Despite rapid advances in genome engineering technologies, inserting genes into precise locations in the human genome remains an outstanding problem. It has been suggested that site-specific recombinases can be adapted towards use as transgene delivery vectors. The specificity of recombinases can be altered either with directed evolution or via fusions to modular DNA-binding domains. Unfortunately, both wild-type and altered variants often have detectable activities at off-target sites. Here we use bacterial selections to identify mutations in the dimerization surface of Cre recombinase (R32V, R32M and 303GVSdup) that improve the accuracy of recombination. The mutants are functional in bacteria, in human cells and in vitro (except for 303GVSdup, which we did not purify), and have improved selectivity against both model off-target sites and the entire E. coli genome. We propose that destabilizing binding cooperativity may be a general strategy for improving the accuracy of dimeric DNA-binding proteins.
S

afe delivery of transgenes into the human genome remains an open problem of critical importance to clinical genetics. Many existing technologies have major limitations. For instance, retroviruses, lentiviruses and transposons integrate non-specifically and can therefore cause cancer by mutagenesis1,2. Transgenes can also be integrated using the endogenous homologous repair pathways, although this process must be stimulated by generating double-stranded breaks at the target site using programmable nuclease technologies, such as meganucleases3, zinc finger nucleases4, TALE nucleases5,6 or the RNA-guided Cas9 protein7,8. This technique is limited by the fact that homologous recombination in humans is less efficient than the competing mutagenic non-homologous end-joining pathway9,10. Site-specific recombinases, which catalyse recombination at precise sites, have properties that make them promising candidates for use as safe gene-delivery vectors. For instance, many do not require host-encoded factors11. In addition, the size of the integrated cassette is less restricted than for other methods. Recombinases’ specificities can be altered either by direction evolution or by fusion to modular DNA-binding domains12–20. Unfortunately, many reprogrammed variants are promiscuous in their activity. This problem is not restricted to artificial variants, as activity at off-target human genomic loci has been reported for some wild-type (WT) recombinases21–24. If recombinases are to vary the ratio of specific to non-specific DNA–protein interactions25. Although powerful, this approach can be inconvenient if the goal is to generate variants of a protein with different specificities: a specificity change alters the DNA–protein interaction, requiring re-optimization of accuracy. Therefore, there is a need for methods to systematically enhance accuracy without changing the DNA–protein interface.

In this work, we attempt to discern such principles using Cre recombinase of the phage P1 as a model system. Cre catalyses a reversible, directional recombination between two 34 bp loxP sites that consist of a pair of 13 bp inverted repeats flanking asymmetrically spaced spacer26–28. Mutagenic studies of loxP have shown that many mutations have non-catastrophic effects on recombination efficiency29–31. Using a theoretical model, we predict that reducing the cooperativity of binding should increase accuracy. We mutagenize a region involved in the formation of Cre dimers and perform bacterial selections for functional and accurate mutants. We isolate three mutants, all of which were able to recombine loxP sites with high efficiency and exhibited improved accuracy with respect to both model off-target sites and the entire Escherichia coli genome.

Results
A theoretical model of DNA-binding accuracy. Under the currently accepted mechanism of Cre recombination, the binding of a Cre monomer to one half of a loxP site is followed by the formation of an asymmetrical homodimer when a second Cre molecule binds to the other half of loxP. Next, the two loxP-bound dimers associate to form a tetramer, and recombination proceeds via a Holiday Junction intermediate28. We reasoned that the formation of the dimer of dimers is not site specific in the sense that it involves no new DNA-binding events, leading us to conclude that the precision of dimer formation determines the accuracy of recombination. Assuming that the protein–protein affinity is negligible in the absence of DNA, dimer formation on target sites is described with:

\[ P + D \xrightarrow{2K_D} P \cdot D \]  
\[ P + P \cdot D \xrightarrow{K_{on}} 2P \cdot D \]

where \( P \) is the unbound protein monomer, \( D \) is the full DNA-binding site, \( K \) is the affinity of each monomer for half of the binding site and \( K_{dim} \) is the affinity of the protein dimer for the full binding site. If it is assumed that the cooperative energy is sequence-independent then \( K_{dim} = K_{coop} \). where \( K_{coop} \) is the protein–protein affinity. A competing set of binding events occurs between off-target DNA and the protein:

\[ P + D_{OT} \xrightarrow{2K_D} P \cdot D_{OT} \]
\[ P + P \cdot D_{OT} \xrightarrow{K_{on}} 2P \cdot D_{OT} \]

where \( D_{OT} \) is the off-target DNA concentration and \( K_{OT} \) is the affinity of the protein for off-target DNA. Accuracy can be defined as the ratio of on-target protein dimers to off-target protein dimers:

\[ A = \frac{[2P \cdot D]}{[2P \cdot D_{OT}] = \left( \frac{K_{OT}}{K_D} \right)^2 \cdot \left( \frac{D}{D_{OT}} \right)  \]

The free DNA concentration \( [D] \) is related to the total DNA concentrations \([D_{total}]\) via:

\[ [D_{total}] = [D] + [P \cdot D] + [2P \cdot D] \]
\[ = [D] \left( 2K_{coop}K_D^2[P] + 2K_P + 1 \right) \]

An analogous expression relates \([D_{OT}]\) and \([D_{OT,total}]\). Equation (5) can be expressed in terms of total DNA concentrations:

\[ A = \left( \frac{K_{OT}}{K_D} \right)^2 \cdot \left( \frac{[D_{total}]}{[D_{OT,total}] - 2K_{coop}K_D^2[P] + 2K_P + 1} \right) \]

To model accuracy of Cre in E. coli, we used the in vitro affinity coefficients \( K_{D} = 1.5 \times 10^3 \) M\(^{-1}\) and \( K_{coop} = 1.7 \times 10^3 \) M\(^{-1}\), with both values obtained from previous in vitro measurements32. We assumed that \( K_{OT} = 10^4 \), which is in the same order of magnitude as the experimentally determined \( K_{OT} \) of EcoRV and BamHI33. Assuming a single target site in an E. coli cell of a 0.5 μm radius gives a \([D_{total}]\) of 3.2 × 10\(^{-9}\) M. If off-target sites exist at 1 bp windows along both strands of the 4.6 Mbp E. coli genome, then \([D_{OT,total}]\) is \( \sim 3.2 \times 10^{-9}\) M × 2 × 9.2 × 10^6 bp = 2.9 × 10^{-2} M. This value of \([D_{OT,total}]\) may be an overestimate owing to competitive binding of other DNA-associated proteins to the genome.

We plotted the predicted accuracy of dimer formation as a function of total protein concentration for both WT Cre and for mutants with reduced cooperativity (Fig. 1). The model predicts that accuracy increases with both a reduction in cooperativity and a decrease in protein expression levels. This is an intuitive result, because accuracy should increase as the two monomer binding events become more independent from each other. We conclude that although a reduction in cooperative binding will affect both target and off-target binding, off-target binding will be destabilized to a greater degree.

Identifying candidate mutations using bacterial selection. Our theoretical model predicted that accuracy could be improved by decreasing the cooperative binding moment. We therefore targeted our mutagenesis towards a domain directly involved in the dimer interaction but distant from the Cre–DNA interaction: the α-helix closest to the amino terminus34. To find mutations that improve accuracy while maintaining proper function with respect to loxP, we performed two rounds of bacterial selection. The first round was designed to identify functional mutants, whereas the second round selected accurate mutants. To select functional mutants, we used a resistance marker flanked by loxP sites in an inverted orientation relative to each other. The reading frame was inverted with respect to the promoter, such that Cre-mediated inversion would result in gain of antibiotic resistance.
An out-of-frame toxic *ccdB* was used to apply a selective pressure against inaccurate recombination reactions that produced frameshifts (Fig. 2a). To minimize false negatives due to reversal of the inversion, we placed the selection cassette on a high copy plasmid. The selection resulted in the recovery of 1,690 library-transformed colonies, which corresponds to 38% of the total transformation efficiency. The library produced fewer clones than the positive WT control, suggesting that the selection is functional (Fig. 2b).

To identify accurate constructs, we first found an off-target site to serve as bait in the negative selection. To achieve a high selective pressure, we wanted a bait sequence that would be recombined with a high efficiency. At the same time, we wanted to select for improved accuracy across the entire protein–DNA interaction and, therefore, we wanted a bait sequence with little similarity to *loxP*. We found such a site by performing a selection for pseudo-*loxP* sites and characterizing their *in vitro* recombination efficiency. The site, which we named *loxBait*, is recombined with 37% the efficiency of *loxP*, despite differences in 9 out of 13 bases within a single inverted repeat (Fig. 3a,b).

We performed the counterselection by flanking an in-frame antibiotic resistance marker with *loxBait* and *loxP* oriented in the same direction. The toxic *ccdB* gene was placed at the 3'0-end of *loxP*. Cre-mediated excision would result in both loss of the resistance marker and expression of the toxic gene (Fig. 2c). We subjected the expression plasmids recovered from the positive selection to one round of counterselection. In this case, the catalytically inactive mutant, Y324F, served as the control for growth without selection (Fig. 2d).

We randomly isolated two of the recovered mutants, R32V and R32M, for further characterization. R32 is involved in an intermonomer salt bridge with E69; thus, its disruption in the two mutants can be expected to reduce the protein–protein affinity (Fig. 4a). Two WT colonies survived the counterselection, one of which contained a clone with a de novo duplication of residues 303–305 (303GVSdup). This region is a loop that makes close contact to the other monomer in the dimer structure (Fig. 4b).

**Mutants better discriminate a model off-target site.** We measured the activity of Cre and the isolated mutants on *loxP* and *ψlox h7q21*, a known human off-target site21, using a model predicting an increase in Cre recombinase dimer binding accuracy as cooperativity decreases. The solid line indicates the accuracy predicted for WT Cre, whereas the dashed lines correspond to the expected accuracy of mutants in which the cooperativity has been reduced by the indicated amount.

![Figure 1](image-url)

**Figure 1 | Model predicting an increase in Cre recombinase dimer binding accuracy as cooperativity decreases.** The solid line indicates the accuracy predicted for WT Cre, whereas the dashed lines correspond to the expected accuracy of mutants in which the cooperativity has been reduced by the indicated amount.

![Figure 2](image-url)

**Figure 2 | Selecting functional and accurate Cre variants.** (a) The substrate used to select for functional variants; proper recombination would place the ampicillin resistance gene (*ampR*) under the lac promoter (P_{lac}) conferring resistance. The *ccdB* gene is crossed out to indicate that it’s out of frame with respect to the *ampR* start codon. (b) Ratio of ampicillin-resistant colonies to total colonies isolated following the positive selection. (c) The substrate used to select for accurate variants—recombination of *loxP* and *loxBait* sites would result in loss of ampicillin resistance and would place the toxic *ccdB* gene in frame with the promoter. (d) Number of ampicillin-resistant colonies recovered from the negative selection.
plasmid-based inversion assay. As in the selections, the proteins were expressed from the P\text{bad} promoter. As our theoretical model suggests that greater accuracy is achieved at low protein expression levels, we grew the cells on repressive LB/glucose medium. Indistinguishable results were achieved with growth on LB in the absence of glucose. To test the mutants for improved accuracy, we analysed their ability to recombine a known human pseudo-loxP site \(\psi Lox\ h7q21\) and \(\psi Core\ h7q21\) (which consists of inverted repeats from \(loxBait\) but a spacer that matches \(\psi Lox\ h7q21\)).

Following recombination, substrate plasmids were isolated and digested in a way that resulted in the substrate producing a 631-bp band and its inverted product producing a 386-bp band (Fig. 5a). Y324F produced only the 631-bp parent product, whereas R32V, R32M and 303GVSdup resulted in both inverted and parental plasmids (Fig. 5b). WT unexpectedly produced the 898-bp band that corresponds to a deletion; there were no detectable bands for the inverted and parental forms. We also noted that the WT produced a 1,795-bp band and its inverted product producing a 386-bp band. This product could be ignored for Y324F, R32V, R32M and 303GVSdup, because the cutting efficiency of ScaI digests also produced a 2,573-bp product that is formed via an intermolecular insertion followed by inversion between recombination sites that originated from different molecules (Supplementary Fig. S1). We excluded this product from our analysis.

Achieving equilibrium in the inversion assay should result in 50% inversion. Recombination of two \(loxBait\) sites by R32V, R32M and 303GVSdup resulted in \(~50\)% inversion and no detectable deletion products. In contrast, WT resulted in deletion in 98.7% (95% confidence interval (CI): 97.5–100%) of plasmids (Fig. 5c). Unlike inversions, which can be reversed, excisions are selected for in a dividing cell population, because the excised products cannot replicate. Therefore, the data should not be taken to mean that WT incorrectly excises nearly 100% of the time, but rather that the vast majority of substrate plasmids experienced at least one excision event during the 12 h of growth. Recombination by WT of pseudo-loxP sites \(\psi Lox\ h7q21\) and \(\psi Core\ h7q21\) resulted in improper excision with a frequency of 97.5% (95% CI: 96.5–98.4%). R32V, R32M and 303GVSdup produced no detectable recombination products (Fig. 5d). In aggregate, the data provided evidence that the isolated

Figure 3 | Activity and nomenclature of pseudo-loxP sites. (a) Linear fragments with \(loxBait\) on one end and the indicated sites on the other were treated with Cre and the products were quantified on an agarose gel. All new bands were counted towards the recombination efficiency. No recombination was observed for any of the sites in the absence of Cre. Bolded positions correspond to differences from \(loxBait\). Sites are listed in order of similarity with \(loxBait\). Asterisks mark sequences generated randomly; all others were obtained from a selection for functional pseudo-loxP sites. The box indicates \(loxBait\). (b) The names and sequences of recombination sites used in this study. Positions that are different from \(loxBait\) are shown in bold. Error bars correspond to 95% CI (\(n=\) two to three experimental replicates).
mutants are better able to distinguish on-target and off-target sites than WT.

Mutants are functional in vitro and in human cells. Given that in vivo recombination between two loxP sites reached equilibrium (~100% deletion for WT and ~50% inversion for R32V, R32M and 303GVSdup), there are two explanations for the improved selectivity against pseudo-loxP sites observed for the mutant recombinases; either the mutants were better able than WT to discriminate against pseudo-loxP sites or the mutants were equally less efficient at recombining loxP and pseudo-loxP sites than WT, but the high enzyme concentration and long reaction times were sufficient to drive loxP recombination to completion. These two hypotheses can be distinguished by comparing recombination of pseudo-loxP sites under conditions at which WT and mutants have comparable, subequilibrium efficiencies of loxP recombination. As such conditions can be easily found in vitro, we attempted to purify the WT and mutant Cre recombinases. Affinity purification of using a maltose-binding protein (MBP)
domain fusion followed by scarless protease cleavage of the MBP domain allowed isolation of WT, R32V and R32M (Supplementary Fig. S2). We attribute our failure to purify 303GSdup to its lack of solubility following MBP removal.

We measured recombination kinetics using an intramolecular excision assay on linearized plasmid substrates. To test off-target activity, we used a pseudo-loxP site lox80 (Fig. 3b), which preliminary tests revealed to be efficiently recombinated by WT in vitro. WT, R32V and R32M were all able to recombine loxP within the 8-h assay. In contrast to WT, R32V and R32M did not produce detectable loxP × lox80 reaction products (Fig. 6a–c). WT recombination of the loxP × lox80 substrate produced a small amount of an unexpected product that migrated at approximately twice the size of the linear recombination product. This product, which we excluded from our quantitative analysis, may represent recombination with a cryptic pseudo-loxP site in the vector backbone.

Cre recombined loxP × loxP and lox80 × loxP reactions with comparable kinetics, although it was significantly less effective recombining lox80 at the 5-min time point (Fig. 5d,e). R32V and R32M recombined loxP sites significantly more slowly than WT, with both mutants failing to achieve a steady state within the 8-h experiment. The kinetic data enabled pseudo-loxP recombination efficiencies of the different mutants to be compared at similar loxP × loxP recombination efficiencies. For instance, after 5 min WT had recombined loxP sites with 21% efficiency, after 8 h R32V had recombined loxP sites with 29% efficiency and after 4 h R32M had recombined the loxP substrate at an efficiency of 27%. Despite the slightly higher loxP recombination efficiencies, R43V and R32M had no detectable off-target activity, whereas WT recombined lox80 and loxP with 4% efficiency. A similar conclusion can be reached by comparing the 46% loxP × loxP efficiency and 35% lox80 × loxP achieved by WT after 30 min to the 44% loxP × loxP efficiency and absence of off-target recombination seen with R32M after 8 h. In aggregate, these data support the model under which the mutations we identified destabilize both loxP and pseudo-loxP recombination, but with a much greater reduction in pseudo-loxP efficiency than in loxP efficiency.

Our observation that R32V and R32M are functional in vitro suggested that the mutants, such as Cre, could catalyse loxP recombination in mammals. To test the functionality of Cre mutants in human HEK293 cells, we expressed R32V, R32M and 303GSdup under the control of the cytomegalovirus promoter. The substrate plasmid contained a loxP-flanked GFP with downstream lacZ gene, such that recombination excised GFP and made the cells turn blue when stained with X-gal (Supplementary Fig. S3a). Two and a half days following transfection with R32V, R32M and 303GSdup, the cells had lost the GFP signal and stained positive for β-galactosidase activity (Supplementary Fig. S3b). As in E. coli, the in vivo reaction approached completion; the transfection efficiency-normalized recombination efficiencies were 95.2% for R32V (95% CI: 85.9–100%), 99% for R32M (95% CI: 97.1–100%) and 100% for 303GSdup (with no variation across replicates; Supplementary Fig. S3c).

Mutants have less off-target activity genome wide. One explanation for the improved accuracy we observed with R32V, R32M

Figure 6 | In vitro recombination of plasmids by mutants of Cre. (a) WT Cre purification of loxP × loxP (left) and lox80 × loxP substrates was performed for the indicated amount of time and resolved on an agarose gel. The linearized substrate plasmid contained two recombination sites such that catalysis resulted produced a circular and a short linear fragment. The same conditions were used to test (b) R32V and (c) R32M. The band intensities were quantified and plotted for (d) loxP × loxP and (e) lox80 × loxP recombination. Error bars correspond to 95% CI (n = two to four experiments using different protein preparations).
Table 1 | Genome-wide off-target integration frequency.

| loxP + | loxP − | 95% confidence intervals |
|-------|-------|------------------------|
| ori − | ori + | insertion frequency |
| colonies | colonies | colonies |

Y324F  1 | 6.2 × 10^6 | 1.6 × 10^−7 | 0–4.8 × 10^−7 |
WT  1.025 | 8.1 × 10^6 | 1.3 × 10^−4 | 1.2 × 10^−4–1.4 × 10^−4 |
R32V  2 | 9.4 × 10^6 | 2.1 × 10^−7 | 0–5.1 × 10^−7 |
R32M  3 | 6.3 × 10^7 | 4.8 × 10^−8 | 0–1 × 10^−7 |
303GVSdup  46 | 2.6 × 10^7 | 1.7 × 10^−6 | 1.2 × 10^−6–2.2 × 10^−6 |

Insertion frequency corresponds to the ratio of loxP + to loxP − colonies. 95% confidence intervals were computed using the Poisson variance. The data represent pooled observations from three to four independent experiments.

and 303GVSdup is that the mutations altered the preferred off-target sites without changing the overall accuracy. To test this possibility, we measured the efficiency of off-target insertions across the entire E. coli genome. Strains carrying the arabinose-inducible recombinase expression plasmids were transformed with a plasmid containing a loxP site, an R6k-γ origin of replication and a kanamycin resistance gene. As the R6k-γ origin cannot replicate in the pir − strain we used, only insertion of the plasmid into the genome would result in the replication of the kanamycin resistance gene. To control for variable transformation efficiencies, we normalized the number of R6k-γ colonies by the number of colonies arising from transformation with a plasmid lacking loxP but containing a functional origin of replication.

We were unable to obtain Cre-mediated integration on LB/glucose. However, when we briefly pulsed the cells with arabinose before the growth on LB/glucose, we obtained a WT-mediated integration frequency of 1.3 × 10^−4 (Table 1). 303GVSdup had an integration frequency ∼100-fold lower than WT. The integration frequencies of R32V and R32M were lower than that of 303GVSdup and could not be distinguished from the Y324F background. These data strongly suggest that the higher accuracy of the mutants was not restricted to loxBait1, ψLox h7q21 and lox80.

Discussion

Our data suggest that R32V, R32M and 303GVSdup can efficiently recombine loxP in bacteria and in human cells. We also observed that the mutants exhibit better directionality and better accuracy with respect to the pseudo-loxP site ψLox h7q21 than WT Cre. R32V and R32M were functional in vitro and were better able to discriminate against the lox80 site than WT Cre. All three mutants more rarely integrated a loxP-carrying plasmid into the E. coli genome than WT Cre. Crystal structures of Cre strongly suggest that R32V and R32M disrupt a strong salt bridge in the dimer interface (Fig. 4a). The duplicated residues in the de novo 303GVSdup mutant are also located at the dimer interface, although the exact biochemical consequences of the mutation are unclear.

Despite our model’s prediction that higher protein concentrations should reduce the accuracy of recombination (Fig. 1), we saw identical results in our bacterial plasmid recombination assay when the Pbad promoter controlling protein expression was repressed with glucose (Fig. 5c,d) and when it was weakly induced by LB without glucose. Despite the apparent contradiction, the data do not disprove the model. Even under glucose repression, WT Cre was able to drive both the loxP and pseudo-loxP reactions to equilibrium within the time frame of the experiment; therefore, any increases in recombination efficiency caused by a higher protein concentration could not be measured. The same logic applies to R32V, R32M and 303GVSdup recombining loxP sites. The model predicts that higher protein concentrations should cause Cre mutants to recombine pseudo-loxP sites more efficiently. The data neither confirm nor deny this prediction, as both expression conditions resulted in a recombination frequency that was below the detection level.

The main limitation of our accuracy model is that it considers only binding, ignoring catalysis. The binding site sequence likely contributes to both establishing the proper alignment of the catalytic site and creating a tertiary DNA structure within the recombinase complex that is energetically favourable for recombination. The DNA site may also contribute to protein folding. It is therefore possible that R32V, R32M and 303GVSdup mutations contribute to improved accuracy in steps of the catalytic pathway other than DNA binding. If that is the case, there may exist pseudo-loxP sites that are recombined by WT Cre but not at all by the mutants.

Although R32V and R32M had high in vivo recombination efficiencies, the mutants catalysed recombination slower than WT in vitro (Fig. 6d). The slower kinetics may explain some of the decrease in apparent off-target integration in the genome-wide assay (Table 1). However, during the genome integration assay, the cells were induced for 30 min and then recovered for 1 h in glucose media before antibiotic selection. If it is assumed that recombination occurred for 1 h (an underestimate, as recombination can continue to happen after the addition of antibiotics), then the in vitro data predict that R32V and R32M will produce 10- to 20-fold fewer transformants than WT. In fact, R32V and R32M were at least 200 times more accurate than WT. We therefore believe that the majority of the decrease in the frequency of genomic integration was caused by an improvement in selectivity against pseudo-loxP sites and not by the mutants’ slowed reaction kinetics.

Our observation that Cre make frequent deletions on a substrate containing inverted loxP has been made previously and is consistent with the fact that Cre can occasionally recombine sites with non-matching spacers. This looseness in the directionality of recombination may interfere with synthetic circuit designs that rely on Cre-mediated inversion as a form of genetic memory. The three Cre mutants we isolated seem to have an improved directionality over WT and may therefore be of use in synthetic biology applications.

As the binding specificity of nucleases is the major determinant of their toxicity, it may be possible to minimize nucleases’ toxicity by reducing their DNA-binding cooperativity. The main limitation of this approach is the fact that it inevitably decreases affinity for the target site. Strong monomer affinities, such as the 1–10 nM Kd of Cre for half of loxP (refs 32,40), are a likely prerequisite. For comparison, DNA-binding domains composed of three Cys–His zinc fingers have been reported to bind to their 9 bp recognition sequence with a KD as low as 400 pM, whereas one 17.5 TALE repeat domain interacts with DNA with a KD of 160 pM (refs 41,42). Although only the better zinc finger and TALE designs achieve such high binding energies, the reported values suggest that destabilizing the cooperativity of DNA binding may be a viable strategy for increasing the accuracy of designer nucleases.

Changing the dimer interface has been previously used to improve the specificity of nucleases engineered to cut asymmetrical DNA sites. As targeting asymmetric sites necessitates coexpression of two different DNA-binding domains fused to a constant nuclease domain, catalytically active dimers can form at four different sites: the desired asymmetric site, an off-target asymmetric site and the two symmetric sites targeted by the two possible homodimers. Converting the nuclease dimerization interface into a heterodimeric interface reduces activity at the
two symmetric off-target sites\(^{43–45}\). The approach described in this paper is different from the obligate heterodimer strategy, because destabilizing the cooperativity of a dimeric protein improves accuracy by increasing the energy difference between binding to the target and the off-target sites—not by reducing the number of possible target sites. The obligate heterodimer and the destabilized cooperativity strategies should be compatible with each other, as it should be possible to modulate the binding cooperativity of heterodimers.

Because of its high efficiency and the lack of necessary cofactors, Cre has found widespread use in genetics research\(^{46}\). However, Cre toxicity in the absence of loxP sites has been observed in a variety of animal and cell culture systems\(^{47–51}\). The source of this toxicity is not known. However, a number of observations, including the absence of toxicity from catalytically inactive Cre mutants, an increase in the frequency of chromosomal rearrangements and evidence of activation of DNA damage response pathways, all point at recombination at sites has been identified in this study may be useful for alleviating the toxic phenotypes associated with WT Cre. This approach should be compatible with existing strategies for reducing Cre toxicity, which include placing the Cre gene in a self-excissable cassette, regulating Cre activity with a hormone-binding domain of a steroid receptor, or using a drug-regulated fragment complementation strategy\(^{50,52–54}\).

**Methods**

### Plasmid construction

The positive selection substrate pCR-loxP-amplR-loxP\(^{455}\) was built by amplifying ampR from pQPL13 (ref. 55) with primers ampR\(_f\) and ampR\(_r\) (primer sequences are provided in Supplementary Table S1), performing an extension PCR on the amplicon with ampR\(_{loxP}\_f\) and ampR\(_{loxP}\_rev\_r\), and TOPO cloning the product into pCR-Blunt II-TOPO (Life Technologies). Cre\(_{TA}\_r\) and TOPO TA cloned into pcDNA3.3-TOPO (Life Technologies).

### DNA damage response pathways, all point at recombination at sites has been identified in this study may be useful for alleviating the toxic phenotypes associated with WT Cre. This approach should be compatible with existing strategies for reducing Cre toxicity, which include placing the Cre gene in a self-excissable cassette, regulating Cre activity with a hormone-binding domain of a steroid receptor, or using a drug-regulated fragment complementation strategy\(^{50,52–54}\).

### Negative and positive selections

The pCR-(loxP-ampR-loxP\(^{455}\)) and pCR-loxBait-ampR-loxP\(^{455}\) selection plasmids were maintained in NEB 1b-beta cells (New England Biolabs; recA1 ada13 D7079 A1571 C1767) and were electrocompetent using standard techniques\(^{59}\). The library of Cre variants was generated by mutagenic PCR using a pool of 19 oligonucleotides that substituted each of the 19 codons encoding S20–S38 for NNN (first round of PCR: primers Cre\(_{mut\_1}\)_r and Cre\(_{PCR}\_r\); second round of PCR: primers Cre\(_{mut\_2}\) and Cre\(_{PCR}\_r\)). The library was introduced into pARC8 using Gibson assembly, desalted by drop dialysis and electroporated into competent cells carrying pCR-loxP-amplR-loxP\(^{455}\). Control transformations were performed with 100 pg pARC8-Cre or pARC8-Y234F. Transformed cells were recovered in low-salt 2\(\times\) 2\% bacto-tryptone, 20\% NaCl, 37 \(^\circ\)C for 30 min, induced with 0.2% arabinose at 37 \(^\circ\)C for 30 min and recovered in SOC with 200 \(\mu\)M IPTG (isopropyl-\(\beta\)-thiogalactoside) at 37 \(^\circ\)C for 1 h. The cells were then grown overnight at 37 \(^\circ\)C on 0.2% glucose, 100 \(\mu\)M IPTG, 12.5 \(\mu\)g ml\(^{-1}\) chloramphenicol, 50 \(\mu\)g ml\(^{-1}\) kanamycin and 100 \(\mu\)g ml\(^{-1}\) carbencillin LB plates. To ensure clonality, the isolated variants were amplified and re-cloned into pRCS via Gibson assembly.

### Bacterial recombination assay

Efficiencies were measured by cotransferring 75 ng of a pARC8-based expression plasmid (WT, Y34F, R23F, R23M or 303GVdup) and an equimolar amount of either pZEO-loxPloxBait or pZEO-\(\psi\)Cre H7q21/loxP H7q21 into 50 \(\mu\)l of One Shot Top10 chemically competent cells. The cells were recovered in LB with 0.2% glucose at 37 \(^\circ\)C for 1 h, then grown at 37 \(^\circ\)C for 12 h in 0.2% glucose, 12.5 \(\mu\)g ml\(^{-1}\) chloramphenicol and 50 \(\mu\)g ml\(^{-1}\) kanamycin LB. The plasmids were isolated using the QiAprep Spin Miniprep kit, eluting into 50 \(\mu\)l of 10 mM Tris-HCl pH 7.8. The purified plasmids were digested for 20 \(\mu\)l at 37 \(^\circ\)C in 60-\(\mu\)l reactions containing all of the DNA collected and 20 \(\mu\)l of both ScaI-HF and NcoI-HF (both from New England Biolabs) in NERBuffer 4 (1 mM dithiothreitol (DTT), 50 mM potassium acetate, 20 mM tris-acetate, 10 mM magnesium acetate, pH 7.9). The digests quantified on 1% agarose gels following purification using the QiAprep Spin Miniprep kit.

### In vitro recombination assay

The pMAL-based recombination expression plasmids were maintained in One Shot Top10 chemically competent cells. An overnight culture of each expression clone was diluted to OD\(_{600}\) of 0.02 in 55 ml 0.2% glucose, 50 \(\mu\)g ml\(^{-1}\) kanamycin LB and grown for 2.5 h at 37 \(^\circ\)C. IPTG was added to a final concentration of 300 \(\mu\)M and the cultures were grown at 37 \(^\circ\)C for an additional 2\(\times\)h. The cultures were then pelleted, resuspended in 20 \(\mu\)l of 1 mM EDTA, 1 mM DTT, 200 \(\mu\)g ml\(^{-1}\) chloramphenicol and 0.1 M Tris-HCl pH 8 to supplement with Halt Protease Inhibitor Cocktail (Thermo Scientific). The cell suspension was lysed using eight 10\(\times\) sonication cycles using a power setting of 4 on a Misonix Sonicator 3000. The lysate was centrifuged for 20 min at 17,000 \(\times\)g at room temperature. The supernatants were filtered through 0.22 \(\mu\)m filters. The protein was then purified using one 50 \(\mu\)l aliquot of Sephacryl 600. The elution was washed with 4 \(\mu\)l wash buffer (200 \(\mu\)l NaCl 20 mM Tris-HCl pH 7.5, 125 \(\mu\)l of DNA removal buffer) and then washed 4 \(\mu\)l wash buffer. The protein was eluted into 500 \(\mu\)l elution buffer (0.5% maltose, 2 \(\mu\)g ml\(^{-1}\) BSA, 100 \(\mu\)g ml\(^{-1}\) kanamycin, 100 \(\mu\)g ml\(^{-1}\) chloramphenicol) at 37 \(^\circ\)C and digested with 200 \(\mu\)g Factor Xa (New England Biolabs) for 13–14 h at room temperature. The protease activity was stopped by adding 1 \(\times\) Halt Protease Inhibitors and 1 \(\mu\)l DTT. Precipitated protein was removed by centrifuging at 16,000 g for 10 min and the supernatant was then incubated with an Amicon Ultracel 100K filter. The flowthrough was concentrated using an Amicon Ultra 100K filter. Glycoler was added to 27% final volume. NaCl was added to adjust the concentration to 200 \(\mu\)M and the proteins were stored at -20 \(^\circ\)C. Catalytic assays were performed within 2 days of purification. Protein concentration was determined using a Qubit 2.0 fluorometer (Life Technologies). The concentration of protease-cleaved recombimase was determined by quantifying band intensity on a SimplyBlue (Life Technologies)-stained PAGE gel (typical values 15–30%) and then by normalizing the total protein concentration by the obtained value.

### Human cell culture recombination assay

pCMV:GFP (loxP)lacZ (ref. 60; Addgene: 31125) was used as the recombination substrate plasmid and pcDNA3.3-
TOPO/lacZ. (Life Technologies) was used as a lacZ staining control. Except for pcDNA3.3-TOPO/lacZ, which was prepared by the manufacturer, all substrate and expression plasmids were prepared for transfection using the HEPaD Plasmid Maxi Kit (Qagen).

Lenti-X 293T cells (Clontech) were grown on poly-D-lysine–coated tissue-culture-treated polystyrene in DMEM High Glucose with GlutaMAX (Life Technologies) supplemented with 10% fetal bovines serum, 50 U ml⁻¹ penicillin and 50 μg ml⁻¹ streptomycin at 37 °C and 5% CO₂ in a humidified incubator. Cells were subcultured using TrypLE Express (Life Technologies). Cells in 12-well plates were transfected with 500 ng pCMV-GFP(loxP)lacZ and 500 ng pcDNA3.3-TOPO-based expression plasmid using Fugene HD (Promega). Substrate-only controls were transfected with 500 ng recombinase substrate and the lacZ controls were transfected with 500 ng pcDNA3.3-TOPO/lacZ. Media was changed 1 day following transfection. Cells were subcultured 2 days following transfection. Staining and imaging was performed 2.5 days following transfection. Cells were fixed using 2% formaldehyde and 0.1% glutaraldehyde in PBS for 10 min at room temperature. After being washed with PBS, the cells were stained with 1 μg ml⁻¹ X-gal, 4 mM potassium ferricyanide, 4 mM potassium ferrocyanide and 2 mM MgCl₂ in PBS for 30 min at 37 °C. Following staining, the cells were washed with PBS and imaged.

Genome-wide off-target integration assay. Electroporation of competent cells were prepared from each of the pARC8-based expression plasmids (WT, Y324F, R32V, R32M or 303GVSdup) cloned in NEB 10-beta cells using 40 ml of culture per transformation. Each expression strain was transformed either with 200 ng pUNI10 (loxP) or 303GVSdup) cloned in NEB 10-beta cells using 40 ml of culture per transformation. Each expression strain was transformed with 500 ng recombination substrate and the lacZ control with altered DNA binding properties. J. Biol. Chem. 273, 22884–22891 (1998).

Keravala, A. et al. A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. Mol. Genet. Genomics 276, 713–724 (2006).

Semprini, S. et al. Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. Nucleic Acids Res. 35, 1402–1410 (2007).

Nelson, H. C. M. & Sauer, R. T. Interaction of mutant λ repressors with operator and non-operator DNA. J. Mol. Biol. 192, 27–38 (1986).

Almendros, I. et al. Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach. Nucleic Acids Res. 41, 2394–2403 (2012).

Thyagarajan, B., Guimaraes, M. J., Groth, A. C. & Calos, M. P. Mammalian genomes contain active recombinase recognition sites. Gene 244, 71–78 (2000).

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. & Calos, M. P. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol. Cell Biol. 21, 3926–3934 (2001).

Keravala, A. et al. A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. Mol. Genet. Genomics 276, 713–724 (2006).

Semprini, S. et al. Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. Nucleic Acids Res. 35, 1402–1410 (2007).

Nelson, H. C. M. & Sauer, R. T. Interaction of mutant λ repressors with operator and non-operator DNA. J. Mol. Biol. 192, 27–38 (1986).

Almendros, I. et al. Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach. Nucleic Acids Res. 41, 2394–2403 (2012).

Thyagarajan, B., Guimaraes, M. J., Groth, A. C. & Calos, M. P. Mammalian genomes contain active recombinase recognition sites. Gene 244, 71–78 (2000).

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. & Calos, M. P. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol. Cell Biol. 21, 3926–3934 (2001).

Keravala, A. et al. A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. Mol. Genet. Genomics 276, 713–724 (2006).

References
1. Stocking, C. et al. Distinct classes of factor-independent mutant can be isolated after retroviral mutagenesis of a human myeloid stem cell line. Growth Factors 8, 197–209 (1993).
2. Hacetine-Abya, S. et al. A serious adverse event after successful gene therapy for X-Linked severe combined immunodeficiency. N. Engl. J. Med. 348, 253–260 (2003).
3. Rouet, P., Smith, F. & Jasim, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell Biol. 14, 8096–8106 (1994).
4. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Nat. Protoc. 6, 1156–1160 (2011).
5. Li, T. et al. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res. 39, 359–372 (2010).
6. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757–761 (2010).
7. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013).
8. Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).
9. Mao, Z., Bozella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. J. Mol. Biol. 7, 1765–1771 (2008).
10. Hartlerode, A. J. & Scully, R. Mechanisms of double-strand break repair in somatic mammalian cells. Biochim. J. 423, 157–168 (2009).
11. Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombinase for improved genomic integration at a native human sequence. Protein Eng. Des. Sel. 14, 227–236 (2001).
12. Buchholz, F., Stewart, A. A. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nat. Biotechnol. 19, 1047–1052 (2001).
13. Siciliano, C. R., Thyagarajan, B. & Calos, M. P. Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res. 29, 5044–5051 (2001).
14. Santoro, S. W. & Schultz, P. G. Directed evolution of the site specificity of Cre recombinase. Proc. Natl Acad. Sci. 99, 4185–4190 (2002).
15. Akerman, A., He, J., Boocock, M. R. & Stark, W. M. Chimeric recombinases with designed DNA sequence recognition. Proc. Natl Acad. Sci. 100, 8688–8691 (2003).
16. Gordin, R. M., Smith, J. D., Graslund, T. & Barbas, III C. F. Evolution of programmable zinc finger-recombinases with activity in human cells. J. Mol. Biol. 367, 802–813 (2007).
17. Asokar, L., Hauber, J., Hauber, J. & Buchholz, F. HIV-1 proviral DNA excision using an evolved recombinase. Science 316, 1912–1915 (2007).
18. Keravala, A. et al. Mutational derivatives of phiC31 integrase with increased efficiency and specificity. Mol. Ther. 17, 112–122 (2009).
19. Gersbach, C. A., Gaj, T., Gordley, R. M. & Barbas, C. F. Directed evolution of recombinase specificity by split gene reassembly. Nucleic Acids Res. 38, 2026–2048 (2010).
20. Abi-Ghanem, J. et al. Engineering of a target site-specific recombinase by a combined evolution- and structure-guided approach. Nucleic Acids Res. 41, 2394–2403 (2012).
21. Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. & Calos, M. P. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol. Cell Biol. 21, 3926–3934 (2001).
22. Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. & Calos, M. P. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol. Cell Biol. 21, 3926–3934 (2001).
49. Loonstra, A. et al. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl Acad. Sci.* 98, 9209–9214 (2001).
50. Silver, D. P. & Livingston, D. M. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol. Cell* 8, 233–243 (2001).
51. Huh, W. J., Mysorekar, I. U. & Mills, J. C. Inducible activation of Cre recombinase in adult mice causes gastric epithelial atrophy, metaplasia, and regenerative changes in the absence of 'floxed' alleles. *Am. J. Physiol.–Gastrointest. Liver Physiol.* 299, G368–G380 (2010).
52. Feil, R. et al. Ligand-activated site-specific recombination in mice. *Proc. Natl Acad. Sci.* 93, 10887–10890 (1996).
53. Kellendonk, C. et al. Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res.* 24, 1404–1411 (1996).
54. Jullien, N., Sampieri, F., Enjalbert, A. & Herman, J.-P. Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids. Res.* 31, e131 (2003).
55. Liu, Q., Li, M. Z., Leibham, D., Cortez, D. & Elledge, S. J. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* 8, 1300–1309 (1998).
56. Choe, W., Chandrasegaran, S. & Ostermeier, M. Protein fragment complementation in M.HhaI DNA methyltransferase. *Biochem. Biophys. Res. Commun.* 334, 1233–1240 (2005).
57. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345 (2009).
58. Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* 22, 841–847 (2004).
59. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press; 3rd edition, 2001).
60. Werdien, D., Peiler, G. & Ryffel, G. U. FLP and Cre recombinase function in Xenopus embryos. *Nucleic Acids Res.* 29, E53–E53 (2001).

**Acknowledgements**

This work was supported by the US Office of Naval Research Multidisciplinary University Initiative and the National Human Genome Research Institute Center for Excellence in Genomics Science. We thank G. Meister for pARC8-HpaII[51–358] and S. Elledge for pQL123 and pUNI10. We also thank P. Mali, J. DiCarlo and A. McElhinney for a critical reading of the manuscript, and S. Kosuri, S. Raman, M. Lajoie and X. Rios for helpful advice and inspiring discussions.

**Author contributions**

N.E. conceived the study, designed and performed all experiments, and wrote the manuscript. G.M.C. supervised the project and provided critical advice.

**Additional information**

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Eroshenko, N. and George, M.C. Mutants of Cre recombinase with improved accuracy. *Nat. Commun.* 4:2509 doi: 10.1038/ncomms3509 (2013)