in vivo expression patterns because weaker reporter signals may reflect lower Cre levels, despite being expressed from the same ubiquitous Rosa26 locus, which will be relevant for e.g. positive/negative threshold decisions.

CRediT authorship contribution statement

Andreas M. Bapst: Conceptualization, Investigation, Methodology, Formal analysis, Visualization, Writing - original draft. Sophie L. Dahl: Resources, Visualization. Thomas Knöpfel: Resources. Roland H. Wenger: Conceptualization, Supervision, Visualization, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors wish to thank Parmjit Jat, Inder Verma, Connie Cepko and Hongkui Zeng for the kind gifts of plasmids, and Patrick Spielmann for expert technical assistance.

Funding

This work was supported by the National Centre of Competence in Research “Kidney.CH” and the Swiss National Science Foundation (310030_184813).
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2020.194568.

References

[1] N. Sternberg, Demonstration and analysis of PI site-specific recombination using lambda-P1 hybrid phages constructed in vitro, Cold Spring Harb. Symp. Quant. Biol. 43 (Pt 2) (1979) 1143–1146, https://doi.org/10.1101/sqb.1979.043.01.128.

[2] N. Sternberg, D. Hamilton, Bacteriophage PI site-specific recombination. I. Recombination between loxP sites, J. Mol. Biol. 150 (1981) 467–486, https://doi.org/10.1006/jmbi.1996.0079.

[3] S. Austin, M. Ziese, N. Sternberg, A novel role for site-specific recombination in maintenance of bacterial replicons, Cell 25 (1981) 729–736, https://doi.org/10.1006/jmbi.1996.0079.

[4] K. Abremski, R. Hoess, N. Sternberg, Studies on the properties of PI site-specific recombination: evidence for topologically unlinked products following recombination, Cell 32 (1983) 1301–1311, https://doi.org/10.1006/jmbi.1996.0079.

[5] B. Sauer, N. Henderson, Cell-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome, Nucl. Acids Res. 17 (1989) 147–161, https://doi.org/10.1093/nar/17.1.147.

[6] P.C. Orban, D. Chui, J.D. Marth, Tissue- and site-specific DNA recombination in transgenic mice, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 6861–6865, https://doi.org/10.1073/pnas.89.15.6861.

[7] M. Lakos, B. Sauer, B. Mosinger Jr., E.J. Lee, R.W. Manning, S.H. Yu, K.L. Mulder, H. Westphal, Targeted oncogene activation by site-specific recombination in transgenic mice, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 6232–6236, https://doi.org/10.1073/pnas.89.14.6232.

[8] C.S. Branda, S.M. Dymecki, Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice, Dev. Cell 6 (2004) 7–28, https://doi.org/10.1016/s1534-5807(03)00399-x.

[9] P. S.oriano, Generalized lacZ expression with the ROSA26 Cre reporter strain, Nat. Genet. 21 (1999) 79–71, https://doi.org/10.1038/s0007.

[10] L. Madisen, T.A. Zwingman, S.M. Sunkin, S.W. Oh, H.A. Zarivcalsa, H. Gu, L.L. Ng, R.D. Palmer, M.J. Hawrylycz, A.R. Jones, E.S. Lein, H. Zeng, A robust and high-throughput Cre reporting and characterization system for the whole mouse brain, Nat. Neurosci. 13 (2010) 133–140, https://doi.org/10.1038/nn.2674.

[11] J.H. Maxwell, G.S. Harrison, W.M. Wood, F. Maxwell, A DNA cassette containing a trimerized SfiI polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription, BioTechniques 7 (1989) 276–280.

[12] F. Imeri, K.A. Nolan, A.M. Bapst, K. Abremski, R. Hoess, N. Sternberg, et al., A DNA cassette containing a hypoxia-inducible factor-1α activation, Blood 96 (2000) 1558–1565.

[13] T. Matsuda, C.L. Cepko, Controlled expression of transgenes introduced by in vivo electroporation, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1027–1032, https://doi.org/10.1073/pnas.0610151104.

[14] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D.J. White, V. Hartenstein, K. Eliezer, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis, Nat. Meth. 9 (2012) 676–682, https://doi.org/10.1038/nmeth.1901.

[15] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Meth. 9 (2012) 671–675, https://doi.org/10.1038/nmeth.2089.

[16] B. Thyagarajan, M.J. Guimaraes, A.C. Groth, M.P. Calos, Mammalian genomes contain active recombinase recognition sites, Gene 244 (2000) 47–54, https://doi.org/10.1016/s0378-1119(00)00008-1.

[17] S. Nilsson, R.H. Hoess, M. Ziese, N. Sternberg, P1 site-specific recombination: nucleotide sequence of the recombining sites, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 3398–3402, https://doi.org/10.1073/pnas.79.11.3398.

[18] B. Sauer, Identification of cryptic lox sites in the yeast genome by selection for Cre-mediated chromosome translocations that confer multiple drug resistance, J. Mol. Biol. 223 (1992) 911–928, https://doi.org/10.1016/0022-2836(92)90252-f.

[19] K. Abremski, R. Hoess, N. Sternberg, Demonstration and analysis of PI site-specific recombination using lambda-P1 hybrid phages constructed in vitro, Cold Spring Harb. Symp. Quant. Biol. 43 (Pt 2) (1979) 1143–1146, https://doi.org/10.1101/sqb.1979.043.01.128.

[20] H. Westphal, Targeted oncogene activation by site-specific recombination in transgenic mice, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 5037–5042, https://doi.org/10.1073/pnas.79.11.5037.

[21] B. Thyagarajan, M.J. Guimaraes, A.C. Groth, M.P. Calos, Mammalian genomes contain active recombinase recognition sites, Gene 244 (2000) 47–54, https://doi.org/10.1016/s0378-1119(00)00008-1.

[22] P.I. Missirlis, D.E. Smailus, R.A. Holt, A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination, BMC Genomics 7 (2006) 73, https://doi.org/10.1186/1471-2164-7.73.

[23] B. Thyagarajan, M.J. Guimaraes, A.C. Groth, M.P. Calos, Mammalian genomes contain active recombinase recognition sites, Gene 244 (2000) 47–54, https://doi.org/10.1016/s0378-1119(00)00008-1.

[24] P.I. Missirlis, D.E. Smailus, R.A. Holt, A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination, BMC Genomics 7 (2006) 73, https://doi.org/10.1186/1471-2164-7.73.

[25] S. Semprini, T.J. Troup, N. Kotelevtseva, K. King, J.R. Davis, L.J. Mullins, K.E. Chapman, D.R. Dunbar, J.J. Mullins, Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombination techniques, Nucl. Acids Res. 35 (2007) 1402–1410, https://doi.org/10.1093/nar/gkl1108.

[26] X. Mao, Y. Fujiwara, S.H. Orkin, Improved reporter strain for monitoring Cre recombination-mediated DNA excisions in mice, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 5037–5042, https://doi.org/10.1073/pnas.96.9.5037.

[27] T.C. Metzger, L.S. Khan, J.M. Gardner, M.L. Mouchess, K.P. Johannes, A.K. Krawisz, K.M. Skrzyczynska, M.S. Anderson, Lineage tracing and cell ablation identify a post-Aire-expressing thymic epithelial cell population, Cell Rep. 5 (2013) 166–179, https://doi.org/10.1016/j.celrep.2013.08.038.

[28] M. Hermann, P. Stillhard, H. Woldner, D. Seruggia, V. Kapp, H. Sanchez-Iranzo, N. Mercader, L. Montoliu, H.U. Zehlhofer, P. Pelczar, Binary recombinase systems for high-resolution conditional mutagenesis, Nucl. Acids Res. 42 (2014) 3894–3907, https://doi.org/10.1093/nar/gkt1361.

[29] K.J. Kaufman, M.A. Oehler, J.R. Dorkin, J.C. Kaczmarek, S. Bhadani, J.C. Kaczmarek, S. Bhadani, J. Wycko, M. Gassmann, R.H. Wenger, Epolones induce erythropoietin expression via hypoxia-inducible factor-1α activation, Blood 96 (2000) 1558–1565.
### Supplementary Table 1

**RT-qPCR primers**
- tdTomato_fwd 5'-acatggccgctcatcaaga-3'
- tdTomato_rev 5'-cttgtacagctgctccatgc-3'
- L28_fwd 5'-gcaaaaggggtcggtgtagt-3'
- L28_rev 5'-ttctgcttcgaaggatggc-3'

**qPCR/Subcloning primers** *(restriction sites in italics)*
- Ai14_recomb_fwd 5'-gcaacgtgctggttattgtg-3'
- Ai14_recomb_rev 5'-ggccgaattcgatctagctt-3'
- Ai14_recombSubcl_fwd 5'-tagccctcgaggcaacgtgctggtatttg-3'

**DNA sequencing primers**
- M13 5'-tgtaaaacgacggcagc-3'
- Ai14_recomb_fwd 5'-gcaacgtgctggttatttg-3'
- Ai14_recomb_rev 5'-ggccgaattcgatctagtt-3'
Supplementary Figure 1

Direct sequencing of the PCR products following amplification of the STOP cassette in FAK1-10, TK-REPD4 and AB-REPD2-22 cells. Yellow, loxP sequence; blue/green/violet, identical SV40 poly(A) signal sequences.

FAIK1-10 cells (tdT null)
GTTATTGTGCTGTCATCATTTTGGGCAAGAAATTGATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA GTTATTGTGCTGTCATCATTTTGGGCAAGAAATTGATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA GTTATTGTGCTGTCATCATTTTGGGCAAGAAATTGATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA

TK-REPD4 cells (tdT high)
GATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA GTTATTGTGCTGTCATCATTTTGGGCAAGAAATTGATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA

AB-REPD2-22 cells (tdT low)
GTTATTGTGCTGTCATCATTTTGGGCAAGAAATTGATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA GTTATTGTGCTGTCATCATTTTGGGCAAGAAATTGATTTGATACCGCGGGCCCTAAGAAGTCTCTT ATTCTCTGAGAAGTTAGGAGGATCTGCGACAATTTAATCTATTAATATAAACCTCCTGCGAGACTCTGCA

Detected by sequencing

| PCR products | Cell lines |
|--------------|------------|
| naive STOP cassette | 1072 bp | FAK1-10 |
| minus 2 poly(A) repeats | 564 bp | AB-REPD2-22 |
| recombined STOP cassette | 201 bp | TK-REPD4 |
Supplementary Figure 2

Sequencing of gel-isolated and sub-cloned PCR products following amplification of the STOP cassette in ERT2-Cre-ERT2-transfected FAIK1-10 cells. Grey, amplification/subcloning PCR primers (bold, restriction sites); yellow, loxP sequence; blue/green/violet, identical SV40 poly(A) signal sequences.

ERT2-Cre-ERT2-transfected FAIK1-10 cells

TAGCCTCGAGGCAACCTGCTGGTTATTGTGCTGCTCATCATTTTGGCAAGAATTGATTTTGATACCGCGGGGCCCTAAGAAGTTCTCTATTTCTAGAAAGTATAGGAACCTTCGCTGCACATTTAAATCTATTATAAACTTCGGCTCAACCCCACTCCCCCCTGGAACTGCAAATGAAATGCAATTGCATCACAATAATTACAAA

TAGCCTCGAGGCAACCTGCTGGTTATTGTGCTGCTCATCATTTTGGCAAGAATTGATTTTGATACCGCGGGGCCCTAAGAAGTTCTCTATTTCTAGAAAGTATAGGAACCTTCGCTGCACATTTAAATCTATTATAAACTTCGGCTCAACCCCACTCCCCCCTGGAACTGCAAATGAAATGCAATTGCATCACAATAATTACAAA

Detected by sequencing | PCR products | Times detected
--- | --- | ---
naive STOP cassette | 1082 bp | 3
minus 1 poly(A) repeat | 828 bp | 1
minus 2 poly(A) repeats | 574 bp | 15
recombined STOP cassette | 211 bp | 4
**Supplementary Figure 3**

Sequencing of gel-isolated and sub-cloned PCR products following amplification of the STOP cassette in CHO cells co-transfected with the pAi9 plasmid together with expression vectors for ERT2-Cre-ERT2. Grey, amplification/subcloning PCR primers (bold, restriction sites); yellow, loxP sequence; blue/green/violet, identical SV40 poly(A) signal sequences.

ERT2-Cre-ERT2-transfected CHO cells

| Detected by sequencing | PCR products | Times detected |
|------------------------|--------------|----------------|
| naive STOP cassette    | 1082 bp      | 2              |
| minus 1 poly(A) repeat | 828 bp       | 3              |
| minus 2 poly(A) repeats| 574 bp       | 4              |
| recombined STOP cassette| 211 bp      | 1              |

```
TAGCCTTCGAGGCAACGTGCTGGTTATTGTCTGTCTCATATTTTTGCAAGAATTTGATTGGATACCGGCGGGCCTAAGAAGTTCTCCTACTGAAAGTTATAGGAACTTCTCGTCGACATTTAATCATTTAATATAACTTCGTATATAGTATGCTATACGAAATTATTCGCGATGAATATTGAGGTTTACTTGCCTTTA
TAGCCTTCGAGGCAACGTGCTGGTTATTGTCTGTCTCATATTTTTGCAAGAATTTGATTGGATACCGGCGGGCCTAAGAAGTTCTCCTACTGAAAGTTCTCCTACTGAAAGTTATAGGAACTTCTCGTCGACATTTAATCATTTAATATAACTTCGTATATAGTATGCTATACGAAATTATTCGCGATGAATATTGAGGTTTACTTGCCTTTA
```
Supplementary Figure 4

Alignment between the STOP cassette (entire sequence between the two flanking loxP sites) and 34 cryptic loxP sites (#2 to #35) previously reported in the indicated publications. For reference, the canonical 34 bp loxP sequence (#1) is indicated in bold. The canonical recognition sequence with all bases irrelevant for Cre-binding replaced by the ambiguous base N, as well as a 34 bp sequence with all bases except the TATA motifs flanking the 8 bp spacer region replaced by N, are included as #36 and #37, respectively. Small letters denote mismatches between the cryptic and the canonical loxP sites. NCBI Blast software was used with a minimal length of 7 nt, allowing for maximally 2 mismatches and 0 gaps. Nucleotides of the reported cryptic loxP sites matching the STOP cassette are underlined, and the longest possible alignments (Length) as well as the corresponding % sequence identity (Similarity) are indicated.

| #  | Recognition                   | Spacer       | Recognition                                                                 | Length | Similarity | Reference |
|----|-------------------------------|--------------|------------------------------------------------------------------------------|--------|------------|-----------|
| 1  | ATAACTTCGTATA                 | ATGTATGC     | TATACGAAGTTAT                                                                | 7 bp   | 20.1%      | [19]      |
| 2  | ATcggCTTCGgATA                | ActTccTGc    | TATACGAAGTTAT                                                                | 0 bp   | 00.0%      | [19]      |
| 3  | ATAACTTCGTATA                 | ATGTATgt     | TATcCGAaacata                                                                | 12 bp  | 35.3%      | [19]      |
| 4  | ATcggCTTCGgATA                | ActTcccggt   | TATcCGAacata                                                                | 8 bp   | 22.9%      | [19]      |
| 5  | ATAACTTCGTATA                 | ATGTatAaa    | TATAtGctaaTTT                                                                | 8 bp   | 23.5%      | [20]      |
| 6  | ATAACTTCGTATA                 | ATGTATGa     | TATATGttGgagc                                                                | 10 bp  | 29.4%      | [20]      |
| 7  | ATAACTTCGTATA                 | ATGTgatga    | TATACcctttTTT                                                                | 17 bp  | 50.0%      | [20]      |
| 8  | ATAACTTCGTATA                 | ATGTATGg     | aAcAaAGgaAg                                                                  | 7 bp   | 20.1%      | [20]      |
| 9  | ATAACTTCGTATA                 | ATGTATGg     | aActCttttTTgT                                                                | 7 bp   | 20.1%      | [20]      |
| 10 | ATAACTTCGTATA                 | ATGTATGg     | TgTcCacAGgcAa                                                                | 8 bp   | 20.1%      | [20]      |
| 11 | ATAACTTCGTATA                 | ATGTATGg     | TaacctAtaTggT                                                                | 7 bp   | 20.1%      | [20]      |
| 12 | ATAACTTCGTATA                 | ATGTATGg     | TATatgAccccAa                                                                | 8 bp   | 23.5%      | [20]      |
| 13 | ATAACTTCGTATA                 | ATGTATGg     | TATcattAaTata                                                                | 14 bp  | 41.2%      | [21]      |
| 14 | AcAACCcatTATA                 | ATaTATaa     | TATATGatGTatT                                                                | 7 bp   | 20.1%      | [21]      |
| 15 | AgAACCcatTATA                 | ATaTATaa     | TATATGatGTatT                                                                | 7 bp   | 20.1%      | [21]      |
| 16 | AgAACCcatTATA                 | ATaTATaa     | TATATGatGTatc                                                                | 7 bp   | 20.1%      | [21]      |
| 17 | ATAcTaCTGATA                  | tatgtata     | TATACatAtaTAT                                                               | 0 bp   | 00.0%      | [21]      |
| 18 | ATAcTaCTGATA                  | tatgtata     | TATACatAtaTAT                                                               | 0 bp   | 00.0%      | [21]      |
| 19 | caAAACagGTATA                 | tgccggtata   | TATACGAaTggT                                                                | 9 bp   | 26.5%      | [21]      |
| 20 | ATAcTCgCaTAaA                 | cTtAcCca     | TATAaGgtGatT                                                                | 7 bp   | 20.1%      | [21]      |
| 21 | gTAACTgaaGTATA                | tgcata       | TATACGAatATAT                                                               | 0 bp   | 00.0%      | [21]      |
| 22 | catAAcCTGATG                  | tgcata       | TATACGAAGTTAT                                                               | 0 bp   | 00.0%      | [21]      |
| 23 | ATATgtaCaTAaA                 | tTaTaAaag    | TATAaGatGTatT                                                                | 7 bp   | 20.1%      | [21]      |
| 24 | ATAACTTCGTATA                 | ATGTATaC     | TATACGAAGTTAT                                                               | 7 bp   | 20.1%      | [22]      |
| 25 | ATAACTTCGTATA                 | ATGTgTaC     | TATACGAAGTTAT                                                               | 7 bp   | 20.1%      | [22]      |
| 26 | ATAACTTCGTATA                 | AaGTATCcc    | TATACGAAGTTAT                                                               | 0 bp   | 00.0%      | [22]      |
| 27 | ATAACTTCGTATA                 | AgaaAcca     | TATACGAAGTTAT                                                               | 0 bp   | 00.0%      | [22]      |
| 28 | ATAACTTCGTATA                 | aTaaAcca     | TATACGAAGTTAT                                                               | 0 bp   | 00.0%      | [22]      |
| 29 | ATAACTTCGTATA                 | AaTAgaa      | TATACGAAGTTAT                                                               | 8 bp   | 23.5%      | [22]      |
| 30 | ATAACTTCGTATA                 | cgaTAcca     | TATACGAAGTTAT                                                               | 0 bp   | 00.0%      | [22]      |
| 31 | taccgTTGTATA                  | ATGTATGc     | TATACGAAGTTAT                                                               | 7 bp   | 20.1%      | [22]      |
| 32 | ATAACTTCGTATA                 | ATGTATGc     | TATACGAacgta                                                                | 7 bp   | 20.1%      | [22]      |
| 33 | AaAtAgatCaTATA                | tgtgtata     | TATAtatAtaTAT                                                               | 7 bp   | 20.1%      | [23]      |
| 34 | ATAcAaaaagTATA                | tatatacana   | TATAtatGgaAtgAa                                                             | 7 bp   | 20.1%      | [23]      |
| 35 | ATTaTcTTCtTtTgTg              | cttcctttt    | TAAaAtttttTaAT                                                              | 7 bp   | 20.1%      | [23]      |
| 36 | ATNACNNCNTATA                 | NNNNNNNNN    | TATANGNNGNTAT                                                               | 20 bp  | 58.8%      | [23]      |
| 37 | NNNNNNNNNNTATA                | NNNNNNNN     | TATANNNNNNNNN                                                                | 15 bp  | 44.1%      | [23]      |
Supplementary Figure 5

Alignment between the STOP cassette (entire sequence between the two flanking loxP sites) and the canonical Cre recognition site (#1, bold) as well as 35 cryptic Cre recognition sites (#2 to #36) previously reported in the indicated publications. The canonical Cre recognition sequence with all bases irrelevant for Cre binding replaced by the ambiguous base N, as well as the Cre recognition sequence with all bases except for the TATA motif replaced by N, are included as #37 and #38, respectively. Bases required for Cre binding [23] are highlighted in blue. Small letters denote mismatches between the cryptic and the canonical loxP sites. LALIGN software was used with a minimal length of 4 nt, allowing for maximally 2 mismatches and 0 gaps. Nucleotides of the reported cryptic loxP sites matching the STOP cassette are underlined, and the longest possible alignments (Length) as well as the corresponding % sequence identity (Similarity) are indicated.

| # | Recognition Sequence | Length | Similarity | Reference |
|---|----------------------|--------|------------|-----------|
| 1 | ATAACTTCGTAATA       | 6 bp   | 46.2%      | [19]      |
| 2 | ATcgCTGGATA          | 7 bp   | 53.9%      | [19]      |
| 3 | TAACGAAGacata        | 10 bp  | 76.9%      | [19]      |
| 4 | TATAacCctaaTcT      | 8 bp   | 61.5%      | [20]      |
| 5 | TATAtGttGqgac       | 9 bp   | 69.2%      | [20]      |
| 6 | TATACctttTcT        | 10 bp  | 76.9%      | [20]      |
| 7 | aAcAaGAAGagaG       | 10 bp  | 76.9%      | [20]      |
| 8 | aActCttttTcT        | 7 bp   | 53.9%      | [20]      |
| 9 | TAacctAataTggT       | 8 bp   | 61.5%      | [20]      |
| 10 | TcCcacaAGGcAa        | 9 bp   | 69.2%      | [20]      |
| 11 | TATACGAAGcTg         | 7 bp   | 53.9%      | [20]      |
| 12 | TATAatgAcccAa        | 9 bp   | 69.2%      | [20]      |
| 13 | TATcatAtaTata       | 9 bp   | 69.2%      | [20]      |
| 14 | AaAACcattTATA        | 8 bp   | 61.5%      | [21]      |
| 15 | TATATGaGTATAT       | 10 bp  | 76.9%      | [21]      |
| 16 | AgAACcattTATA        | 9 bp   | 69.2%      | [21]      |
| 17 | TATATGaGTATac       | 7 bp   | 53.9%      | [21]      |
| 18 | ATAcaTaCGTATA        | 6 bp   | 46.2%      | [21]      |
| 19 | TATACatAtaTAT        | 9 bp   | 69.2%      | [21]      |
| 20 | caAACaagGTATA        | 8 bp   | 61.5%      | [21]      |
| 21 | TATACGAaTgTgT        | 6 bp   | 46.2%      | [21]      |
| 22 | ATAtCTgCaTaaA        | 8 bp   | 61.5%      | [21]      |
| 23 | TATAaGgTGTATAT       | 10 bp  | 76.9%      | [21]      |
| 24 | gTATACGaGTATA        | 7 bp   | 53.9%      | [21]      |
| 25 | TATACQatAtaTAT       | 5 bp   | 38.5%      | [21]      |
| 26 | catAacCTCGA         | 9 bp   | 69.2%      | [21]      |
| 27 | ATAtggaCaTATA        | 8 bp   | 61.5%      | [21]      |
| 28 | TATAaGtATGTTAT       | 6 bp   | 46.2%      | [21]      |
| 29 | taccgTTCGTATA        | 4 bp   | 30.8%      | [22]      |
| 30 | TATACGAaggtgta       | 6 bp   | 46.2%      | [22]      |
| 31 | AaAtggaCaTATA        | 7 bp   | 53.9%      | [23]      |
| 32 | TATAtatAtaTAT        | 4 bp   | 30.8%      | [23]      |
| 33 | ATAAaaaagTATA        | 9 bp   | 69.2%      | [23]      |
| 34 | TATAaGatagaTAT       | 10 bp  | 76.9%      | [23]      |
| 35 | ATAtTTCTTggTg        | 8 bp   | 61.5%      | [23]      |
| 36 | TAAatttttTaTAT       | 10 bp  | 76.9%      | [23]      |
| 37 | ATNCNNCNGTATA        | 12 bp  | 92.3%      | [23]      |
| 38 | NNNNNNNNNNTATA       | 13 bp  | 100.0%     | [23]      |