Published online 9 May 2023

Nucleic Acids Research, 2023, Vol. 51, No. 10 ***5285–5297***
https://doi.org/10.1093/nar/gkad366

Discovery and characterization of novel Cre-type tyrosine site-specific recombinases for advanced genome engineering

Milica Jelicic[1], Lukas Theo Schmitt[1], Maciej Paszkowski-Rogacz[1], Angelika Walder[1], Nadja Schubert[1,2], Jenna Hoersten[1], Duran Sürün[1] and Frank Buchholz[1,*]

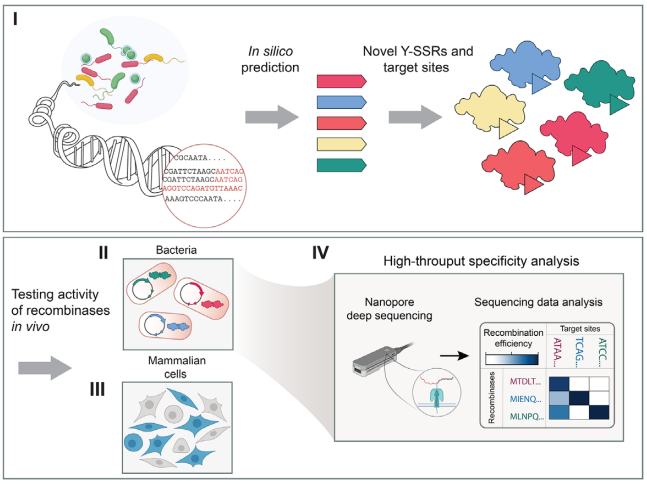
[1]Faculty of Medicine and University Hospital Carl Gustav Carus, UCC Section Medical Systems Biology, TU Dresden, 01307 Dresden, Germany and [2]DKMS Life Science Lab gGmbH, St. Petersburger Str. 2, 01069 Dresden, Germany

Received November 08, 2022; Revised April 21, 2023; Editorial Decision April 21, 2023; Accepted April 27, 2023

ABSTRACT

Tyrosine-type site-specific recombinases (Y-SSRs) are versatile tools for genome engineering due to their ability to mediate excision, integration, inversion and exchange of genomic DNA with single nucleotide precision. The ever-increasing need for sophisticated genome engineering is driving efforts to identify novel SSR systems with intrinsic properties more suitable for particular applications. In this work, we develop a systematic computational workflow for annotation of putative Y-SSR systems and apply this pipeline to identify and characterize eight new naturally occurring Cre-type SSR systems. We test their activity in bacterial and mammalian cells and establish selectivity profiles for the new and already established Cre-type SSRs with regard to their ability to mutually recombine their target sites. These data form the basis for sophisticated genome engineering experiments using combinations of Y-SSRs in research fields including advanced genomics and synthetic biology. Finally, we identify putative pseudosites and potential off-targets for Y-SSRs in the human and mouse genome. Together with established methods for altering the DNA-binding specificity of this class of enzymes, this work should facilitate the use of Y-SSRs for future genome surgery applications.

GRAPHICAL ABSTRACT



INTRODUCTION

Site-specific recombinases (SSRs) are reliable tools for targeted modification of genomes, with a variety of applications in research, medicine, and biotechnology. SSRs are divided into two evolutionarily and mechanistically distinct families of enzymes: the tyrosine and the serine recombinases (1). Nevertheless, SSRs can catalyze both cleavage and immediate resealing of DNA strands without the help of additional proteins, unlike DNA-altering systems such as CRISPR-Cas and other nuclease-based technologies (1). Although nuclease-based approaches are being intensively developed to expand their utility (i.e. base editors and prime editors (2–5)), these systems still typically introduce DNA-nicks and rely on cell intrinsic repair mechanisms (1). In contrast, SSRs can operate autonomously and the outcome of recombination is very predictable. Furthermore, most

*To whom correspondence should be addressed. Tel: +49 351 463 40277; Fax: +49 351 463 40289; Email: frank.buchholz@tu-dresden.de
Present address: Lukas Theo Schmitt, Seamless Therapeutics GmbH, Tatzberg 47/49, 01307 Dresden, Germany.

SSRs are relatively small in size, rendering them more convenient for different delivery vectors (1).

Tyrosine recombinases such as Cre and Flp are used extensively for genome engineering due to their simplicity and their ability to conduct efficient genome modifications in heterologous hosts (6). Hence, these SSRs were extensively used to model and understand how these types of enzymes work. The Cre/*loxP* system is derived from bacteriophage P1 and consists of the recombinase Cre and the 34-bp target sites *loxP* (7). loxP sites are palindromic sequences with two 13-bp inverted repeats separated by an 8-bp spacer. Each half-site is bound by one recombinase monomer forming a tetrameric complex on two *loxP* sites, which together catalyze strand exchange within the spacer region (8). Depending on the relative orientation of the spacers, Cre is able to perform a variety of reactions such as excision, integration, inversion and translocations (1). Furthermore, when presented with two heterospecific target sites in the genome, Cre alone, or in combination with other SSRs, can perform recombinase-mediated cassette exchange (RMCE) for precise replacement of a DNA fragment (1,9,10).

The functionality of Y-SSRs is particularly evident in the fact that they can be used in animal models of various human diseases (11,12). In addition, several elegant strategies for temporal, spatial and cell type-specific control of recombination in animals and plants have been developed, allowing SSRs to contribute to the study of embryonic development and lineage tracing (13–17). The combination of different SSRs in the same cell or organism have made it possible to establish ever increasing sophisticated systems (14,16–20), and several different enzymes have recently been utilized in synthetic biology to build complex genetic circuits (21–24).

Although all these features make SSRs very attractive genome editing tools, their application is limited due to several factors. First and foremost, SSRs are highly dependent on a specific target sequence that is typically not present in foreign genomes, and depending on the position even a single base change in their target site can eliminate efficient function of the system. To overcome this limitation, efforts have been made to alter recombinase target site specificity by means of directed molecular evolution (25–27). Several designer recombinases have indeed been developed to perform genomic surgery with high promise for therapeutic applications (28–31). However, their development is still laborious and time-consuming, especially if the new target site differs substantially from the original sequence. In this case many cycles of directed evolution are required to obtain a functional recombinase with activity on the new site. Naturally occurring SSR variants, each with their own unique substrate specificity, provide a more rational starting point to facilitate the process of altering the recombinase sequence specificity. The SSR system with the highest native target similarity to the novel chosen target can be selected. As we characterize more Cre-type SSRs, options for targetable substrates increase substantially. Moreover, we predict that the collection of natural Y-SSRs with defined specificities could ultimately help researchers to improve the rational design of Y-SSRs with tailored specificities.

Here, we expand the Y-SSR toolbox by computationally identifying and experimentally characterizing Y-SSRs and their target sites, thus broadening the sequence space that one can target with these tools. In total, we describe eight new Y-SSR systems, explore their specificity and finally validate their activity and utility in mammalian cells.

MATERIALS AND METHODS

Identification of putative recombinases and their target sites

Potential new Y-SSRs were identified by using tblastn from BLAST + 2.10.1 (https://www.ncbi.nlm.nih.gov/books/NBK131777/) with the protein sequences of Cre, Vika (32), Nigri, and Panto (33) as references to search the NCBI nucleotide collection database (v5). The results were filtered for below 90% identity and a sequence length of 300 to 400 amino acids with GNU awk. Protein sequences were acquired with efetch (https://dataguide.nlm.nih.gov/edirect/efetch.html). Full genome sequences of the potential SSRs were gathered using bastdbcmd (part of BLAST+). Potential target sites were identified by searching the genome sequences 1000 bp upstream and downstream of the potential Y-SSRs for palindromes. Palindrome search was performed with EMBOSS palindrome (v6.6.0.0) (34) with a minimum palindromic length of 13–15 bp and a gap limit of 8 with one mismatch allowed. The program output was then converted to a tabular format with GNU awk and combined with the protein data in R with the dplyr package. Potential SSRs with a Levenshtein distance (stringdist R package, https://journal.r-project.org/archive/2014/RJ-2014-011/index.html) below 10 to the references were removed. Potential SSRs were clustered with complete hierarchical clustering (base R) based on Levenshtein distances and cluster groups were formed with a cut-off distance of 11. The same clustering method was also used on the potential half-sites of the palindromic sequences, here the cluster cut-off was a distance of 2. The top candidates for testing were chosen considering the clustering, their organism they were found in and their distance to the reference recombinases (Supplementary Table S3, Supplementary Figure S1B).

The phylogeny tree of known and putative recombinases was generated by performing an all-against-all pairwise sequence alignment of the protein sequences using EMBOSS needle (35), followed by complete hierarchical clustering of the sequence dissimilarities. Visualization of the tree was done with R packages tidygraph and ggraph. For brevity, the tree was cut at the 98% sequence similarity, to represent almost identical proteins as a single node.

Human and mouse genomic sequences with high similarity to potential target sites were identified using PatMaN (36). The search was performed on half-sites only, allowing for up to 2 mismatches. If two genomic sequences matching the same half-site were found to be located on opposite strands, with a distance of 8 bp between them, they were called a potential target site of the respective recombinase. Genomic coordinates manipulation and sequence extraction steps were performed with the BEDTools suite (37).

Plasmid construction

For expression in *Escherichia coli*, codon-optimized DNA sequences of 17 predicted candidates were synthetized by

Twist Biosciences and cloned into pEVO vector via BsrGI and XbaI restriction sites (Supplementary Figure S3A) (25). Target sites were introduced via primers that were designed to carry the desired target sites and an overlap with the pEVO vector (Supplementary Table S1). The PCR fragment, that was generated when using the pEVO vector as template, was then cloned via Cold Fusion into a BglII digested pEVO backbone (System Biosciences).

For expression of recombinases in mammalian cells, a lentiviral PGK-NLS-BFP plasmid was used (Supplementary Figure S7A). Recombinase sequences were amplified from pEVO vectors and cloned via BsrGI and XbaI restriction sites (Supplementary Table S1). The before mentioned lentiviral vector, harboring recombinases was either transfected into HEK293T cells for transient expression, or used for virus production and infection for continuous expression.

For the construction of the recombination reporters, pCAG-loxP-mCherry-loxP-GFP 'traffic light' vector was used (Supplementary Figure S7A) (28). Oligonucleotides containing respective target sites were used to amplify mCherry cassette and the fragment was later ligated to the pCAG plasmid via NheI and HindIII restriction sites (Supplementary Table S1).

Recombination reporter assays

To visualize the recombination activity of recombinases on their predicted target sites, a plasmid-based assay was used as previously described (32,33) (Figure 1C). In short, expression of the recombinases from the pBAD promoter was induced with L-arabinose (Sigma-Aldrich Chemie GmbH). Single clones containing the pEVO plasmid with the recombinase and recombination target sites were cultured overnight in 6 ml LB medium with 25 μg/ml Cm and either 0 or 100 μg/ml L-arabinose at 37°C and 200 rpm. The recombinase mediated excision event was detected by agarose gel electrophoresis after digestion with BsrGI and SbfI restriction enzymes. The recombined plasmid is smaller in size compared to the non-recombined plasmid. Therefore, after gel electrophoresis a slower migrating non-recombined band (∼5.0 kb), and a faster migrating (∼4.3 kb) band for recombined plasmids can be seen (Figure 1C).

To compare the recombination efficiency of the active recombinases on their native target sites, recombinase expression was induced with increasing concentrations of L-arabinose (0, 1, 10 or 100 μg/ml medium) overnight in 6 ml culture volume. The test digest was done for each induction level and recombination efficiency was estimated by agarose gel electrophoresis. To quantify the recombinase activity, the ratio of band intensities was determined using Fiji-ImageJ for image processing. The quantified recombination was plotted in R 4.0.3 with dplyr v1.0.7 and visualized with ggplot2 v3.3.5. All test digests were done in triplicates ($n = 3$).

For the mammalian recombination reporter assay, HEK293T cells were plated at a density of 2×10^5 cells per well in 24-well dishes and cultured in glucose Dulbecco's Modified Eagle's Medium (DMEM, Gibco®), supplemented with 10% fetal bovine serum (Invitrogen), 1% Penicillin-Streptomycin (10000 U/ml, Thermo Fisher). At a confluency of 70–80%, cells were co-transfected with pPGK-NLS-Recombinase-P2A-BFP plasmids expressing a recombinase and pCAG-lox-mCherry-lox-GFP traffic light reporters using Lipofectamine® 2000 Transfection Reagent (Invitrogen) according to manufacturer´s instructions. Per well 0.5 μg of DNA (0.25 μg of each plasmid) and 2.5 μl of Lipofectamine® 2000 reagent diluted in 100 μl Opti-MEM® Reduced Serum Media each were used. On the next day, the media was changed and the cells were further cultured at 37°C and 5% CO_2. Upon recombination between the target sites, the mCherry cassette is excised and CAG promotor starts driving the expression of downstream green fluorescent protein (GFP). The cells were analyzed 2 days after transfection with fluorescent activated cell analysis and were then imaged with a fluorescent microscope (EVOS FL imaging system; Thermo Fisher Scientific).

Western blotting

Sample preparation for western blot analysis was performed by trichloroacetic acid (TCA) precipitation. pEVO-N-FLAG plasmids carrying 8 newly described recombinases and Vika and Cre as controls were transformed in *E. coli* and grown ON with 10 μg/ml L-arabinose to induce recombinase expression. Next day, the cultures were diluted and grown until $OD_{600} = 0.5$. The cultures were then incubated with ice-cold TCA for 30 min and then the pellets were centrifugated and washed with cold acetone. Remaining pellets were resuspended in 2× Laemmli buffer and incubated at 95°C for 5 min. 20 μl of each sample were separated by electrophoresis on NuPAGE™ Novex 4–12% Bis–Tris Gel (Invitrogen) with 1× NuPAGE™ MOPS SDS Running Buffer (20×) (Invitrogen) in an Invitrogen™ XCell SureLock™ Mini-Cell and XCell II™ Blot Modul (Invitrogen) at 120 V for 80 min. Western blotting was performed using standard procedures. The Nitrocellulose Blotting Membrane was stained using 1× Amido black solution to visualize total protein fraction and then photographed as to use it for loading control. For the detection of recombinase expression monoclonal ANTI-FLAG® M2 mouse antibody (1:5000, Sigma-Aldrich) was used. The membranes were then incubated with IRDye® 800CW Donkey anti-Mouse IgG Secondary Antibody (1:10 000, LI-COR Biosciences) and protein signal was then detected using the Odyssey® Classic Imaging System (LI-COR) and Image Studio™ Software (LI-COR). The band intensities were quantified by using Fiji-ImageJ and visualized with GraphPad Prism. The quantification was done from three independent biological replicates.

Fluorescent activated cell analysis

HEK293T were washed once with PBS and then detached using Trypsin (Gibco). The cells were then resuspended in Dulbecco's modified Eagle's medium (DMEM, Gibco®) and analyzed with the MACSQuant® VYB Flow Cytometer (Miltenyi). Analysis of the data was performed using FlowJo™ 10 (BD). All gating strategies are displayed in Supplementary Figure S7.

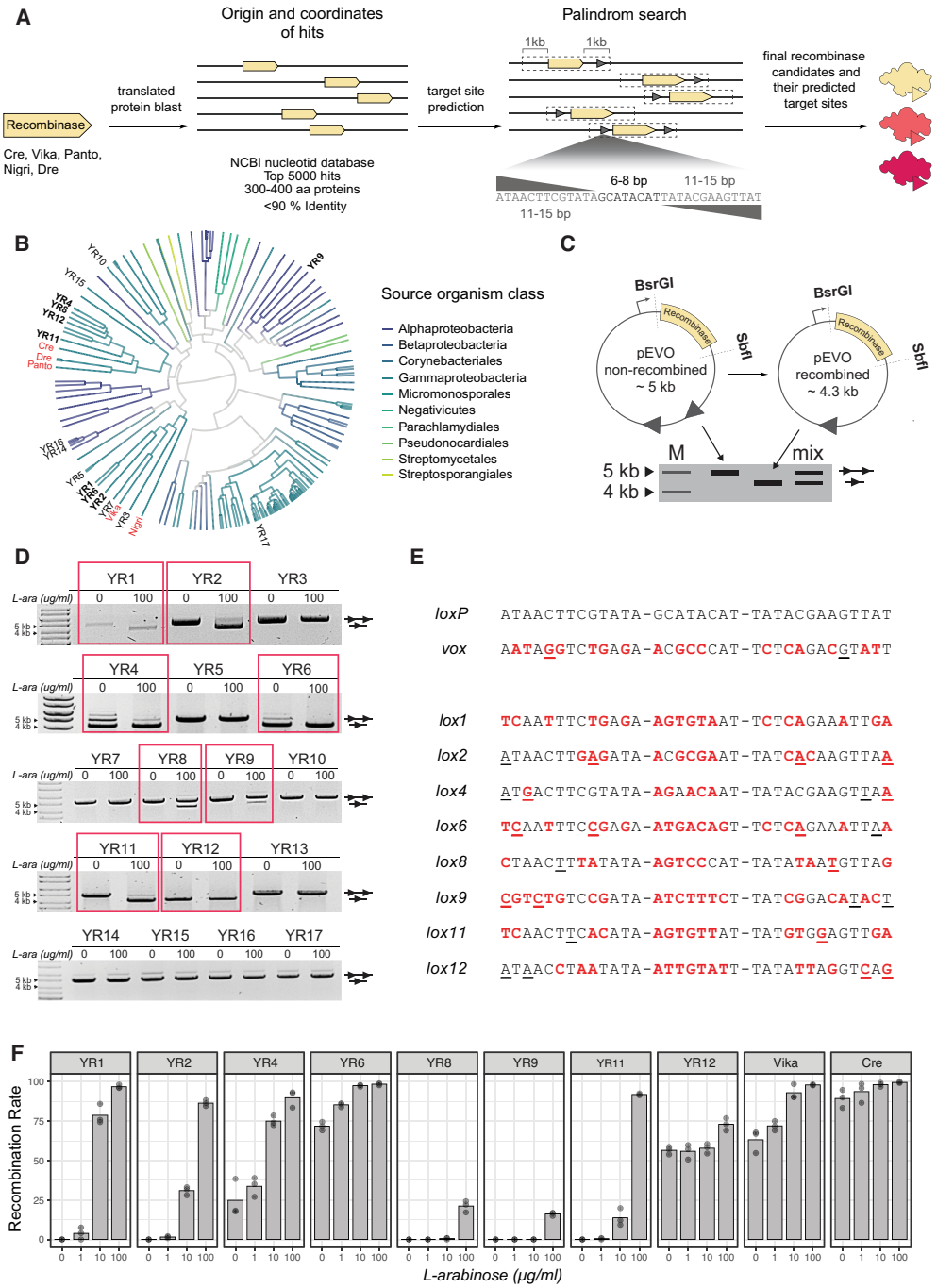


Figure 1. Discovery and experimental validation of recombinases and their target sites. (**A**) Schematic of computational workflow for prediction of novel Y-SSR candidates and their putative target sites in bacterial genomes. Dashed boxes present the palindrome search region while the triangles represent putative target site candidates. Zoom in depicts the criteria used for target site search: 11–15 bp palindromic half sites separated by 8 bp spacer with up to 2 asymmetries. (**B**) Phylogenetic tree of discovered Y-SSR candidates, with branches color coded based on the class of the source organism. Candidates chosen for experimental validation are labeled as well as previously known recombinases (in red). New Y-SSRs active on their predicted target sites are depicted in bold. (**C**) Overview of plasmid recombination assay. Important features, such as restriction sites, recombinase coding sequence and target sites (triangles) are shown. A schematic representation of expected recombination products on an agarose gel is shown below. Marker and expected sizes of recombined and unrecombined plasmids are indicated separately or together as usually seen on the gels. (**D**) Recombination activity of seventeen tested putative Y-SSR/target site pairs. Each sample was tested with or without L-arabinose (100 µg/ml) added to the growth medium for recombinase expression, indicated with '0' or '100'. Recombination is indicated by the band aligned with the single triangle and non-recombined plasmids are indicated by two triangles. Active recombinases are highlighted in a red box. M = GeneRuler™ DNA Ladder Mix (Thermo Fisher). (**E**) Nucleotide sequence alignment of the target sites that were recombined by respective putative recombinases. Full target sequences are separated into two half sites flanking the spacer sequence with a dashed line. Underlined nucleotides in the sequence represent asymmetric positions and the nucleotides in red depict differences to the loxP sequence. (**F**) Quantification and reproducibility of recombination activity of new SSRs in *E. coli*. Recombinase expression was induced with rising concentrations (µg/ml) of L-arabinose indicated along the x-axis. Vika and Cre were included as positive controls. Recombination was calculated from measuring the band intensities from agarose gels shown in Supplementary Figure S4A. Bacterial assays were done in triplicates ($n = 3$).

Overexpression studies in mammalian cells

For the viral delivery, pPGK-Recombinase-P2A-BFP vectors were used to produce lentiviral particles as described previously (38). NIH/3T3 mouse fibroblasts were seeded at a density of 4×10^4 cells per well in 24-well plates and grown at 37°C and 5% CO_2. The next day, fibroblasts were transduced with the different lentiviruses with a MOI of 0.5 in order to achieve about 50% of infection rate. The percentage of BFP expressing cells from at least 2×10^4 cells was tracked over the course of 15 days using MACSQuant® VYB Flow Cytometer (Miltenyi). The difference in the percentage of BFP cells at the last time point (day 15) and first time point was calculated and visualized with GraphPad Prism and statistical significance relative to BFP control was calculated by doing one-way ANOVA test with 95% CI.

Cross-recombination assay: nanopore sequencing

Eight new recombinases and Cre, Vika, Panto, Dre (39), VCre (40) were amplified from pEVO vectors, cleaned using the Isolate II PCR and Gel Cleanup Kit (Bioline) and mixed together in a 1:1 ratio. All respective target sites were cloned into the pEVO vectors with Cold Fusion Cloning kit as previously described and the resulting vectors were also mixed with equal molar ratio. Both, the mix of recombinases and pEVO backbones were digested with BsrGI and SbfI and ligated in a single reaction, thus creating a library of different recombinase/ target site pairs. Plasmids were transformed in XL1-Blue electrocompetent *E. coli* cells and grown overnight with 100 μg/ml L-arabinose to induce recombinase expression. On the next day, plasmids were linearized with BsrGI and ScaI and fragments carrying the recombinase sequence and target sites were isolated by agarose gel excision using the Isolate II PCR and Gel Cleanup Kit (Bioline). These DNA fragments were then prepared for nanopore sequencing with the SQK-LSK110 Kit according to the 'Amplicons by Ligation' protocol on a MinION R9.4.1 Flow Cell (Oxford Nanopore Technologies). Base calling of the sequence data was performed with guppy v5.0.7 on the high accuracy model (Oxford Nanopore Technologies). The sequence reads were then filtered for a read length of at least 1800 bp and a minimum mean phred score of 10 with filtlong (https://github.com/rrwick/Filtlong). To identify the recombinases, the reads were aligned to the reference recombinase sequences with minimap2 v2.17 (41). The target sites were identified using exonerate v2.2.0 using the affine:local model (https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate). The read ID and the matching references were then extracted from both alignments and combined in R with the dplyr package. Visualization of the data was performed with the R package ggplot2.

RESULTS

In silico prediction of putative recombinases and their target sites

In order to identify novel Y-SSRs, we searched the NCBI nucleotide collection for sequences resembling already characterized SSRs. To avoid identifying too similar recombinases with potentially same target sites, we filtered the results of this search for a sequence identity of less than 90% to the references (Cre, Vika, Panto and Nigri) and for a protein sequence length of 300–400 amino acids typical for Cre-type SSRs (Figure 1A) (42). To select top candidates for testing, we considered the sequence similarity to the reference sequences (Supplementary Figure S1B), the composition of the potential target site, and the conservation of catalytical residues (corresponding to R173, H289, R292, K201, W315 and Y324 in Cre) (Supplementary Figure S2).

More challenging than identifying new SSRs is finding their native target site. Previous comparative genomic studies of phages suggest that target sites are typically found directly upstream or downstream of the recombinase coding sequence (43,44). We therefore set the search criteria for target sites in close proximity of recombinase genes, 1 kb upstream and downstream of the coding sequence of the enzyme (Figure 1A). Since Y-SSR target sites are frequently palindromic, we used the EMBOSS palindrome search tool to identify potential target sites. We focused on palindromic sequences that contain 13–15 bp inverted complementary repeats allowing up to two asymmetrical mismatches, that are separated by an 8 bp spacer. (Figure 1A). Furthermore, all the hits were filtered to have at least 3 mismatches (spacer included) from previously known target sites. With these measures, we generated a data set of approximately 500 putative recombinase/ target site pairs providing a rich resource for experimental validation (Figure 1B, Supplementary Figure S1A, Supplementary Table S2).

8 new recombinases recombine their predicted target sites in bacteria

In order to experimentally test the activity of the candidate recombinase on the predicted target sequence, we randomly selected 17 candidates that best fit our criteria and individually cloned their coding sequence into the L-arabinose inducible pEVO recombination reporter vector (25) harboring two copies of the respective predicted target sites (Supplementary Table S3, Supplementary Figure S3A). The plasmids were then transformed into *E. coli* and cultured overnight in medium containing L-arabinose to induce recombinase expression. Upon expression, successful recombination leads to excision of a ∼700 bp DNA fragment from the plasmid. This size difference can be visualized by agarose electrophoresis of linearized plasmids (Figure 1C). Eight out of seventeen candidates showed activity on their predicted target site, evident by the appearance of ∼4 kb recombination bands (Figure 1D, Supplementary Figure S3B). Five of the candidates already showed efficient recombination in the samples without addition of L-arabinose to the medium (YR1, YR2, YR4, YR6 and YR12), indicating that these enzymes are active even when expressed at very low levels. Other recombinases (YR8, YR9 and YR11) recombined the plasmid only when L-arabinose was present in the growth medium, suggesting that they require a higher induction to become active in this assay. Interestingly, target sites of active recombinases shared two prominent features:

a G(T)-A-T(G)-A motif close to spacer sequence and a conserved T at position 7 (Supplementary Figure S3C). These results demonstrate that the established pipeline is able to identify novel Y-SSRs and their respective target sites at a success rate of approximately 50%.

To characterize the active Y-SSRs in more detail, we quantified their recombination efficiencies at different expression levels in order to investigate dose response and to obtain a better side-by-side comparison of their efficiencies compared to the well-established recombinases Cre and Vika (32). pEVO plasmids with desired recombinase/target site pairs were transformed into *E. coli* and were grown over night at different concentrations of L-arabinose to induce expression of the recombinase (45). Extracted plasmid DNA was then assayed for recombination on agarose gels (Supplementary Figure S4A). Quantification of band intensities revealed that the new recombinases have different activity profiles on their respective target sites (Figure 1F). Although, most of the recombinases were highly active when grown at high L-arabinose concentrations, showing recombination rate between 87 and 100% (YR1, YR2, YR4, YR6 and YR11), they behaved quite differently when expressed with low L-arabinose concentrations. While YR1, YR2, YR4 and YR11 showed low recombination rates when induced at 1 or 10 μg/ml of L-arabinose, YR6 was highly active even at these low induction levels, with its activity profile resembling Cre and Vika (Figure 1F). Interestingly, the YR12 recombinase showed a mostly constant recombination rate raging from ∼50% at 0 μg/ml L-arabinose and peaking at 70% when induced with 100 μg/ml of L-arabinose. YR8 and YR9, on the other hand, showed the weakest activity and only recombined their target sites to 25% and 16%, respectively, at the highest L-arabinose concentration. Nevertheless, with just a few rounds of substrate-linked directed evolution (25), we managed to increase the activity of YR9 recombinase significantly (Supplementary Figure S5), indicating that the activity of weaker enzymes can be increased with standard directed molecular evolution methods. The best performing clones showed about seven-fold improvement in activity in bacteria (Supplementary Figure S5C), and eight-fold improvement in mammalian cells (Supplementary Figure S5E).

One possible explanation for different activities of the examined enzymes could be different protein levels, due to varying production- and degradation-rates of the enzymes in the cells. To investigate this possibility, we measured protein concentrations in *E. coli* for each recombinase by Western blotting (Supplementary Figure S4B). We cloned all the recombinases in a modified pEVO vector harboring an N-Flag tag, and transformed and expressed the recombinases at 10 μg/ml of L-arabinose overnight. Using a Flag-antibody, we observed differential protein levels of the recombinases (Supplementary Figure S4C). Cre showed the highest levels, followed by YR4 and YR8 recombinases that showed just slightly lower values (Supplementary Figure S4C). YR9 and YR11 showed the weakest expression levels compared to the other recombinases. Interestingly, the expression levels do not correlate too well with the enzyme activities, suggesting that other factors such as catalytic activity, DNA binding affinity, dimerization or speed of isomerization also contribute to the difference in recombination efficiencies.

Profiling target-site selectivity of Cre-type recombinases

SSRs with different sequence specificity are frequently used in combination to allow sophisticated genomic or synthetic biology experiments (11,46–49) For these experiments it is important to know the specificity of the enzymes and to consider possible cross reactivity (50,51). This information may also help to provide more insight on how these enzymes recombine specific nucleotide sequences. In order to test all possible combinations, we developed a high-throughput sequencing approach where the activity of known (Panto, Dre, Cre, Vika and VCre) and new Y-SSRs (YR1, YR2, YR4, etc.) can be quantified on all target sites in a single experiment. To accomplish this, we started with a two-step cloning scheme to produce all combinations of 13 recombinases and their respective 13 target sites on 169 (13 × 13) individual vectors. The 13 target sequences were cloned individually into the pEVO vector. The resulting constructs were then pooled and linearized to clone in a pool of the 13 recombinase coding sequences in one ligation reaction (Figure 2A). After an overnight culture and induction of recombinase expression, we retrieved the plasmid DNA and cut out the fragments carrying the recombinase sequence on the 3'-end and target site(s) on the 5'-end. Using the Oxford Nanopore Technologies' long read sequencing platform, we obtained a total of 417 769 reads containing both the specified recombinases and target sites. All possible 169 combinations of Y-SSRs and target sites were identified with a minimum coverage of 224 reads (Supplementary Figure S6A). Using this data, we calculated the recombination rates for the individual recombinases on all the target sites, providing a specificity profile for each recombinase (Figure 2B).

Interestingly, we found a wide spectrum of specificities for the different recombinases (Figure 2B). First, we identified that YR2, YR8, YR9, YR11 and Vika are highly specific recombinases, recombining only their own target sites. The other recombinases showed activity on their own, but also on varying numbers of other target sites. Analyzing this data in detail, we made several interesting observations from the recorded cross-recombination events: (i) Three-way cross-recombination between YR1, YR6 and VCre, can be explained by the high similarity of their target sites (Figure 2C, Supplementary Figure S6B) as well as relatively high protein homology (Supplementary Figure S1). (ii) Two-way cross-recombination between Cre and YR4 can be explained by the similarity of their target sites, since the half-sites of loxP and lox4 differ only at one nucleotide position but, interestingly, the recombinases share only 58% similarity (42% identity, Supplementary Figure S1); (iii) Dre recombinase seems to recombine a variety of target sites ranging from ones very similar to its cognate target site rox (pox, lox8 – 2 mismatches) to lox11 with four mismatches to the rox half site (Figure 2C, Supplementary Figure S6B); (iv) three SSRs (YR4, Panto and Dre) cross-recombine targets, which are quite different to their bone-fide target sites. For example, YR4 is able to recombine lox11 even though, its half-site differs in seven positions from lox4 (Figure 2C,

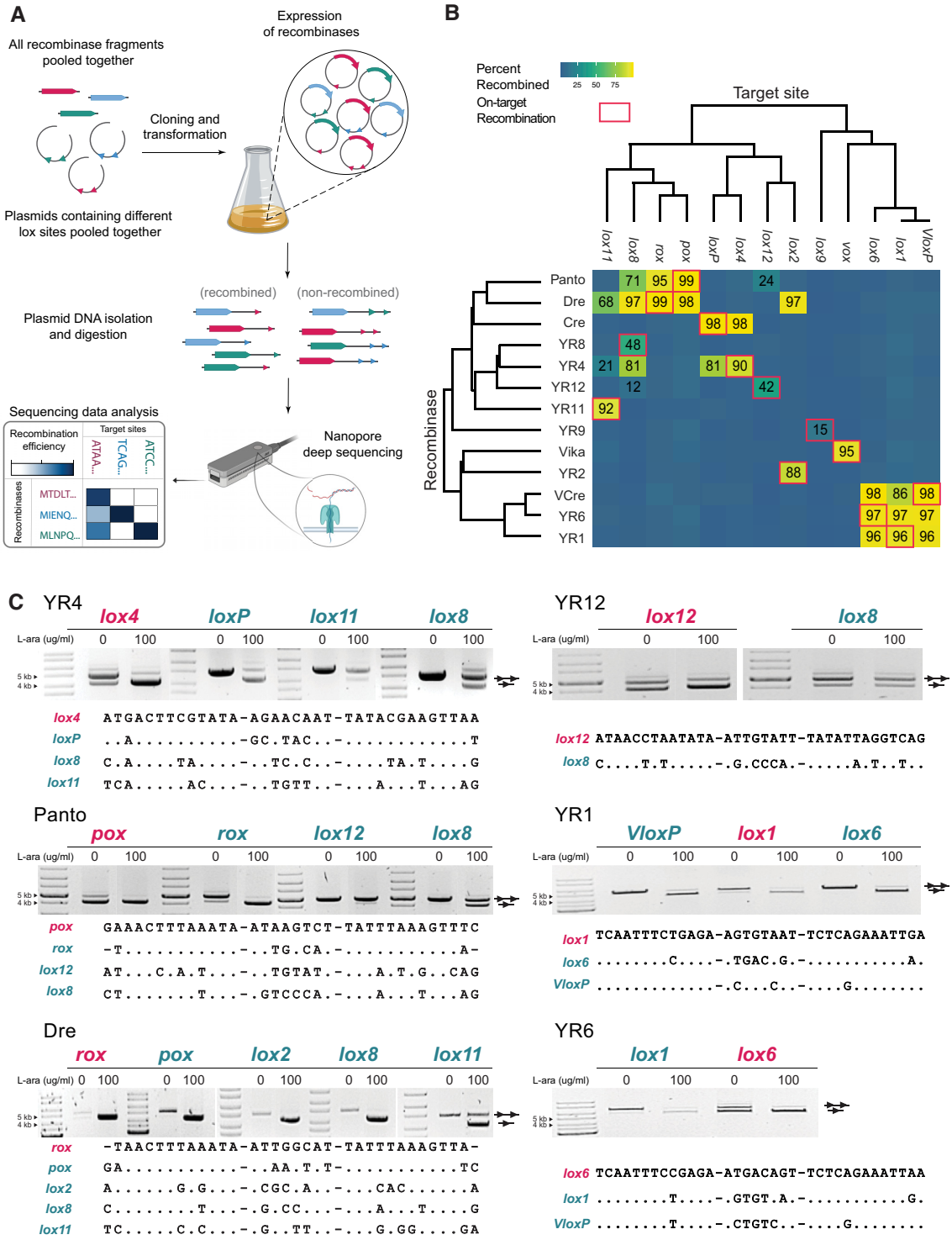


Figure 2. Profiling target-site selectivity of Cre-type recombinases. (**A**) Schematic representation of the experimental workflow. Important steps are indicated by arrows. (**B**) Analysis of deep sequencing results. Cross recombination events are displayed by a heatmap of recombination efficiency for each possible combination. Recombination efficiencies were calculated by dividing the number of recombined reads by the number of total reads for a given combination and expressed as a percentage. For combinations where more than 10% recombination was observed the exact number of recombination percentage is indicated. Target sites are displayed horizontally and ordered based on their similarity. Recombinases tested are aligned based on their homology on the vertical axis. On-target recombination events are boxed with red squares. (**C**) Validation of cross recombination events by plasmid-based recombination assay. Recombination activity of respective recombinases is shown on their on-target sites in magenta, and on off-targets in turquoise. Recombination was assessed by agarose gel electrophoresis. Each sample was tested with or without L-arabinose to induce recombinase expression, indicated with '0' or '100' μg/ml of L-arabinose). Recombination is indicated by the line with the single triangle, whereas a line with two triangles illustrates the non-recombined band. M = GeneRuler™ DNA Ladder Mix (Thermo Fisher). Parts of the figure were created with BioRender.com.

Supplementary Figure S6B). Interestingly, the reverse is not true and YR11 does not show any activity on lox4 (Figure 2B); (v) recombinases that recombine similar target sites have different cross-recombination profiles (e.g. Dre recombines lox11 and lox2, whereas Panto does not, but recombines instead lox12, where Dre and Panto both recombine their respective target sites, rox and pox, respectively). Altogether, these results document the complex relation of recombinase sequences and their activity on target sites, but also serve as guidelines for simultaneous use of these recombinases in complex setups where fine spatial and temporal control of multiple target recombination is needed.

Activity of novel recombinases in human cells

For potential applications in mammalian cells, we tested the activity of the eight new Y-SSRs in a human cell line. For this purpose, we co-transfected HEK293T cells with recombinase expression plasmids alongside recombination reporter plasmids harboring the corresponding target sites (Figure 3A, Supplementary Figure S7A). In the reporter plasmids, the mCherry cassette, driven by a CAG promoter, is flanked by the target (lox) sites. Upon recombination, the mCherry cassette is deleted from the plasmid and the CAG promoter then drives the expression of a GFP cassette (Figure 3A, Supplementary Figure S7A). Hence, the activity of the recombinases on their predicted target sites can be visualized by fluorescent microscopy and quantified by flow cytometry. When co-transfection experiments were analyzed, all of the recombinases tested displayed GFP positive cells (Figure 3B, Supplementary Figure S5E), demonstrating that these recombinases are active in HEK293T cells, whereas no GFP-positive cells were observed when co-transfections were done with an 'empty' expression vector, lacking recombinase coding sequences (Figure 3B). Flow cytometry analyses revealed that in almost all samples more than 90% of the cells that were co-transfected with the recombinase expression plasmid and the reporter (cells that were double, BFP and mCherry positive) were also GFP positive, indicating that these recombinases are highly active in this setting (Figure 3C, Supplementary Figure S7C).

Influence of recombinase expression on cell proliferation

Investigations of SSRs in heterologous hosts are important to define their applied properties. Recombinases could recognize cryptic (pseudo) recombination sites in a genome that might be recombined and lead to off-target effects, potentially resulting in growth arrest or apoptosis. Indeed, active pseudo-loxP sites have been described in the human and mouse genome (52). Consequently, impairment of cell proliferation may occur when Cre is overexpressed in human or mouse cells (53–55). Furthermore, overexpression of any DNA binding protein could lead to impaired regulation of gene expression, which could affect cell growth.

To test for potential effects on cell proliferation of the newly identified Y-SSRs when they are overexpressed, we constructed lentiviral vectors, which allow co-expression of the recombinases and tagBFP (Figure 4A). As controls, we also produced viral particles for overexpression of either Cre, an inactive Cre variant (CreY324F), Vika and tagBFP alone. NIH3T3 cells were then infected and the change in percentage of BFP-positive cells was monitored for 15 days. A reduction in BFP-positive cells over time would indicate a negative effect on cell proliferation due to recombinase overexpression (54,55). Indeed, the number of BFP-positive cells progressively dropped when Cre recombinase was tested in this assay, while the catalytically inactive version of Cre had no effect (Figure 4B). In comparison to Cre, YR4 and YR8 showed a less pronounced decrease in the percentage of BFP positive cells (∼15%; $P < 0.0001$, and ∼17%; $P < 0.0001$, respectively), suggesting that overexpression of these recombinases slightly inhibits cell proliferation (Figure 4B). In contrast, the percentage of BFP-positive cells did not significantly change in cells expressing the other recombinases (Figure 4B), indicating that overexpression of these recombinases is well tolerated in the cells. We conclude that six out of the eight newly identified Y-SSRs can be expressed at high levels without compromising cell proliferation, supporting their utility in mammalian cells.

Prediction of possible recombination target sites in human and mouse genomes

We sought to identify pseudo-lox sites of the investigated SSRs in human and mouse genomes, as these could serve as potential endogenous targets for a variety of genome engineering experiments, but also could lead to unwanted genome rearrangements (52,54). Hence, we searched for highly homologous target sites with unique or same putative spacers to address both possibilities (Figure 5, Supplementary Figure S8 and S9). Criteria that we used were based on sequence homology against putative target sites identified together with the recombinases, filtering for sequences carrying no more than two mismatches per half-site and leaving the spacer region entirely flexible. With the presumption that two target sites could be recombined by a single recombinase only if they possess matching spacer sequences, we searched for target sites with unique spacer to serve as landing pads for integration, while also screening for sites that share identical spacer sequence with at least one other target site to account for potential undesired inter- or intrachromosomal rearrangements.

Our search showed that the pseudo-sites with unique spacers were distributed to all the chromosomes of both genomes (Figure 5A, Supplementary Figure S8, Supplementary Figure S9A). We found pseudo-rox sites of Dre recombinase to be the most common in both of the genomes with 217 in human and 123 target sites in mouse genome (Figure 5B, Supplementary Figure S9B). Out of these, 49 and 22 genomic sequences, respectively, represent the subsets where at least two share the same spacer, whose recombination could lead to potential inter- or intrachromosomal rearrangements. Pseudo-sites with identical spacers, possibly recombined by all other recombinases, were drastically less represented in our search making them more suitable for precise genomic manipulations (Figure 5B, Supplementary Figure S8, Supplementary Figure S9B).

Y-SSRs have the ability to perform recombination bidirectionally, with excision being more thermodynamically

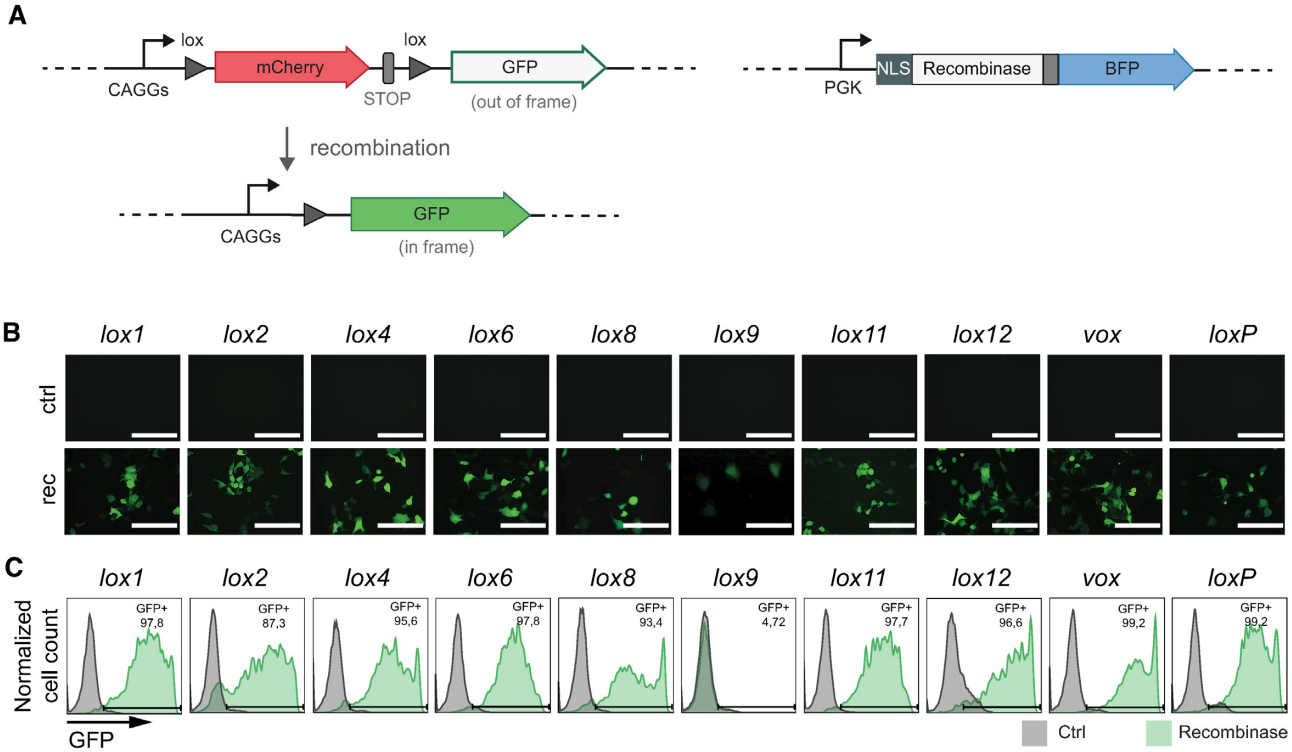


Figure 3. Activity of newly discovered SSRs in mammalian cells. **(A)** Graphical representation of the mammalian recombination reporter and expression constructs. Important features are marked in the reporter and expression vectors. Upon recombination of the reporter vector the mCherry cassette will be excised allowing for the expression of GFP (green) from the pCAG promotor (arrow). Black triangles represent different lox sites for each corresponding recombinase. NLS, nuclear localization signal. (**B**) Fluorescence microscopy analysis in Hek293T cells. Transfected cells with the empty expression plasmids and non-recombined reporter plasmids harboring eight new target sites (top panel), or co-transfection of the reporter with the expression plasmids carrying respective recombinases (lower panel) are shown. Vika/vox and Cre/loxP were included as positive controls. Ctrl, negative controls; Rec, recombinase. (**C**) FACS analysis of the samples shown in (B). Grey histograms depict control samples where non-recombined reporter plasmids were co-transfected with 'empty' expression plasmids, while the green histograms show samples transfected with corresponding recombinases.

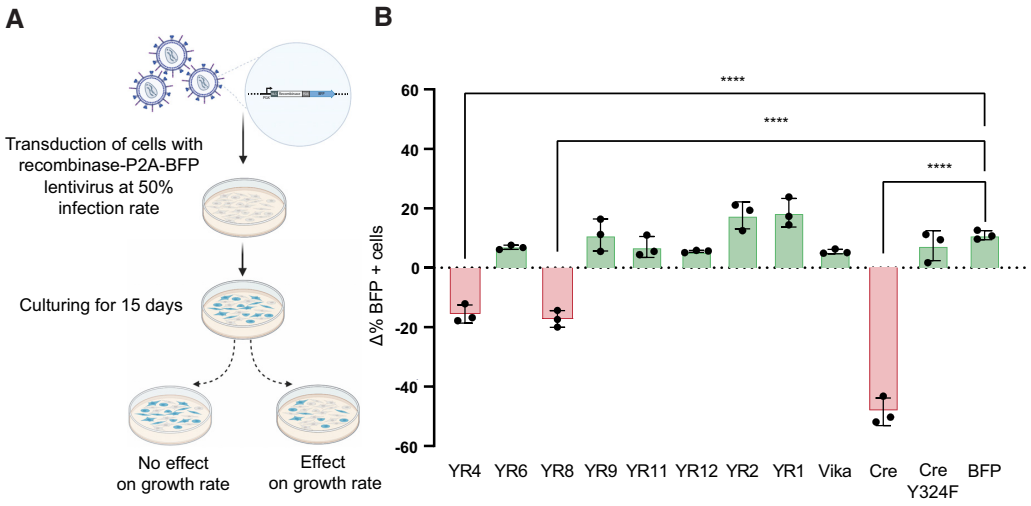


Figure 4. Effect on cell growth upon overexpression of Y-SSRs in mammalian cells. (**A**) Overview of the experimental setup. Important steps are indicated by arrows. Cells were transduced at a rate of ca. 50% with a bicistronic lentivirus expression construct, where the expression of respective recombinases was linked via P2A to BFP expression. Cells were analyzed every 72 h by flow cytometry and the percentage of BFP-positive cells was recorded over the course of 2 weeks. Declining percentages of BFP-positive cells are indicative of a proliferation disadvantage of infected cells. (**B**) Analysis of growth rates. Difference in the percentage of BFP-positive cells between day 3 and day 15 are plotted (biological replicates are shown as dots, $n = 3$). Error bars represent standard deviation of the mean (SD). Statistical significance relative to BFP control was calculated by a one-way ANOVA test. (****): $P \leq 0.0001$. Parts of the figure were created with BioRender.com.

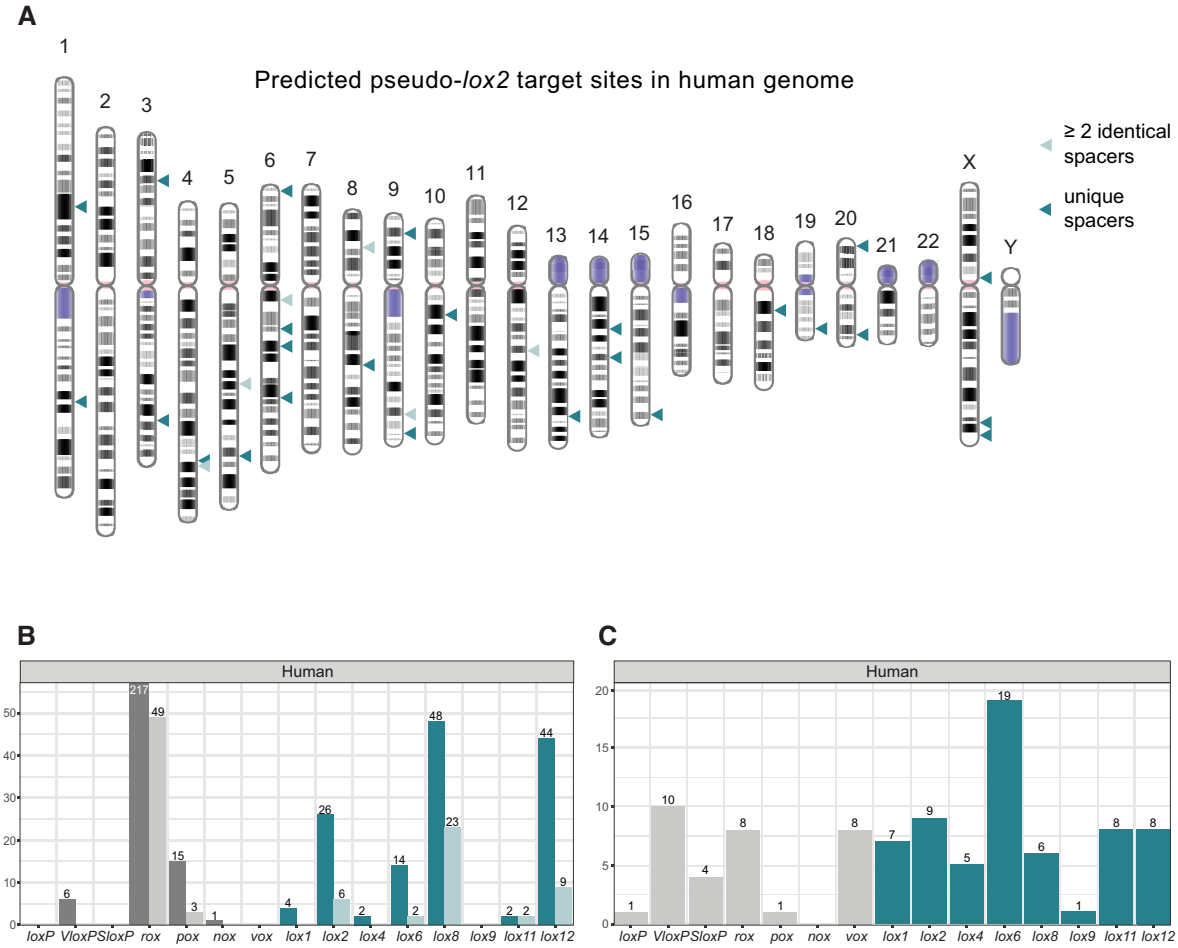


Figure 5. Prediction of pseudo-recombination target sites in human genome. (**A**) Example of chromosomal distribution of human pseudo-lox2 sequences. Positions marked with triangles indicate sequences having no more than two mismatches per half-site of the matching target site sequence. Dark-colored triangles refer to a subset of genomic sequences, where each has a unique spacer (potential integration sites). Light-colored triangles represent a subset of genomic sequences, where at least two share the same spacer (sites for potential inter- or intra-chromosomal rearrangements). (**B**) Counts of human genomic sequences with high similarity, up to two mismatches per half-site, towards already described (grey) and new (colored) Y-SSR target sites. Dark-colored bars refer to a subset of genomic sequences, where each has a unique spacer (potential integration sites). Light-colored bars refer to a subset of genomic sequences, where at least two share the same spacer (sites for potential inter- or intra-chromosomal rearrangements). (**C**) Graphs depicting the number of unique sequences in the human genome with at least one half-site highly similar, with only one mismatch allowed, to target sites of the new (colored) and already described (grey) Y-SSRs.

favored. Hence, their target sites have to be specifically engineered to increase the efficiency of the integration events (56,57). One of the most efficient ways to shift the equilibrium of the reaction towards integration has been shown to be using the RE/LE (RE – right element, LE – left element) strategy (58). Single mutant loxP sites can be modified in opposite half sites to ensure that upon recombination double mutants would be generated that inhibit reverse excision reaction and thereby stabilize the integration product. In order to show the potential utilization of new Y-SSRs for this purpose we also screened the human and mouse genomes for genomic sequences having only one half-site highly similar to the sequence of putative target sites, with only one mismatch allowed (Figure 5C, Supplementary Figure S9C). We found numerous possible entry points with our search, with the pseudo-lox6 site possibly recombined by YR6 being most represented with 19 possible genomic sequences (Figure 5C, Supplementary Figure S9C). By designing the compatible donor target sites multiple specifically targeted transgene insertions could possibly be realized by these Y-SSRs.

DISCUSSION

Diverse microbial organisms and their gene pool represent an almost inexhaustible source for the identification of new DNA editing enzymes (59–62). The discovery and characterization of some of these enzymes has enabled the development of molecular biology techniques for advanced genome engineering. Nevertheless, the multitude of publications in recent years shows that there is a growing need for new and better genome engineering tools. In this study, we performed a bioinformatic-guided genome-wide search to identify novel Cre-type recombinases and their associated target sequences. Based on our bioinformatic search, we identified over 500 putative Y-SSRs candidates of which a selection of 17 candidates were experimentally tested and

8 novel Y-SSRs systems were molecularly characterized in depth. The results show that the pipeline is able to predict new Y-SSRs and their target sites at a success rate of about 50%, indicating that many more active Y-SSR systems should be retrievable from the list. The success of prediction still proves to be largely limited by the nomination of the appropriate target site. Our search was focused on scanning regions 1kb upstream and downstream of the putative recombinase gene as this genomic organization was reported for previously described phage-related Y-SSRs (7,32,43,63). Nevertheless, some of these SSRs come from temperate phages that integrated their genomes into bacterial chromosomes and commonly undergo a complex decay process consisting of inactivating point mutations, genome rearrangements, modular exchanges, invasion by further mobile DNA elements, and DNA deletions. These events could cause the disturbance of the expected organization and lead to misplacement of the native target sites. Given the various mechanisms involved and the diversity within the Y-SSR family, our search criteria were designed to increase the likelihood of successful identification, but also limited the number of putative recombinases that we characterized. Several recently reported methods for identification of prophage elements in bacterial genomes may improve the annotation of new Y-SSRs and their target sites in the future (42,62,64).

To optimally use different Y-SSR systems, it is essential to characterize their applied properties. The determination of the *in vivo* recombinase activity may be crucial for planning an experiment. In this study, the eight newly characterized YRs showed varying activity on their predicted target sites in bacteria, from over 90% in the case of YR1, YR4, YR6 to as low as around 15% in the case of YR8 and YR9. The observed low activity of the latter is possibly due to suboptimal expression of the protein (Supplementary Figure S4B, C), but the need for additional cofactors or optimal temperature for more efficient recombination of the target sites could also play a role. Furthermore, the annotated sequencing sample may contain errors for the target site or for the recombinase, or the recombinase originated from a prophage that accumulated mutations over time, influencing the activity of the enzyme. Because we could quickly improve the activity of YR9 recombinase by directed molecular evolution, the later explanation is the most likely reason for the weak activity we originally observed. Nevertheless, Y-SSRs that have lower activity could be beneficial in a setting where multiple sequential DNA rearrangements are needed.

We profiled the specificity of Cre-type recombinases providing a valuable overview of possible cross recombination events on all target sites. We found that five recombinases showed high specificity on the tested target sites. Hence, these recombinases likely represent good candidates for experiments where high specificity of recombination is desired. On the other hand, different cross recombination properties described in this work could have important implications for the design of various complex experiments requiring the use of multiple Y-SSRs. Interestingly, our results reveal that the correlation between amino acid sequence homology of the Y-SSRs and nucleotide sequence homology of their target sites that they recombine is not straightforward. In addition, cross recombination is not always reciprocal (e.g. YR4 and YR11 as well as YR11, YR2, YR8 and Dre, Figure 2B), suggesting that there are yet unknown determinants that render these recombinases more or less tolerant to sequence variations of the target sites. This data should be useful to investigate key specificity determinants of these enzymes. Indeed, in previous work, detailed comparative analyses of Y-SSRs have identified several amino acids as key players in target site distinction. For example, it was shown that K43, R259, and G263 of Cre are critical residues for the discrimination between the loxP and rox sites (33), and that nearest-neighbor amino acids of these residues influence the activity of Cre-type recombinases (65). Our collection of Cre-like Y-SSRs with characterized specificities, coupled with additional structural data, has the potential to serve as a valuable resource for better understanding of the target site recognition of Y-SSRs. This knowledge could, in turn, accelerate the engineering of new designer recombinases with desired specificities and functionalities.

For their applied use in higher organisms, we tested the new Y-SSRs for their activity and compatibility in mammalian cells. All the recombinases except for YR9 recombinase showed high activity in a plasmid-based assay in HEK293T cells. Interestingly, YR8 showed high recombination rates in mammalian cells, whereas this recombinase had only weak activity in bacteria, suggesting that activity profiles of recombinases can vary in heterologous hosts. For the applied use in heterologous cells, it is also important to consider potential adverse effects they could cause in cells not just by off-target recombination on pseudo-sites that may exist in the host genome, but also by other mechanisms such as a potential impact on DNA replication or gene expression. This could lead to impaired growth of the cells. We did not observe this effect for most recombinases upon overexpression, but some illegitimate recombination events could also lead to 'silent' effects that do not affect cell growth. We therefore bioinformatically screened the human and mouse genome for lox-like sites for all known Cre-type recombinases. This information is useful in two ways: i) It provides an estimation on potential off-target sites that could potentially compromise an experiment. ii) it delivers potential endogenous target sites that might be usefully employed for genome engineering exercises (for instance for targeted delivery of DNA cargo into a safe harbor locus). It would be interesting to address the activity of these recombinases on these sites, which could then potentially serve as entry point for multiple genome manipulation strategies.

Finally, the discovery and characterization of different naturally occurring recombinases should be useful to accelerate the development of novel enzymes via directed evolution. It has been demonstrated that shuffling of related genes can speed up the evolution process (66). Hence, new recombinases with desired properties might become available in shorter time through family shuffling. In addition, by using the wide variety of new naturally occurring recombinases with activity on various target sites, one could start different evolution campaigns depending on similarity to the desired target site. While the present work aims to provide a more diverse collection of starting points for recombinase optimization, the lack of molecular mechanisms

underlying their specificities and target site recognition limits rational design approaches. It would therefore be helpful to obtain additional structural data for these new Y-SSR-systems to gain insights into key structural features and amino acid residues that contribute to target site recognition and specificity.

Furthermore, this data could also improve machine learning algorithms developed to predict recombinase activity on desired target sites (67). By using data from evolved Y-SSRs, in combination with data from naturally occurring SSR systems, we expect that a smarter design of synthetic SSRs can be generated to rapidly target novel loci in desired genomes.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the members of Buchholz group for fruitful discussions and valuable advice they provided.

FUNDING

European Union [ERC 742133, H2020 UPGRADE825825]; German Research Council [DFG BU 1400/7-1 and PI 600/4-1]; Bundesministerium für Bildung und Forschung GO-Bio (BMBFGO-Bio) [031B0633]. Funding for open access charge: TU Dresden.

Conflict of interest statement. Technical University (Technische Univeristät) Dresden has filed a patent application based on this work, in which M.J., L.T.S. and F.B. are listed as inventors.

REFERENCES

1. Meinke,G., Bohm,A., Hauber,J., Pisabarro,M.T. and Buchholz,F. (2016) Cre recombinase and other tyrosine recombinases. *Chem. Rev.*, **116**, 12785–12820.
2. Komor,A.C., Kim,Y.B., Packer,M.S., Zuris,J.A. and Liu,D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.
3. Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.
4. Anzalone,A.V., Gao,X.D., Podracky,C.J., Nelson,A.T., Koblan,L.W., Raguram,A., Levy,J.M., Mercer,J.A.M. and Liu,D.R. (2022) Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.*, **40**, 731–740.
5. Anzalone,A.V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
6. Sauer,B. and Henderson,N. (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci.*, **85**, 5166–5170.
7. Sternberg,N. and Hamilton,D. (1981) Bacteriophage P1 site-specific recombination I. Recombination between loxP sites. *J. Mol. Biol.*, **150**, 467–486.
8. Duyne,G.D.V. (2001) A structural view of Cre- loxP site-specific recombination. *Annu. Rev. Biophys. Biom.*, **30**, 87–104.
9. Anderson,R.P., Voziyanova,E. and Voziyanov,Y. (2012) Flp and Cre expressed from Flp–2A–Cre and Flp–IRES–Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange. *Nucleic Acids Res.*, **40**, e62.
10. Minorikawa,S. and Nakayama,M. (2011) Recombinase-mediated cassette exchange (RMCE) and BAC engineering via VCre/VloxP and SCre/SloxP systems. *BioTechniques*, **50**, 235–246.
11. Feil,R. (2007) Conditional mutagenesis: an approach to disease models. *Handb. Exp. Pharmacol.*, **178**, 3–28.
12. Justice,M.J., Siracusa,L.D. and Stewart,A.F. (2011) Technical approaches for mouse models of human disease. *Dis. Model. Mech.*, **4**, 305–310.
13. Weng,W., Liu,X., Lui,K.O. and Zhou,B. (2021) Harnessing orthogonal recombinases to decipher cell fate with enhanced precision. *Trends Cell Biol.*, **32**, 324–337.
14. Liu,K., Jin,H. and Zhou,B. (2020) Genetic lineage tracing with multiple DNA recombinases: a user's guide for conducting more precise cell fate mapping studies. *J. Biol. Chem.*, **295**, 6413–6424.
15. Jeong,J., Kim,T.H., Kim,M., Jung,Y.K., Kim,K.S., Shim,S., Jang,H., Jang,W.I., Lee,S.B. and Choi,D. (2022) Elimination of reprogramming transgenes facilitates the differentiation of induced pluripotent stem cells into hepatocyte-like cells and hepatic organoids. *Biology*, **11**, 493.
16. Jin,H., Liu,K. and Zhou,B. (2021) Dual recombinases-based genetic lineage tracing for stem cell research with enhanced precision. *Sci. China Life Sci.*, **64**, 2060–2072.
17. Wang,H., He,L., Li,Y., Pu,W., Zhang,S., Han,X., Lui,K.O. and Zhou,B. (2022) Dual Cre and Dre recombinases mediate synchronized lineage tracing and cell subset ablation in vivo. *J. Biol. Chem.*, **298**, 101965.
18. Karimova,M., Baker,O., Camgoz,A., Naumann,R., Buchholz,F. and Anastassiadis,K. (2018) A single reporter mouse line for Vika, Flp, Dre, and Cre-recombination. *Sci. Rep-uk*, **8**, 14453.
19. Liu,K., Yu,W., Tang,M., Tang,J., Liu,X., Liu,Q., Li,Y., He,L., Zhang,L., Evans,S.M. *et al.* (2018) A dual genetic tracing system identifies diverse and dynamic origins of cardiac valve mesenchyme. *Development*, **145**, dev167775.
20. Voziyanova,E., Anderson,R.P., Shah,R., Li,F. and Voziyanov,Y. (2016) Efficient genome manipulation by variants of site-specific recombinases R and TD. *J. Mol. Biol.*, **428**, 990–1003.
21. Duplus-Bottin,H., Spichty,M., Triqueneaux,G., Place,C., Mangeot,P.E., Ohlmann,T., Vittoz,F. and Yvert,G. (2021) A single-chain and fast-responding light-inducible Cre recombinase as a novel optogenetic switch. *Elife*, **10**, e61268.
22. Takao,T., Yamada,D. and Takarada,T. (2022) Mouse model for optogenetic genome engineering. *Acta Med. Okayama*, **76**, 1–5.
23. Siuti,P., Yazbek,J. and Lu,T.K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.*, **31**, 448–452.
24. Ba,F., Liu,Y., Liu,W.-Q., Tian,X. and Li,J. (2022) SYMBIOSIS: synthetic manipulable biobricks via orthogonal serine integrase systems. *Nucleic Acids Res.*, **50**, 2973–2985.
25. Buchholz,F. and Stewart,A.F. (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.*, **19**, 1047–1052.
26. Abi-Ghanem,J., Chusainow,J., Karimova,M., Spiegel,C., Hofmann-Sieber,H., Hauber,J., Buchholz,F. and Pisabarro,M.T. (2012) Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach. *Nucleic Acids Res.*, **41**, 2394–2403.
27. Santoro,S.W. and Schultz,P.G. (2002) Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 4185–4190.
28. Karpinski,J., Hauber,I., Chemnitz,J., Schäfer,C., Paszkowski-Rogacz,M., Chakraborty,D., Beschorner,N., Hofmann-Sieber,H., Lange,U.C., Grundhoff,A. *et al.* (2016) Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat. Biotechnol.*, **34**, 401–409.
29. Sarkar,I., Hauber,I., Hauber,J. and Buchholz,F. (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science*, **316**, 1912–1915.

30. Lansing,F., Mukhametzyanova,L., Rojo-Romanos,T., Iwasawa,K., Kimura,M., Paszkowski-Rogacz,M., Karpinski,J., Grass,T., Sonntag,J., Schneider,P.M. *et al.* (2022) Correction of a Factor VIII genomic inversion with designer-recombinases. *Nat. Commun.*, **13**, 422.
31. Lansing,F., Paszkowski-Rogacz,M., Schmitt,L.T., Martin Schneider,P., Romanos,T.R., Sonntag,J. and Buchholz,F. (2019) A heterodimer of evolved designer-recombinases precisely excises a human genomic DNA locus. *Nucleic Acids Res.*, **48**, 472–485.
32. Karimova,M., Abi-Ghanem,J., Berger,N., Surendranath,V., Pisabarro,M.T. and Buchholz,F. (2012) Vika/vox, a novel efficient and specific Cre/loxP-like site-specific recombination system. *Nucleic Acids Res.*, **41**, e37.
33. Karimova,M., Splith,V., Karpinski,J., Pisabarro,M.T. and Buchholz,F. (2016) Discovery of Nigri/nox and Panto/pox site-specific recombinase systems facilitates advanced genome engineering. *Sci. Rep.*, **6**, 30130.
34. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
35. Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.
36. Prüfer,K., Stenzel,U., Dannemann,M., Green,R.E., Lachmann,M. and Kelso,J. (2008) PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics*, **24**, 1530–1531.
37. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
38. Sürün,D., Schneider,A., Mircetic,J., Neumann,K., Lansing,F., Paszkowski-Rogacz,M., Hänchen,V., Lee-Kirsch,M.A. and Buchholz,F. (2020) Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. *Genes*, **11**, 511.
39. Anastassiadis,K., Fu,J., Patsch,C., Hu,S., Weidlich,S., Duerschke,K., Buchholz,F., Edenhofer,F. and Stewart,A.F. (2009) Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. *Dis. Model Mech.*, **2**, 508–515.
40. Suzuki,E. and Nakayama,M. (2011) VCre/VloxP and SCre/SloxP: new site-specific recombination systems for genome engineering. *Nucleic Acids Res.*, **39**, e49–e49.
41. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
42. Houdt,R.V., Leplae,R., Lima-Mendez,G., Mergeay,M. and Toussaint,A. (2012) Towards a more accurate annotation of tyrosine-based site-specific recombinases in bacterial genomes. *Mobile DNA*, **3**, 6.
43. Casjens,S. (2003) Prophages and bacterial genomics: what have we learned so far?*Mol. Microbiol.*, **49**, 277–300.
44. Wang,Z., Xiong,G. and Lutz,F. (1995) Site-specific integration of the phage ΦCTX genome into the *Pseudomonas aeruginosa* chromosome: characterization of the functional integrase gene located close to and upstream of attP. *Mol. Gen. Genet.*, **246**, 72–79.
45. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
46. Sheets,M.B., Wong,W.W. and Dunlop,M.J. (2020) Light-inducible recombinases for bacterial optogenetics. *Acs Synth. Biol.*, **9**, 227–235.
47. Merrick,C.A., Zhao,J. and Rosser,S.J. (2018) Serine integrases: advancing synthetic biology. *Acs Synth. Biol.*, **7**, 299–310.
48. Livet,J., Weissman,T.A., Kang,H., Draft,R.W., Lu,J., Bennis,R.A., Sanes,J.R. and Lichtman,J.W. (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, **450**, 56–62.
49. Snippert,H.J., Flier,L.G.v., Sato,T., Es,J.H.v., Born,M.v., Kroon-Veenboer,C., Barker,N., Klein,A.M., Rheenen,J.v., Simons,B.D. *et al.* (2010) Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell*, **143**, 134–144.
50. Fenno,L.E., Mattis,J., Ramakrishnan,C., Hyun,M., Lee,S.Y., He,M., Tucciarone,J., Selimbeyoglu,A., Berndt,A., Grosenick,L. *et al.* (2014) Targeting cells with single vectors using multiple-feature Boolean logic. *Nat. Methods*, **11**, 763–772.
51. Weinberg,B.H., Pham,N.T.H., Caraballo,L.D., Lozanoski,T., Engel,A., Bhatia,S. and Wong,W.W. (2017) Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.*, **35**, 453–462.
52. Thyagarajan,B., Guimarães,M.J., Groth,A.C. and Calos,M.P. (2000) Mammalian genomes contain active recombinase recognition sites. *Gene*, **244**, 47–54.
53. Loonstra,A., Vooijs,M., Beverloo,H.B., Allak,B.A., Drunen,E.v., Kanaar,R., Berns,A. and Jonkers,J. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 9209–9214.
54. Schmidt,E.E., Taylor,D.S., Prigge,J.R., Barnett,S. and Capecchi,M.R. (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 13702–13707.
55. Pugach,E.K., Richmond,P.A., Azofeifa,J.G., Dowell,R.D. and Leinwand,L.A. (2015) Prolonged Cre expression driven by the α-myosin heavy chain promoter can be cardiotoxic. *J. Mol. Cell Cardiol.*, **86**, 54–61.
56. Araki,K., Araki,M. and Yamamura,K.-i. (1997) Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nucleic Acids Res.*, **25**, 868–872.
57. Araki,K., Okada,Y., Araki,M. and Yamamura,K. (2010) Comparative analysis of right element mutant lox sites on recombination efficiency in embryonic stem cells. *BMC Biotechnol.*, **10**, 29.
58. Thomson,J.G., Rucker,E.B. and Piedrahita,J.A. (2003) Mutational analysis of loxP sites for efficient Cre-mediated insertion into genomic DNA. *Genesis*, **36**, 162–167.
59. Pausch,P., Al-Shayeb,B., Bisom-Rapp,E., Tsuchida,C.A., Li,Z., Cress,B.F., Knott,G.J., Jacobsen,S.E., Banfield,J.F. and Doudna,J.A. (2020) CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science*, **369**, 333–337.
60. Harrington,L.B., Burstein,D., Chen,J.S., Paez-Espino,D., Ma,E., Witte,I.P., Cofsky,J.C., Kyrpides,N.C., Banfield,J.F. and Doudna,J.A. (2018) Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*, **362**, 839–842.
61. Karvelis,T., Druteika,G., Bigelyte,G., Budre,K., Zedaveinyte,R., Silanskas,A., Kazlauskas,D., Venclovas,Č. and Siksnys,V. (2021) Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*, **599**, 692–696.
62. Durrant,M.G., Fanton,A., Tycko,J., Hinks,M., Chandrasekaran,S.S., Perry,N.T., Schaepe,J., Du,P.P., Lotfy,P., Bassik,M.C. *et al.* (2022) Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat. Biotechnol.*, **41**, 488–499.
63. Casjens,S.R. and Hendrix,R.W. (2015) Bacteriophage lambda: early pioneer and still relevant. *Virology*, **479**, 310–330.
64. Smyshlyaev,G., Bateman,A. and Barabas,O. (2021) Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol. Syst. Biol.*, **17**, e9880.
65. Soni,A., Augsburg,M., Buchholz,F. and Pisabarro,M.T. (2020) Nearest-neighbor amino acids of specificity-determining residues influence the activity of engineered Cre-type recombinases. *Sci. Rep.*, **10**, 13985.
66. Crameri,A., Raillard,S.-A., Bermudez,E. and Stemmer,W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, **391**, 288–291.
67. Schmitt,L.T., Paszkowski-Rogacz,M., Jug,F. and Buchholz,F. (2022) Prediction of designer-recombinases for DNA editing with generative deep learning. *Nat. Commun.*, **13**, 7966.