Vika/*vox*, a novel efficient and specific Cre/*loxP*-like site-specific recombination system.

Madina Karimova[1], Josephine Abi-Ghanem[3], Nicolas Berger[1], Vineeth Surendranath[2], M. Teresa Pisabarro[3], Frank Buchholz[1]*

[1]Medical Systems Biology, University Hospital and Medical Faculty, Carl Gustav Carus, University of Technology, Fetscherstrasse 74, 01307,Dresden, Germany
[2]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
[3]Structural Bioinformatics, BIOTEC TU Dresden, Tatzberg 47-51 01037 Dresden, Germany

Supplementary Table 1. Primers used for vectors' construction.
The first column shows the primer names used in PCR reactions. The vectors that served as recipients are highlighted in bold in this column. The second column lists the sequences of the oligos used in the PCR reactions

Primer name	Primer sequence 5'→3'
pEVO vectors construction	
pEVO-*VloxP*-up	ATGCTCGAGTCAATTTCCGAGAATGACAGTTCTCAGAAATTGAAAGCTTGCATGCCTGCAGA
pEVO-*VloxP*-low	TTGAGATCTCAATTTCTGAGAACTGTCATTCTCGGAAATTGATCGAACTGTACCGGTTGTTAGTG
pEVO-*rox*-up	ATGAGATCTTAACTTTAAATAATGCCAATTATTTAAAGTTAAAGCTTGCATGCCTGCAGATCGAG
pEVO-*rox*-low	ttgagatcttaactttaaataattggcattatttaaagttatcgaactgtaccggttgttagtga
pEVO-*vox*-up	gtGACAGTAGATCTAATAGGTCTGAGAACGCCCATTCTCAGACGTATTGGTCTGACGCTCAGTGGAAC
pEVO-*vox*-low	gtGACAGTCTCGAGgcAATACGTCTGAGAATGGGCGTTCTCAGACCTATTCCGATTTCGGCCTATTGG
Catalytic mutant VikaY243F	
VikaY343F_up	cgaaaaaatggttctgcgttttggtcgtcatctgcatg
VikaY343F_low	catgcagatgacgaccaaaacgcagaaccatttttttcg
NLS protein tagging (N-terminal)	
NLS-VCre-up	AATTGTACAGCCACCATGaagaagaagaggaaggtgATCGAGAACCAGCTGAGC
NLS VCre-low	GACTCTAGAGTTTAATTAATCATTATATGC
FLAG protein tagging (C-terminal)	
Flag-Vika-up	CTCTGTACAATGACCGATCTGACCCCGT
Flag-Vika-low	ACCTCTAGATTActtgtcatcgtcatccttgtaatcACGCTGACGACGTTTCTGTGCC
pSV vectors construction	
1-pSV-up	ACAATTAATAGACTGGATGGAGG
1-pSV-*VloxP*-low	TGCAAGCTTCAATTTCTGAGAACTGTCATTCTCGGAAATTGATCGAACTGTACCGGTTGTTAG
2-pSV-*VloxP*-up	ctggatcctcaatttccgagaatgacagttctcagaaattgaACCGGGCCACCATGGTCGCGAGTAGC
2-pSV-low	CAGGATATCCTGCACCATCGTCTGCTCATC
1-pSV-*vox*-low	TGCAAGCTTAATACGTCTGAGAATGGGCGTTCTCAGACCTATTTCGAACTGTACCGGTTGTTAG
2-pSV-*vox*-up	ctggatccAATAGGTCTGAGAACGCCCATTCTCAGACGTATTACCGGGCCACCATGGTCGCGAGTAGC
pD vectors	
R6K-*loxP*-up	actTCTAGAATAACTTCGTATAGCATACATTATACGAAGTTATggtctgacgctcagtggaac
R6K-*VloxP*-up	actTCTAGATCAATTTCcGAGAATGACAGTTCTCAGAAATTGAggtctgacgctcagtggaac
R6K-*rox*-up	actTCTAGATAACTTTAAATAATTGGCATTATTTAAAGTTAggtctgacgctcagtggaac
R6K-*vox*-up	actTCTAGAAATAGGTCTGAGAACGCCCATTCTCAGACGTATTggtctgacgctcagtggaac
R6K-low	actCTCGAGGAAATGTGCGCGGAACCCC
pRK-eGFP vectors	
pRK-*VloxP*-up	ggaagatctTCAATTTCCGAGAATGACAGTTCTCAGAAATTGAAAGCTTAGGTGGCACTTTTCG
pRK-*VloxP*-low	TCCctcgagTCAATTTCTGAGAACTGTCATTCTCGGAAATTGATATCGACAGAGTGCCAGCC
pRK-*vox*-up	ggaagatctAATAGGTCTGAGAACGCCCATTCTCAGACGTATTAAGCTTAGGTGGCACTTTTCG
pRK-*vox*-low	TCCctcgagAATACGTCTGAGAATGGGCGTTCTCAGACCTATTTATCGACAGAGTGCCAGCC
pBabe retroviral vectors	
pBabe-psi-mid-up	GGTACTAGTTAGCTAACTAGCTCTGTATCTGG
pBabe-Xhoi-EcoRI-low	GGCGAATTCatacagtatctcgagCTACGTACCACCACACTGG
pNPK-rec-xhoI-up	gtcCTCGAGGCTCAGGAGGAATTGTACAG
Cre/173-XhoI-up	TAGctcgagtccaatttactgaccgtac
Cre-EcoRI-low	AGAGAATTCGACTCTAGAGCTAATCGCCATCTTCCAGCAG
pNPK-Vika-EcoRI-low	AGAGAATTCgacTCTAGAttaCCGCTGTCTCCGCTTC

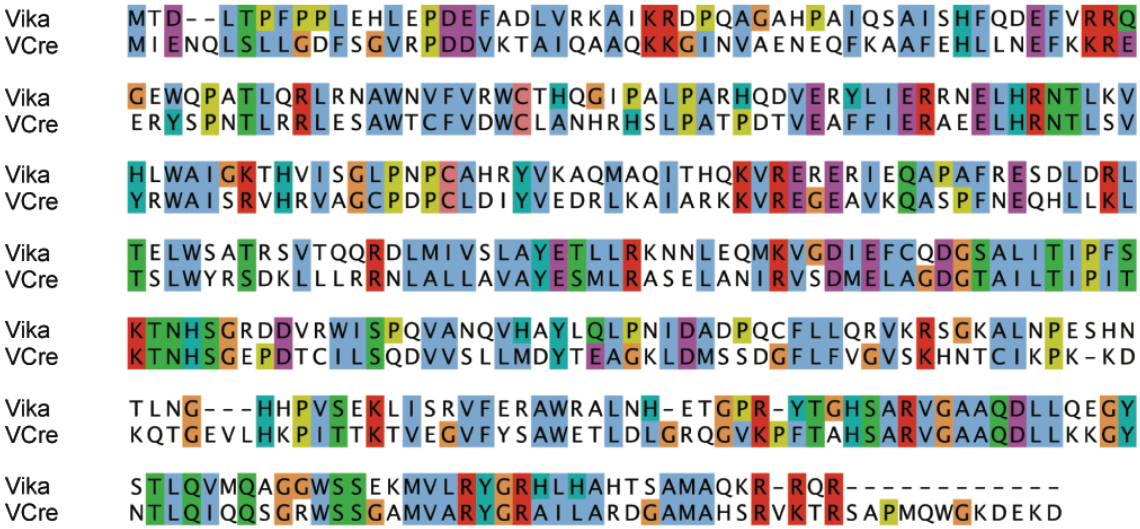
Supplementary Table 2. Sequences of the proposed recombinases and their target sites.
Amino acid sequences and accession numbers of putative tyrosine site-specific recombinases are depicted in the first column of the table. Target sites for the putative recombinases were predicted and their DNA sequences are shown corresponding to each recombinase in the second column. The origin of each system is shown in the third column as the name of the organism, strain and name of plasmid, if applicable. Amino acid identity of proposed recombinases was compared to Vika (A) or Cre (B) and is shown as percentage in the last column.

GenBank Accession number Amino acid sequence	Putative target site	Origin	Sequence identity (%) to (A)Vika (B)Cre
EGU56467.1 MTTLSVLSEVPFERLLPHEFAEGLAAAQRAGEALEG HPLVEAAITHYQGEFFRRAERLQPASLVRLKSAWAT FVAWCCEQDRCALPASPQTVEAYLIAEQDRLHRNTL KVQLWAIGKTHQISGCPDPCHNDYVKAQLQQIHHRK VRQREVIRQAVALRESHLNALADLWDRPEASLTECR DLLIVSMLYETLLRKSNLETLRVGDVDWQADGSGLI KVFVTKTDKSGDVKYSYVSPSTMDLLARYLGHADIV DNPEAFLIQRVKLSSQQLKGSARTQAAISPVSAKLI GRVCAKAAKTLGLSTDRPFTGHSARVGATQDLLAEG FSSLQVQQAGGWSSERMVLRYGGSVLASESAMAQRR QRKSPK	CATACGTCCTAGA**ATGGCAGT**TCTAGGACGTATT	Vibrio tubiashii ATCC 19109	(A) 50 (B) 23
YP_003065675.1 MELVATDSAAEPQRDAFNPPVPFADALPPGLELLIE RLEQHARAARGAFADNTVRALAADSRIFAAWCREEG RAMLPATPETVAAFIDAQGETKARATVERYRSSIAA LHRAAGLPNPCADEIVRLAVKRMNRARGRRQKQAEP LNRASIERMLEVKTPGRLHRRVTEAKRETPLIALRN AALVAVAYDTLLRRSELVSLYIGDLHKGADGSGTVL VRRSKADQEGEGAIKYLAPDTMAHIEAWLSAAHLES GPLFRPLTKGGQVGTVALGGGEVARVFRDLATAAGL KLARLPSGHSTRVGATQDMFAAGFELLEVMQAGSWK TPAMPARYGERLRAQRGAARKLATLQNRA	ATTTCCCGCGATA**GATGGTGT**TATCGCAGGCAAT	Methylobacterium extorquens DM4 plasmid p1METDI	(A) 23 (B) 24
YP_003280920.1 MTEHDQGEVVDAELVDDQLPALRNQAQAPAVPAPKN DPDAWLSDQAREDVKAGIADGTRDGYKGDMERFAAW CTSAGRRPMPAAPQTVTEYLSYLKHTPRPRTNKPYG PNSMDRIIAAIRSAHRAAGHEPPDTMGARKVVLGYR AELSERKDPAAKPRKATPADRAVLRRALAELDRATL AGQRDAALMLLGHALASRGSELVPLNIPDSFTDLPD GGFSVAVYRKKRKCWQDVTVVLDPDPDLCAVRAVRR LVATLADNGHHTGPLFLRMDRWGYLAPPMHRNGKPI GDPTGRMTVEAASDIVQRSIERTGIPGRWRSHSSRR GFVKSARQAGVDIVQIGRHGGWDDKSKALIGYIDEE DAQGDNNPLVQIGRKAALPPDAASGT	GTTGCCCCCGTCG**CGCGGTCG**CGTTGGGGGCAAC	Streptomyces sp. W9 plasmid pCQ3	(A) 14 (B) 15
ZP_06822377.1 MAIRRGALTSGPDRAKLSAGAVAAMEKGIPPETRRG YAGDWQRFEAWAFGEGACPLPCSAETLTEYVTFLTV FPRPRTGMPYEPAPIERAMAAIAVAHKAAGFAPPDT TGARLVLRGYERELKETKDPRGRVAKAAAATPLILR TMIAHTDLTTPIGLRDAAAMTNGFALAARSSEAKLL DWEDTADVEQGLEYDLYRPKVNNDQPLGVPYGAYPS TCPVRRLHAWRQCLLDLGYPVSGPIYVRINRHGHIN PPMTRRGLPIGDPSGRMTTEGIAEIVTRAAKRAGLT AVPDDLLPSLPPRWSGHSLRRGYAKAAREAGKDMLE SGRHGGWADGSRAFAGYFDRAAIWDEDLNPLFGIGL	CTGGCTCTTGGTA**AGGCACGT**TATCAAGAGCCAA	Streptomyces sp. SPB74	(A) 8 (B) 13
NP_395953.2 MTDQDVETLRHLVNQGMGDNTLRALTSDLAYLEAWG LATTGSSLPWPAPEALLLKFVAHHLWDPEKRATDPD HGMPAAVDENLRRQGFLRSVGPHAPSTVRRRLANWS TLTRWRGLHGAFASPALKSAIRLAVRAVPRTRARKS AKAVTGDVLAKLLATCESDSLRDLRDKAILMVAFAS GGRRRSEIAGLRREQLTIEAPIETEGGPPLPSLAIH LGRTKTTSGEEDDTVFLTGRPVEALNAWLAAAKIDK	AGCCATCAAGATG**GCAGACGC**CATCTTGATGGCT	Agrobacterium tumefaciens str. C58 plasmid At	(A) 14 (B) 15

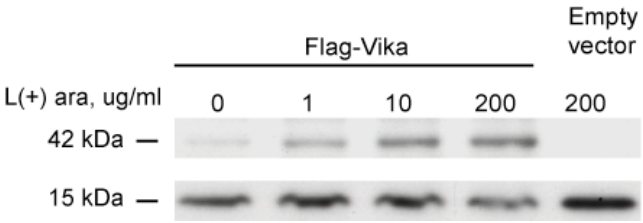
GSVFRGIGRWGTVSRRALDPQSVNAILKQRAEMAGL EAGQFSAHGLRSGYLTEAANRGIPLPEAMEQSRHRS VQQASSYYNSATRRSGRAARLL*			
YP_666181.1 MTRIAAFDGRSAEFVAPRLRLPNHARISTMTNTVHQ PADDLPDIVDLVKEMCRPTQLERQSGSDKPNPPALP AAHRAENQIPSHLDGLADRARGYVEAASSSNTRRAY ASDWKHFASWCRRQGFSLMPPDPQTVGLYITAQASA SGRDKKSVSTIERRLSSLTWNYSQRGQPLDRKDRHI ATVMAGIRNKHASPPRQKEAILRDDLVAMLETLDRG SLRGLRDRAMLLLGFAGGLRRSEIVGLDVARDQTED GRGWIEILDKGMLVSLRGKTGWREVEIGRGSSDATC PIVALETWMKFARIAHGPVFRRVTGQSKAVGADRLK DQEVARLVKRAALAAGVRGDLPEGERGQKFAGHSLR AGLASSAEVDERYVQKQLGHASAEMTRKYQRRRDRF RVNLTKASGL*	ACATCGAGCGGCTCCGCGACGAACCGCGCGATGT	Chelativorans sp_ BNC1 plasmid 3	(A) 15 (B) 18
YP_957160.1 MNENSHKKPPDLTLRNEGSAVSIHMESEALRHYLQA ATTDNTRKAYRSAIRQFEKWGGRLPTDRDTVVRYLL SKAKSLNSRTLNLHLTAIGQWHHYQGITDPVRDPLV RKTMDGIRRTHGQPKRKAKALRLEHIAQMVKHLQRL PDCNKKYRDIAMVLTGFFGAFRRSELVAIRVSDLIW EPEGLIIKMPRSKTDQEAEGLMRALPFGDVAVCPVQ ALKSWLEEAEIREGPVFRPVNRWDQIQPRPLTPSSI NDLLKALGKACDFDFIHELSSHSFRRGLSTSAARER IDFELIKKQGGWRSDATVWAYVEEGQQLSENAAVVL MEKLQALMKPEPNQEHSTGAIIE*	CTAACCCACGATA**ATCAATCT**TATCGCGGGTTAA	Marinobacter aquaeolei VT8 plasmid pMAQU02	(A) 14 (B) 20
NP_943161.1 MSIICGTHGLNRRFVMTAGNNDENLPTRRHEEPTVL ARTPGTLTTPEQLAEQHQRFLAAATTDNTRRTYRSA IRHFLAWGGVLPCDEAALIRYLLSFAEVLNPRTLAL RLTALSQWHRYQGFPDPTASATVGKTLRGIERVNGR PRQKAKALVLEDLERIVVHLNTLDGLATLRDSALLQ VGYFGAFRRSELVTLEMQYLEWEQEGLRITLPRSKT DQEGEGLDKAIPYGDSICCPATALRRWLDAAQIVQG PLFRRISRWGVLGEVALHEGSVNTILTARAEAAGLL YVPELSSHSLRRGLATSAHRAGADFLEIKRQGGWRH DGTVHGYIEEAGAFEENAAGSLLRRKP*	TTGACCCACGATA**AGCGCGGT**TATCGTGAGTTAA	Pseudomonas sp. ND6 plasmid pND6-1	(A) 18 (B) 21
ZP_05884863 (Vika) MTDLTPFPPLEHLEPDEFADLVRKAIKRDPQAGAHP AIQSAISHFQDEFVRRQGEWQPATLQRLRNAWNVFV RWCTHQGIPALPARHQDVERYLIERRNELHRNTLKV HLWAIGKTHVISGLPNPCAHRYVKAQMAQITHQKVR ERERIEQAPAFRESDLDRLTELWSATRSVTQQRDLM IVSLAYETLLRKNNLEQMKVGDIEFCQDGSALITIP FSKTNHSGRDDVRWISPQVANQVHAYLQLPNIDADP QCFLLQRVKRSGKALNPESHNTLNGHHPVSEKLISR VFERAWRALNHETGPRYTGHSARVGAAQDLLQEGYS TLQVMQAGGWSSEKMVLRYGRHLHAHTSAMAQKRRQ R	AATAGGTCTGAGA**ACGCCCAT**TCTCAGACGTATT	Vibrio coralliilyticus ATCC BAA-450	(A) 100 (B) 26

Supplementary Table 3. List of the Cre residues contacting DNA and their corresponding residues in Vika, grouped into 3 categories. The residues before the slash relate to Cre and after the slash to Vika, respectively. Group I contains the catalytic residues. Group II covers the residues interacting specifically either with the minor groove (mG) or the major groove (MG). In group III are the residues interacting with the phosphate backbone (non-specifically) of the DNA. Red labeling highlights the conserved residues in the alignment. In orange are semi-conserved residues. In blue are the conserved and semi-conserved residues identified in a 3D alignment between the Vika model and the Cre template.

Catalytic Residues	Specific Interaction		Non specific Interaction
	mG	MG	
R173/R153	K201/K181	H40/-	F37/W18
H289/H270	R243/K223	K43/R25	S38/Q19
R292/R273	K244/R224	K86/R65	T41/T22
W315/W296	R282/-	Q90/K69	M44/-
Y324/Y305		R259/K246	S47/N28
			R50/-
			R81/R59
			L83/L63
			A84/H64*
			T87/T67
			M97/K77
			R100/-
			R101/-
			R118/-
			R121/H102
			K122/R106
			A131/I111
			K132/R110*
			R154/R133
			Q156/Q137
			R159/R139
			I174/K154
			A175/N155
			R241/R221
			V242/V222
			N245/S225
			S257/S244
			A260/-
			E262/E245*
			K276/R266*
			Y283/-
			S287/T268
			G288/G269
			R326/R307



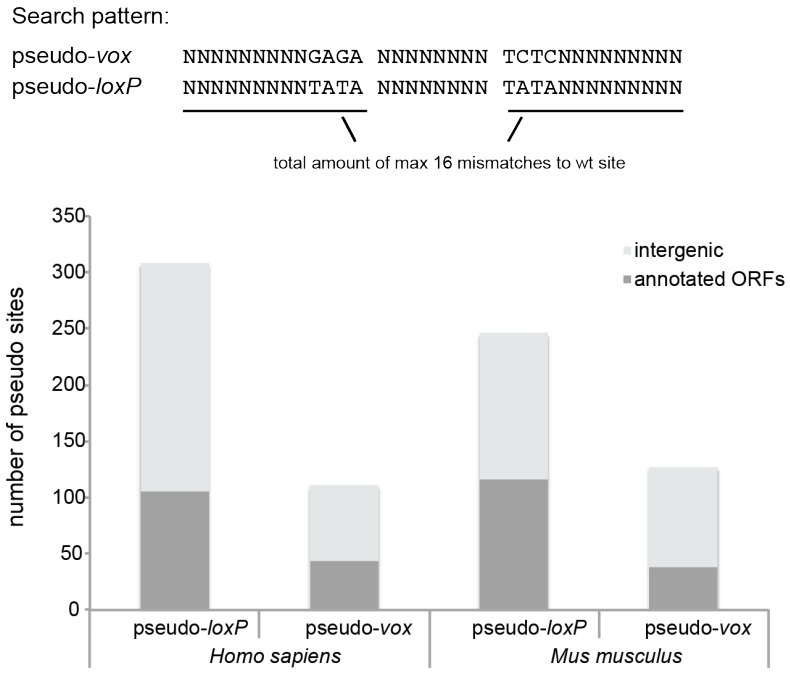
Supplementary Figure 1. Amino acid alignment of Vika and VCre. The ClustalX coloring scheme was used to highlight the conservation between the two recombinases.



Supplementary Figure 2. L(+)-arabinose-induced expression of the Vika recombinase in *E.coli*. Western blot shows concentration-dependent (0, 1, 10, 200 mg/ml L(+) arabinose) increase of the protein amount in the lysate of the bacterial cells. Cells grown in the presence of an empty vector served as control. A non-specific cross-reacting protein served as loading control (lower panel).

Supplementary Table 4. Sequences of the SeLOX-predicted recombination sites for Vika recombinase. Search was done in 10 kb genome sequence surrounding gene coding for Vika. Putative spacer sequences are depicted in bold.

lox-like candidate site	Nucleotide sequence
lox-1	ACAAAAAAAGAGG**CCAATCGG**CCTCTTTTTTTGT
lox-2	TTATTCGATTTAG**TAAAATA**CCAAAACGAATAA
lox-3	CTGCTCGTACCGG**GGCCAC**CCGGTACGAGCCT
lox-4(vox)	AATAGGTCTGAGA**ACGCCCAT**TCTCAGACGTATT
lox-5	AACCGTGTTCTTT**CATCGT**CAAAGCACACCGT
lox-6	CTAGGCGGTGGGT**TGACAC**ACCCACTGCCTAT



Supplementary Figure 3. Nucleotide analysis of the mouse and human genomes for the presence of the *loxP*- and *vox*-like sequences according to the search pattern depicted (see Results section for description).

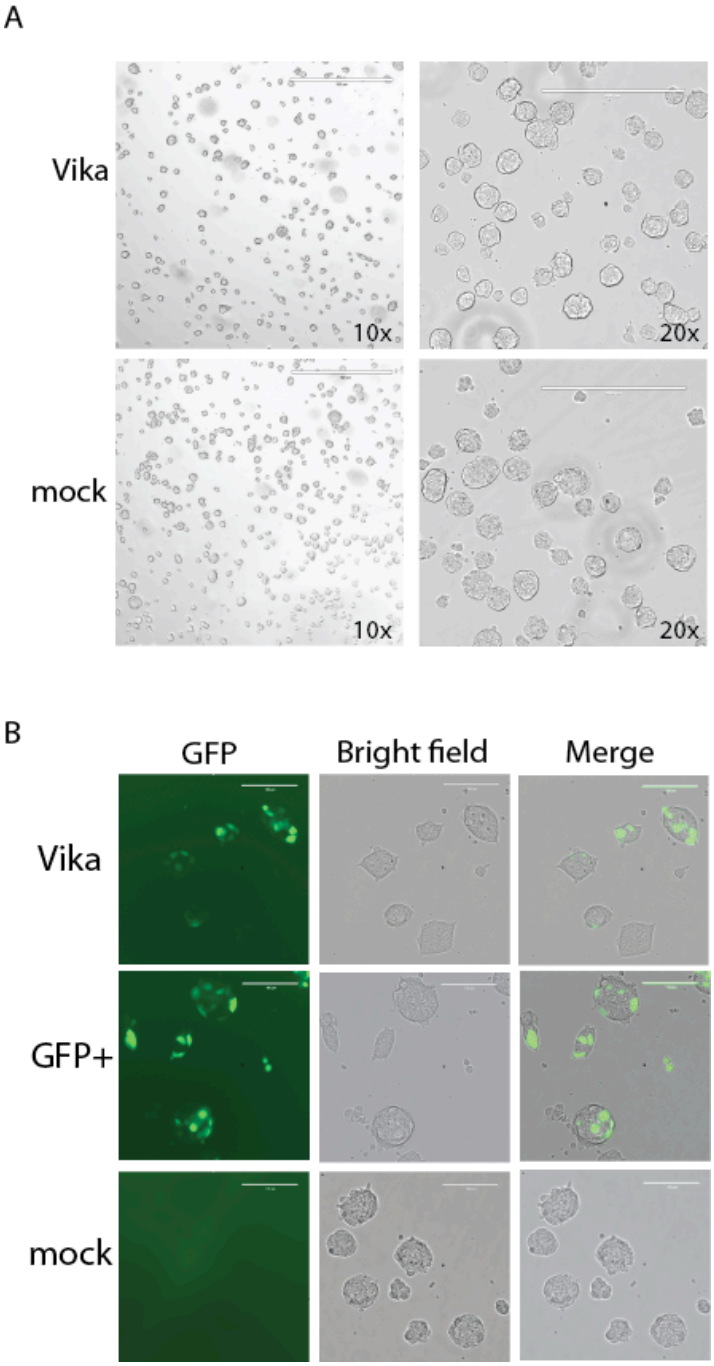
vox-like cryptic chromosomal sites

Site	Alignment	#of mutations
voxCH18	CAGAGCTCTGAGA **CTTTGTGT** TCTCAAAGATATC	7
vox	AATAGGTCTGAGA **ACGCCCAT** TCTCAGACGTATT	
	* ** ******* * ***** * ***	
voxCH21	AAGAGGTGTGAGA **CTGAATTT** TCTCAGTCAGGTT	6
vox	AATAGGTCTGAGA **ACGCCCAT** TCTCAGACGTATT	
	** **** ***** * * ****** * **	
*vox*CHX	AAGAGAACTGAGA **AAATATTC** TCTCAGAGGGAAT	6
vox	AATAGGTCTGAGA **ACGCCCAT** TCTCAGACGTATT	
	** ** ****** * ******* * * *	
voxCMp92	AACAGCTTTGAGA **GCTGTTGC** TCTCAGCTGAATT	6
vox	AATAGGTCTGAGA **ACGCCCAT** TCTCAGACGTATT	
	** ** * ***** * ****** * ***	
voxCHp3	AACAGGACTGAGA **TAAAACAG** TCTCAGACAGCAT	6
vox	AATAGGTCTGAGA **ACGCCCAT** TCTCAGACGTATT	
	** *** ****** ** ******** *	

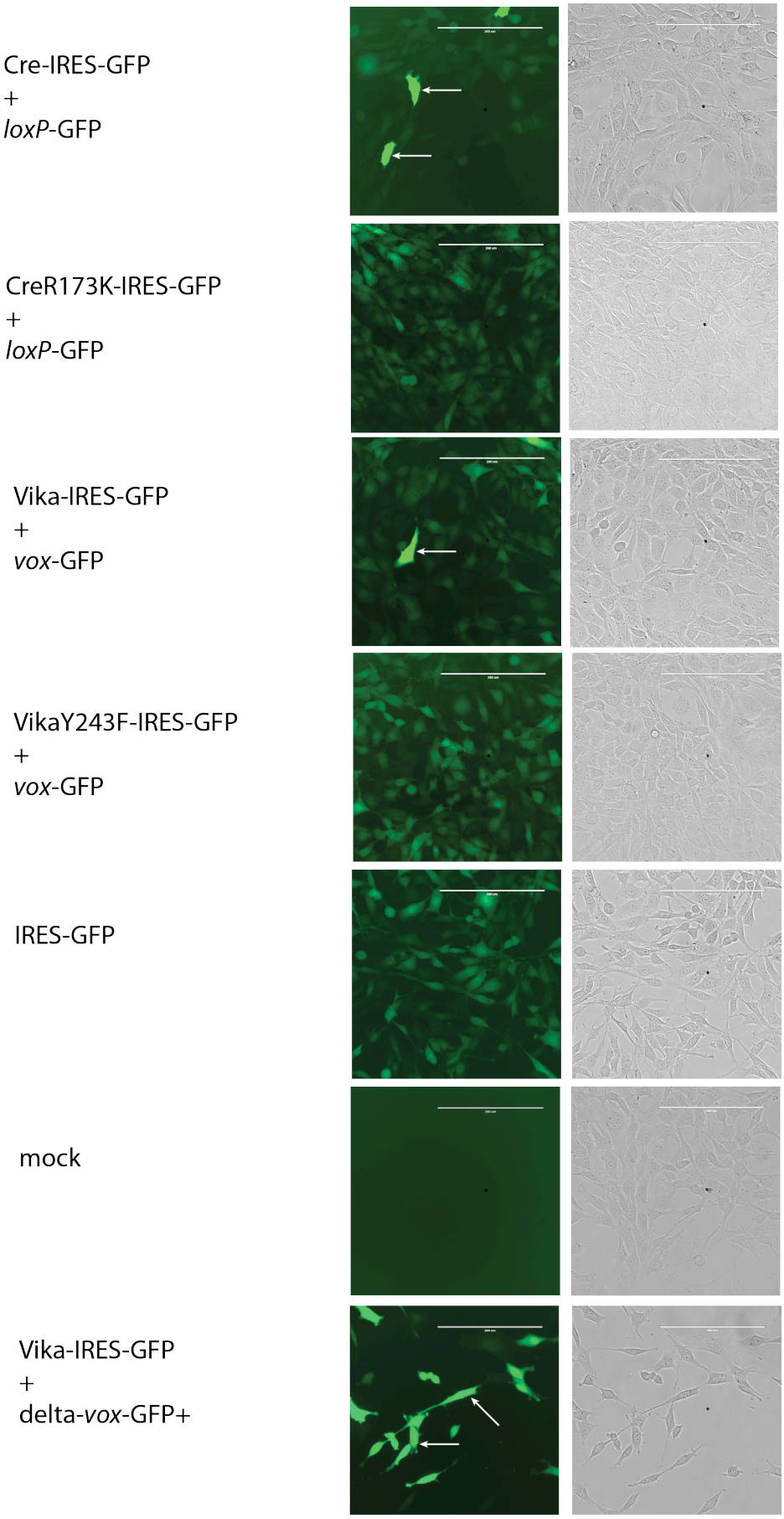
loxP-like cryptic chromosomal sites

loxhXp22	ACAACCATTTATA **ATATATAA** TATATGATGTTAT	7
loxP	ATAACTTCGTATA **GCATACAT** TATACGAAGTTAT	
	* *** **** *** * **** ** *****	
loxM5	GTAACTGAGTATA **TGCATATA** TATACGTATATAT	5
loxP	ATAACTTCGTATA **GCATACAT** TATACGAAGTTAT	
	***** ***** ****** * ***	

Supplementary Figure 4. Nucleotide sequence alignment of the cryptic chromosomal target sites compared to wild-type *vox* or *loxP*. Capital H or M in the name of cryptic *vox* sites indicates human or mouse origin. Putative spacer sequences are depicted in bold. Mismatches in the chromosomal sites are highlighted in red. Total count of mismatches in the inverted repeats is depicted on the right hand side of the alignment.



Supplementary Figure 5. Evaluation of prolonged expression of Vika in mouse ES cells. **(A**) mES cell line with stably integrated Vika recombinase after prolonged passaging (24 days). A representative photo of a clonal culture is depicted. (**B**) Recombination activity of stably expressed Vika recombinase in mES cell line. Images show cells 24 hours after transfection with *vox*-GFP reporter plasmid. Note the apparent Vika-mediated recombination signified through GFP expression. A control of the recombined reporter plasmid (GFP+) was transfected for detecting transfection efficiency.



Supplementary Figure 6. Recombination test of NIH3T3 cells infected with viruses expressing respective recombinases. Reporter plasmids specific to the recombinases carry recombination sites (*loxP* or *vox*). eGFP is expressed upon recombination. A control reporter plasmid (delta-*vox*-GFP+) constitutively expressing eGFP was used as transfection control. Cells were imaged for EGFP expression (green), 24 hours after cotransfection with reporter plasmids. Cells with reporter-mediated eGFP expression are marked with an arrow.