

Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity

Janet Karpinski^{1,2,11}, Ilona Hauber^{2,11}, Jan Chemnitz^{2,11}, Carola Schäfer^{2,3}, Maciej Paszkowski-Rogacz¹, Deboyoti Chakraborty¹, Niklas Beschorner², Helga Hofmann-Sieber^{2,3}, Ulrike C Lange²⁻⁴, Adam Grundhoff^{2,3}, Karl Hackmann⁵, Evelin Schrock⁵, Josephine Abi-Ghanem⁶, M Teresa Pisabarro⁶, Vineeth Surendranath⁷, Axel Schambach⁸, Christoph Lindner⁹, Jan van Lunzen^{2,3,10}, Joachim Hauber^{2,3} & Frank Buchholz^{1,7}

Current combination antiretroviral therapies (cART) efficiently suppress HIV-1 reproduction in humans, but the virus persists as integrated proviral reservoirs in small numbers of cells. To generate an antiviral agent capable of eradicating the provirus from infected cells, we employed 145 cycles of substrate-linked directed evolution to evolve a recombinase (Brec1) that site-specifically recognizes a 34-bp sequence present in the long terminal repeats (LTRs) of the majority of the clinically relevant HIV-1 strains and subtypes. Brec1 efficiently, precisely and safely removes the integrated provirus from infected cells and is efficacious on clinical HIV-1 isolates *in vitro* and *in vivo*, including in mice humanized with patient-derived cells. Our data suggest that Brec1 has potential for clinical application as a curative HIV-1 therapy.

Although substantial advances have been made in the treatment of HIV infection using cART, present therapies do not cure HIV-infected individuals¹⁻³. A small population of latent, transcriptionally silent but still replication-competent viruses persists in long-lived HIV reservoirs, such as resting memory CD4⁺ T cells^{4,5}. As current treatment does not target the integrated provirus, viral rebound is observed when medication is discontinued⁶. The challenges of lifelong cART include drug toxicity⁷, the emergence of drug resistance^{8,9}, adverse drug interactions¹⁰, strict treatment adherence^{11,12} and high costs¹³. Furthermore, a substantial number of patients on cART still exhibit increased levels of immune activation, along with associated long-term adverse effects¹⁴. Given the limitations of cART, there is a growing interest in developing novel approaches to control or preferably cure an HIV infection. Several strategies to excise the integrated provirus using advanced genome engineering tools have been described¹⁵⁻¹⁸. However, none of these target the majority of HIV-1 isolates and efficiently removes the provirus without causing undesired side effects. Our previous work had established the feasibility of redirecting the DNA binding specificity of Cre recombinase to generate Tre recombinase that recombines a sequence present in the LTR (*loxLTR*) of a primary HIV-1 isolate¹⁵. However, the *loxLTR* sequence is not very conserved among different HIV-1 clades and less than 1% of HIV-1 infected patients could potentially benefit from an application of Tre recombinase. Hence, although providing proof of the concept, Tre recombinase is not suitable to eradicate the provirus from most HIV-1 infected individuals.

We now report the development and application of the broad-range recombinase 1 (Brec1) that is active against the majority of HIV-1 primary isolates. Brec1 was evolved by directed molecular evolution and, as a Cre-type recombinase, is characterized by several properties that particularly favor its application in future therapies.

RESULTS

Target site selection

To identify a highly conserved target sequence in HIV-1 long terminal repeats (LTRs) we employed the SeLOX algorithm¹⁹ and screened sequence libraries in the Los Alamos HIV Sequence Databases (<http://www.hiv.lanl.gov/>) for appropriate candidates. We selected a sequence, which we called the Brec target region (*loxBTR*) due to its high conservation rate (90%) throughout the major HIV-1 subtype groups A, B and C, and its remote similarity to *loxP* and target sequences of related site-specific recombinase systems (**Fig. 1a**)^{15,20-22}.

The *loxBTR* sequence is located in the R region of HIV-1 LTRs (**Fig. 1b**) and is a 34-bp asymmetric sequence displaying 32% sequence homology to *loxP*. As 94% (1,297 out of 1,386 isolates) of subtype B viruses, 85% (745 out of 876) of subtype C isolates, and 80% (158 out of 198) of subtype A isolates match the exact *loxBTR* sequence (**Fig. 1c** and **Supplementary Table 1**), this sequence seems to be highly conserved and thus less susceptible to mutational changes than most other sequences within the HIV-1 genome.

¹Medical Systems Biology, UCC, Medical Faculty Carl Gustav Carus, TU Dresden, Germany. ²Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany. ³German Center for Infection Research (DZIF), partner site Hamburg, Germany. ⁴Department of Anesthesiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ⁵Institute for Clinical Genetics, University Hospital and Medical Faculty Carl Gustav Carus, TU Dresden, Germany. ⁶Structural Bioinformatics, BIOTEC TU Dresden, Germany. ⁷Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany. ⁸Institute of Experimental Hematology, Hannover Medical School, Hannover, Germany. ⁹Agaplesion Diakonieklinikum, Gynecology, Hamburg, Germany. ¹⁰Infectious Diseases Unit, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ¹¹These authors contributed equally to this work. Correspondence should be addressed to J.H. (joachim.hauber@hpi.uni-hamburg.de) or F.B. (frank.buchholz@tu-dresden.de).

Received 30 April 2015; accepted 21 December 2015; published online 22 February 2016; doi:10.1038/nbt.3467



Figure 1 Highly conserved recombinase target site *loxBTR* in HIV-1 LTRs. (a) Brec1 target site (*loxBTR*) sequence aligned to the Cre target site (*loxP*) and related recombinase target sites. Sequence differences in the enzyme-bound half sites, compared to *loxP*, are marked in red. The spacer sequences are shown in gray. (b) Schematic representation of proviral DNA with LTRs and the location of the Brec1 target site *loxBTR*. (c) Representations (%) of the target sequences shown in a in HIV-1 subtypes A, B, C (blue discs), and of *loxBTR* in the three individual main subtypes.

Directed evolution of broad-range HIV-1 recombinases

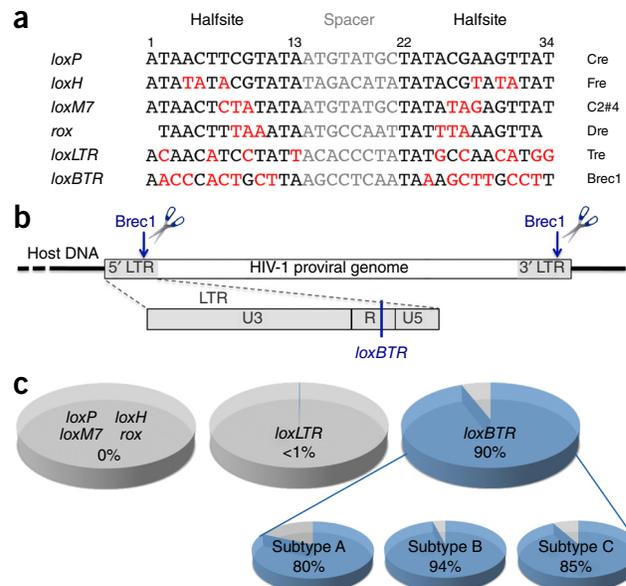
The *loxBTR* site differs in 23 out of the 34 nucleotides of the Cre recognition site *loxP* and 11 alterations in the inverted repeats had not previously been addressed in studies to alter Cre DNA binding specificity^{15,20–23}. This divergence necessitated initiating the substrate-linked evolution process^{20,24} with six different subsites (1A–C and 2A–C), which were successively collapsed during the directed evolution process (Fig. 2a and Supplementary Fig. 1). As a starting library we pooled all available recombinase coding sequences from previous Cre-related evolution experiments^{15,20,25} and inserted them into the evolution vector pEVO containing the starting target sites 1A–C and 2A–C (Fig. 2a and Supplementary Fig. 1). Recombinases with activity on the respective sites recombined the plasmid and thereby removed a unique NdeI restriction site. After NdeI digestion of the isolated plasmid DNA, active recombinase coding sequences were selectively retrieved by PCR and ligated back into the initial pEVO plasmids (Supplementary Fig. 1b). After accumulation of functional recombinases the libraries were pooled and shuffled, and the sites were progressively collapsed to ultimately probe recombinases on the final *loxBTR* sequence (Fig. 2a and Supplementary Fig. 1).

A recombinase library possessing good activity on *loxBTR* was obtained after 126 evolution cycles. To reduce potential off-target activity, evolution cycling was continued for 18 additional generations with counter-selection against recombination of related target sites (Fig. 2a). Finally, the library was transduced into human Jurkat T cells constitutively expressing the recombinase for 4 weeks to ensure that long-term expression is well tolerated (Fig. 2a). Thus, 145 evolution cycles were required to obtain mutants with the desired properties.

We investigated the molecular evolution process at nucleotide resolution and identified evolutionarily conserved mutations by performing deep sequencing on the final library. A total of 331,000 reads of Brec sequences were obtained and aligned to Cre in order to visualize the evolution process (Fig. 2b). The high resolution afforded by these results highlight the divergent evolvability of different regions in the protein. Furthermore, the results unmasked mutations that had been fixed during the evolution process, with 26 residues (N3, V7, P12, P15, M30, H40, M44, S51, Y77, K86, G93, S108, C155, N160, A175, R241, K244, A249, R259, E262, T268, D278, P307, N317, N319, I320) being altered with a mutation frequency of 85% or higher (Fig. 2b and Supplementary Table 2).

To investigate frequently appearing alterations in detail, we mapped the mutations onto the Cre/*loxP* crystal structure^{26,27}. We identified four regions as likely critical for changes in target site specificity (Fig. 2c). These regions are located at the protein–DNA binding interface in the Cre/*loxP* system, and they evolved differently compared to previous Cre evolution studies^{20,21,25}, including the previously evolved Tre recombinase¹⁵ (Fig. 2b,c). Indeed, identification of these four mutational hot spots substantially enhances our understanding of how recombinases evolve to target different sequences and should accelerate future efforts to produce enzymes with novel specificities.

From the final library (cycle 145) we tested more than 100 individual clones for their recombination properties in *Escherichia coli*. One clone (Brec1, Supplementary Fig. 2) was selected for



subsequent experiments based on its promising recombination behavior, showing efficient recombination of the *loxBTR* sequence without recombining *loxP*.

Brec1 specifically targets *loxBTR* and variant sites in bacteria

We first tested the recombination efficiency of Brec1 on its target site upon different induction levels in *E. coli*. Brec1 recombination progressively increased, with full recombination observed at the highest induction level (Fig. 3a). Second, we assessed Brec1 recombination on different target sites (Fig. 3b). Brec1 efficiently recombined the *loxBTR* sites, whereas no recombination was detected in plasmids carrying *loxP*, *loxLTR*, or SINΔlox, a *loxBTR*-like site present in self-inactivating (SIN) lentiviral vectors^{28,29} designed for Brec1 delivery into mammalian cells. Third, we performed experiments using a blue/white reporter assay to signal recombination (Fig. 3c). Consistent with the above results, Brec1 also proficiently recombined *loxBTR* without showing cross-reactivity on target sites of related recombinase systems, demonstrating that Brec1 efficiently and specifically recombines *loxBTR* in bacteria.

Despite the high sequence conservation of *loxBTR*, some viruses with mutations in this target sequence might escape recognition by Brec1. To assess the recombination behavior of Brec1 on *loxBTR* mutants, we selected four recurrently occurring mutations in the *loxBTR* sequence from the Los Alamos HIV Sequence Database (Fig. 3d) and tested the ability of Brec1 to recombine these sites. Notably, the activity of Brec1 on these variants was similar to its activity on *loxBTR* (Fig. 3e), suggesting that certain point mutations, in particular those found in primary isolates, do not compromise the recombination proficiency of Brec1.

Brec1 specifically recombines *loxBTR* in mammalian cells

To examine the recombination properties of Brec1 in mammalian cells, HeLa cells were co-transfected with recombinase expression and reporter plasmids. Recombinase activity was evaluated 48 h after transfection by microscopy (Fig. 4a) and quantified by fluorescence-activated cell sorting (FACS) (Fig. 4b). In the fluorescence-based reporter assay, monomeric red fluorescent protein (mCherry) was constitutively expressed from the nonrecombined reporter plasmid, whereas enhanced green fluorescent protein (EGFP) was expressed only upon excision of the mCherry encoding sequence by recombination.

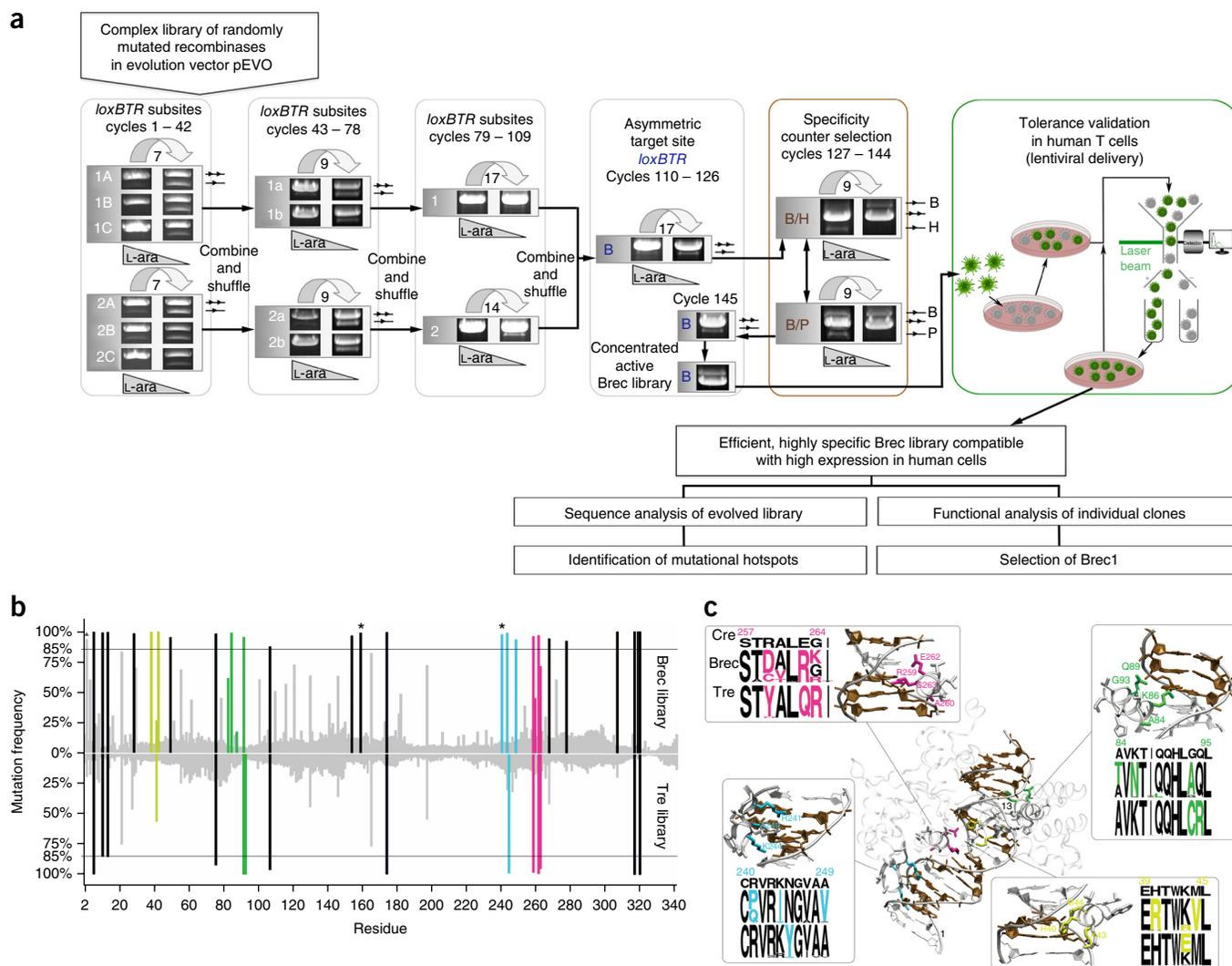


Figure 2 Brec1 evolution. **(a)** Temporal scheme of the evolution process. The recombinase library activities of the first and last cycle of the target sites are shown as images of respective agarose gels. The number of evolution cycles performed on each subset is presented below the arrow. The nonrecombined (line with two triangles) and the recombined (line with one triangle) bands are indicated beside the gels. B, recombination through *loxBTR*; H, recombination through *loxH*; P, recombination through *loxP*. Triangles labeled L-ara indicate reduced levels of L-arabinose for progressive rounds of evolution. **(b)** Deep sequencing of the final Brec library. Each bar represents an amino acid mutation compared to the Cre sequence and its frequency of occurrence (%) in the Brec library. Bars of evolutionarily conserved mutations (>85%) during the directed evolution process are colored black with mutational hotspots shown in yellow, green, blue or magenta, correlating with the colors used in **c**. Asterisks mark high occurrence of a mutation in a specific site but with two different possible amino acids. The triangle marks high occurrence of mutations in amino acid position three, but with no specific amino acid conservation. All conserved amino acid mutations are listed in **Supplementary Table 2**. A comparison to the mutations during Tre evolution¹⁵ is provided. **(c)** Structural analysis of mutational hotspots. The structure of Cre bound to *loxP* is depicted with the four mutational hotspot regions shown enlarged. Frequently mutated residues are shown in magenta, green, blue and yellow, as in **b**. Alignments to Cre and Tre are shown with amino acid frequencies represented by the size of the letters. *loxP* base positions 1 and 13 as shown in **Figure 1a** are labeled.

As in *E. coli*, Brec1 efficiently and site-specifically recombined its *loxBTR* target sites, with no cross-reactivity on target sites of related recombinase systems (**Fig. 4a,b**).

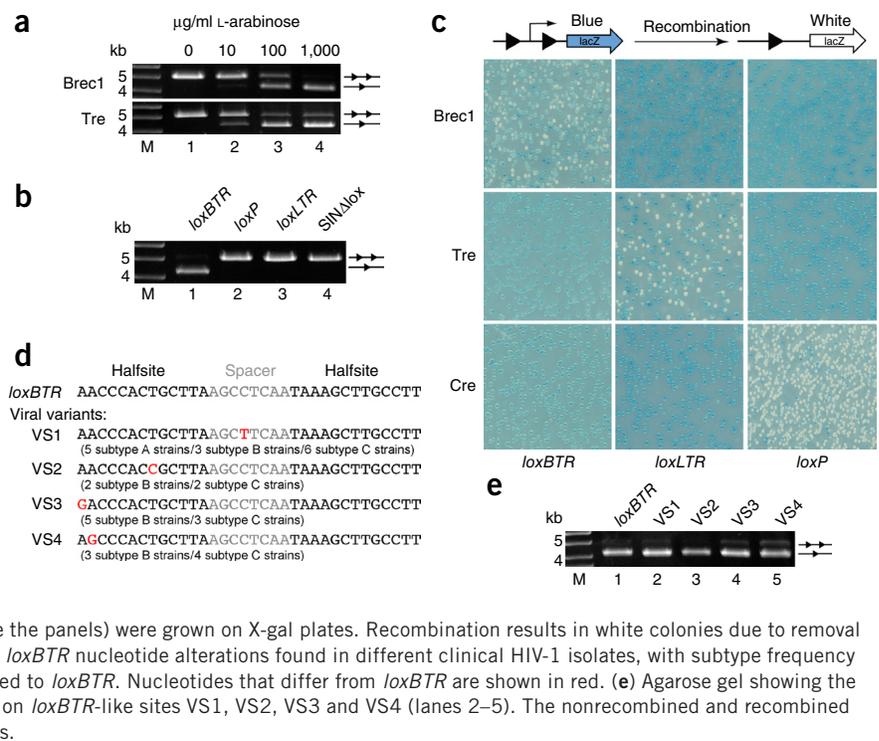
To investigate whether Brec1 can also recombine *loxBTR* sites present in the genome of eukaryotic cells, we tested recombination of stably integrated reporter constructs in HeLa cells (**Supplementary Fig. 3**). Moreover, we analyzed Brec1 activity in HeLa reporter cells that contained stably integrated full-length HIV-1 genomes²⁹ (**Fig. 4c**). Precisely excised circular recombination products were observed only in cell cultures that were transduced with lentiviral or alpharetroviral Brec1- and GFP-expressing vectors (LV-Brec1 or RV-Brec1), but not in cells treated with negative control vectors (LV-Ctr or RV-Ctr; **Fig. 4d,e**).

We next tested whether Brec1 can inhibit HIV-1 in human PM1 T-lymphocyte cultures, a human T-cell line routinely used in HIV-1 research, infected with replication-competent virus. Viral particle release was abrogated specifically in Brec1-expressing cells (**Fig. 4f**), documenting the antiviral effect of Brec1 on HIV-1 infection. Thus, Brec1 efficiently and specifically recombines *loxBTR* in mammalian cells and possesses anti HIV-1 activity in cultured cells.

Safety evaluations

To investigate potential side effects of the Brec1 approach, we performed experiments to assess conceivable cytotoxic, cytopathic or genotoxic effects. First, we constitutively expressed Brec1 in Jurkat T cells by lentiviral vector gene transfer (LV-cBrec1) for up to 10 weeks

Figure 3 BreC1 recombination is efficient and specific in bacteria. **(a)** Recombination efficacy of BreC1 on *loxBTR*. *E. coli* cells harboring the pEVO vector (**Supplementary Fig. 1b**) containing the BreC1-encoding sequence and target *loxBTR* sites were grown at indicated L-arabinose concentrations (lanes 1–4). Recombination of isolated plasmid DNA was assayed by restriction enzyme digestion, resulting in a smaller fragment for recombined (one triangle) and a larger fragment for nonrecombined plasmids (two triangles). For comparison, recombination of Tre recombinase on *loxLTR* is shown in the lower panel. **(b)** Activity of BreC1 on *loxBTR* (lane 1, positive control), related recombinase target sites (lanes 2, 3) and a mutated *loxBTR* site present in a lentiviral expression vector designed for BreC1 delivery in mammalian cells (lane 4). *E. coli* cells harboring the pEVO vector containing the BreC1-encoding sequence and indicated target sites were grown at 1 mg/ml L-arabinose. M, DNA marker. **(c)** LacZ-based recombination reporter assay in *E. coli*. Cells co-transformed with recombinase expression and target site reporter plasmids (as indicated beside the panels) were grown on X-gal plates. Recombination results in white colonies due to removal of the promoter driving lacZ expression. **(d)** Frequent *loxBTR* nucleotide alterations found in different clinical HIV-1 isolates, with subtype frequency indicated below each sequence. Sequences are aligned to *loxBTR*. Nucleotides that differ from *loxBTR* are shown in red. **(e)** Agarose gel showing the activity of BreC1, induced with 1 mg/ml L-arabinose, on *loxBTR*-like sites VS1, VS2, VS3 and VS4 (lanes 2–5). The nonrecombined and recombined bands are indicated as a line with one or two triangles.



and assayed BreC1's potential impact on cell growth and cell cycle progression (**Supplementary Fig. 4a,b**). Likewise, primary human CD4⁺ T lymphocytes were transduced with BreC1 and analyzed with respect to gene expression, apoptosis and immune activation and function (**Supplementary Fig. 4c–f**). No significant difference between BreC1 and control transduced cells were observed in any of these experiments, indicating that prolonged expression of BreC1 is well tolerated in the host cell. Furthermore, analyzing the capacity of BreC1-vector-transduced CD34⁺ hematopoietic stem cells (HSC) to differentiate into various hematopoietic lineages using colony forming unit (CFU) assays demonstrated that BreC1 did not alter the capacity of human HSC to differentiate into the expected lineages (**Supplementary Fig. 4h**).

Second, we analyzed potential BreC1-related genotoxic or off-target effects. We screened the human genome for *loxBTR*-like sites using the SeLOX algorithm¹⁹. Six human genomic sites (HGS 1–6) with the highest sequence similarity to *loxBTR* (**Supplementary Fig. 5a** and **Supplementary Table 3**) were tested for recombination proficiency in *E. coli*. None of these sites were recombined (**Supplementary Fig. 5b**), indicating that the human genome does not contain obvious *loxBTR*-like sequences that are targeted by BreC1. We also tested transduced CD4⁺ T cells for potential genomic alterations by spectral karyotyping (SKY) and array-comparative genomic hybridization (array-CGH) and found no chromosomal alterations in BreC1-expressing cells (**Supplementary Fig. 5c** and data not shown). Furthermore, we performed analyses to investigate potential recombination between multiple *loxBTR* sites integrated in the genome, reflecting integration of two viruses in one cell. No recombination of distant *loxBTR* sites was detected in this scenario (**Supplementary Fig. 6**), indicating that BreC1-mediated genomic deletions in cells harboring several proviruses are rare. Moreover, we performed whole genome sequencing totaling over 1.3 billion reads on primary CD4⁺ T cells, derived from an HIV-1 patient, that had been transduced

with LV-BreC1 or LV-Ctr, and found no signs of inadvertent BreC1-mediated genetic alterations (**Supplementary Table 4**).

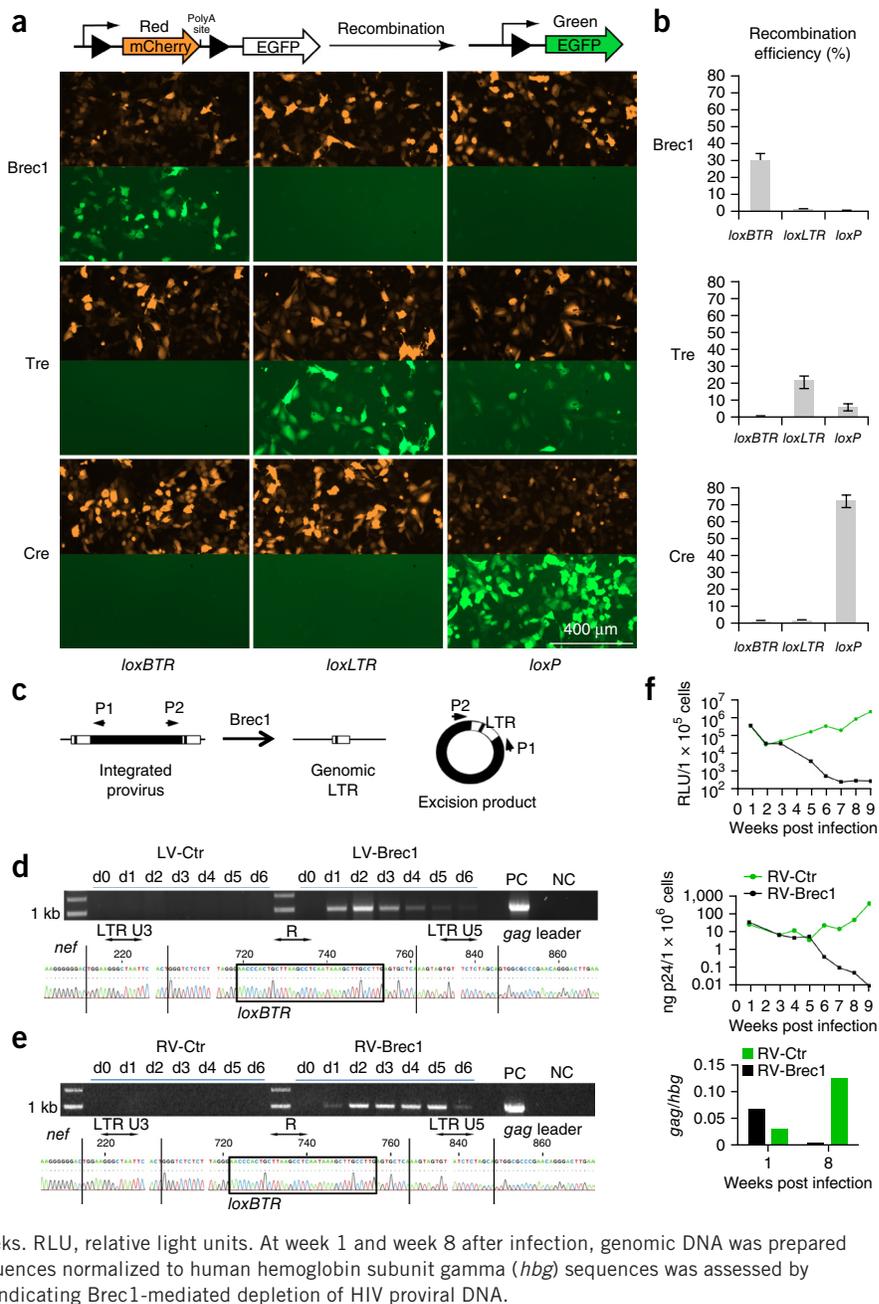
Finally, we generated transgenic mice that express BreC1 from a constitutive promoter (**Supplementary Fig. 7**) and monitored the health of the animals at frequent intervals. The BreC1 allele was inherited at a Mendelian ratio with animals remaining healthy over a period of more than 18 months, demonstrating that BreC1 expression is well tolerated in an organism. Together, the data signify that BreC1 expression is well tolerated in human cells and in a model organism without apparent cytopathic, cytotoxic or genotoxic effects.

BreC1-mediated provirus ablation in patient-derived cells

In antiviral gene therapy research, vector-transduced human hematolymphoid cells are frequently infected with established HIV-1 strains to subsequently monitor viral load^{16–18,29–32}. However, in a clinical setting BreC1 would have to function in an already infected subject. To mirror this situation we isolated CD4⁺ T cells from the peripheral blood of an HIV-1-infected patient and transduced the cells with LV-Ctr or LV-BreC1 (**Fig. 5a**) and subsequently cultured the cells for 20 d (for patient details, see Online Methods). In LV-Ctr-transduced cells, viral loads increased and the number of transduced cells decreased over time (**Fig. 5b**), implying that viral expansion compromised the growth and viability of LV-Ctr-transduced cells. In contrast, the same cells transduced with LV-BreC1 showed decreased viral loads and stable amounts of transduced cells over time (**Fig. 5b**), indicating BreC1-mediated provirus excision in the patient's cells. Deep sequencing of genomic DNA isolated from the cells after the 20-d culture period indeed revealed that the proviral DNA was efficiently depleted in the BreC1-treated cells (**Fig. 5c**). These data demonstrate that BreC1 can efficiently excise proviral DNA of primary HIV-1 isolates *in vitro*.

To test efficacy of this approach in an *in vivo* model, we 'personalized' immunodeficient mice by transplanting Rag2^{-/-}γc^{-/-}

Figure 4 Brec1 recombination in mammalian cells. **(a)** Fluorescence-based recombination reporter assay. Images of HeLa cells co-transfected with indicated recombinase expression and fluorescent reporter plasmids are shown. Prior to recombination, mCherry is constitutively expressed whereas EGFP is silent, resulting in only red fluorescence. Recombination at the target sites (triangles) leads to excision of the mCherry encoding sequence, allowing the promoter to drive expression of EGFP. Images shown are always from the same field of view. **(b)** Cells as shown in **a** were analyzed by FACS and recombination efficiency measured as the ratio of green cell counts to red cell counts. Error bars, mean \pm s.d. of three independent experiments. **(c)** Schematic depiction of Brec1-mediated recombination. Provirus-specific PCR primers (P1 and P2) permit detection of the transient circular excision product. Upon recombination, a single LTR remains at the integration site. **(d)** Genomic DNA isolated from HIV-1 infected HeLa cells transduced with LV-Ctr or LV-Brec1 was analyzed at the indicated days after transduction by PCR using primers P1 and P2 to detect the transient circular excision product (1 kb; upper panel). PC, positive PCR control; NC, negative PCR control. The LTR region was subsequently sequenced, demonstrating the presence of a single LTR flanked by *nef* and *gag*-leader sequences. Brec1 activity resulted in precise *loxBTR* recombination (boxed; lower panel). **(e)** Analysis of BTR recombination upon alpharetroviral Brec1 delivery. Cells were transduced with RV-Ctr or RV-Brec1 and analyzed as described in **d**. **(f)** Analysis of antiviral Brec1 activity in human PM1 T lymphocytes. Cells were transduced with RV-Ctr or RV-Brec1 (MOI = 10). At day 13 after transduction, 1×10^6 cells of each culture were infected with 500 ng of the CCR5-tropic HIV-1_{BaL}Luc2, where the *nef* coding region is substituted by a luciferase open reading frame. Infection was monitored by quantifying luciferase activity (upper panel) as well as the amount of p24 antigen in the supernatant (middle panel) over 9 weeks. RLU, relative light units. At week 1 and week 8 after infection, genomic DNA was prepared from both cultures and the amount of HIV-1 *gag* sequences normalized to human hemoglobin subunit gamma (*hbg*) sequences was assessed by quantitative real-time PCR (lower panel, bar chart), indicating Brec1-mediated depletion of HIV proviral DNA.



animals (hu-Rag2) with CD4⁺ T cell pools derived from an HIV-1 patient and transduced with either LV-Ctr or LV-Brec1. Human cell engraftment was verified by FACS analysis of peripheral blood mononuclear cells (PBMC) 4–8 weeks after transplantation. Viral loads and human CD45⁺CD4⁺ cells were monitored over time. Data obtained from the individual animals revealed robust viremia in the plasma of the animal into which control vector-transduced cells were transplanted (Fig. 5d). By contrast, in the mouse engrafted with Brec1-transduced patient-derived cells, plasma viral loads continuously declined over time to below the limit of detection (<20 HIV-1 RNA copies/ml) at week 18–21 after transplantation (Fig. 5d). When the Brec1 animal was euthanized at week 21, analysis of human cells revealed an enrichment of human (CD45⁺CD4⁺) and vector-transduced (CD4⁺GFP⁺) lymphocytes, particularly in the spleen and liver, indicating that Brec1-expressing cells had been

enriched over time in these organs. Mouse transplantation experiments with cells derived from two additional HIV-1-positive donors showed similar progression, validating that this effect is not patient or animal specific (Supplementary Fig. 8). Collectively, these experiments show that Brec1 can target the provirus in cells isolated from HIV-1-positive patients.

Brec1 activity in animals with stem cell-transplants

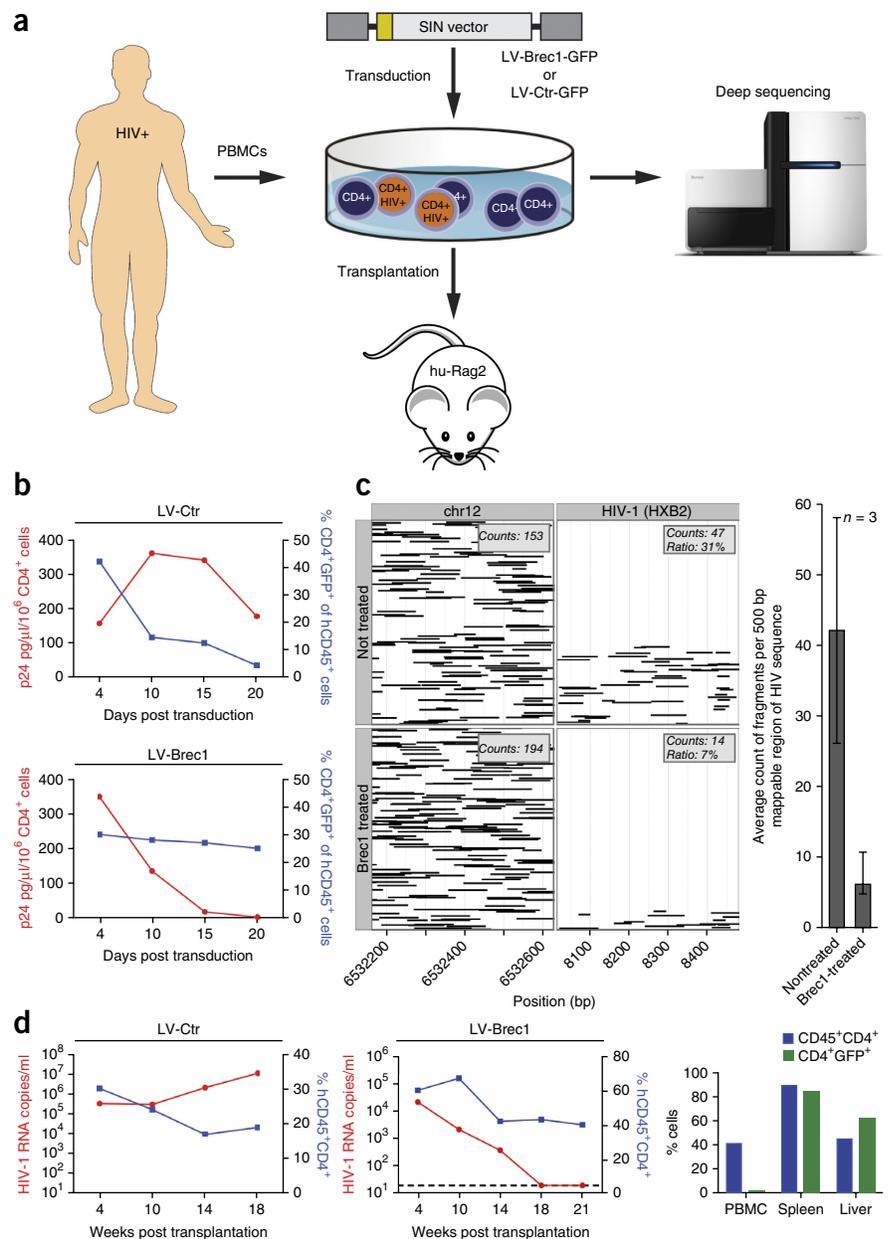
Anti-HIV gene therapies may also include genetic modification of CD34⁺ HSC^{33,34}. Therefore, in another *in vivo* approach to investigate Brec1-based antiviral effects, newborn NOD.Cg-Prkdc^{scid}IL2rg^{tm1Wjl}/SzJ (nonobese diabetic/severe combined immunodeficient/gamma; NSG) mice were irradiated and given transplants as previously described²⁹ by intrahepatic injection of LV-Ctr- or LV-Brec1-transduced human CD34⁺ HSC. In these experiments the transduction

Figure 5 Brec1 deletes the HIV-1 provirus from patient-derived primary T cells. **(a)** Schematic representation of the workflow, with important steps highlighted by arrows. **(b)** Primary CD4⁺ T cells (3×10^6) were isolated from HIV-1 patient-derived PBMC and transduced with 1×10^7 infectious particles of LV-Ctr or LV-Brec1. Viral load (red dots) and percentage of transduced human CD4⁺GFP⁺ cells (blue squares) were determined at the indicated time points. **(c)** Deep sequencing of genomic DNA isolated at day 20 after transduction from the cultures shown in panel **b**. Individual sequencing reads from 500-bp regions on chromosome 12 (control) and of the HIV-1 genome (HXB2) are shown as lines for nontreated and Brec1-treated cells, with the counts and ratios indicated. Quantification of average counts for nontreated and Brec1-treated samples is shown. Error bars, mean \pm s.d. of three nonoverlapping regions. **(d)** BALB/c Rag2^{-/-} γ c^{-/-} (hu-Rag2) mice were given transplants by i.p. injection of 1×10^6 lentiviral vector (LV-Ctr or LV-Brec1)-transduced human patient-derived CD4⁺ T cells. Viral load (red dots) and percentage of human CD4⁺CD4⁺ cells (blue squares) were monitored over time. Percentage of human cells in PBMC, spleen and liver of the Brec1-expressing animal was determined at necropsy by FACS using single-cell suspensions (bar chart).

rate typically resulted in $\sim 20\%$ GFP⁺ cells. Human cell engraftment was verified by FACS analysis at 16 weeks after transplantation and animals with $>10\%$ human CD45⁺ lymphocytes in their peripheral blood were challenged by intraperitoneal injection of the replication-competent CCR5-tropic HIV-1_{JR-CSF} strain (dose equivalent to 200 ng of p24 antigen)³⁵. For LV-Brec1-treated mice, as opposed to the control animal that received LV-Ctr-transduced HSC, the viral load prominently declined over time (**Fig. 6a**). After these animals were euthanized, FACS analysis again revealed a strong accumulation of transduced cells in various tissues, demonstrating that the decline in virus loads observed in the Brec1-treated animals was not due to loss of the human cell engraftment.

To more directly demonstrate antiviral Brec1-mediated effects in humanized mice, additional animals engrafted with either Brec1-vector, or control vector transduced HSC were euthanized at week 24 after HIV-1 infection and human CD4⁺ T cells were collected from the spleen to prepare genomic DNA. As before (**Fig. 4f**), the proportion of HIV-1 *gag* sequences normalized to human *hbg* sequences was reduced in the Brec1-transduced cells as quantified by real-time PCR, indicating Brec1-mediated depletion of HIV proviral DNA *in vivo* (**Fig. 6b**). Likewise, immunohistochemical analysis of human CD3⁺ and HIV-1 p24 antigen-expressing cells in spleens of representative mice demonstrated that p24⁺ cells were distinctly depleted in animals that had received LV-Brec1-transduced HSC (**Fig. 6c**).

Altogether, the combined *in vivo* experiments reveal substantial antiviral activity of Brec1 at the organism level.



DISCUSSION

Complete elimination of replication-competent HIV, including latent viral reservoirs, may be the only way to achieve a genuine cure. We show that Brec1 can remove the provirus from HIV-1-infected cells by specifically recognizing and recombining a highly conserved target site (*loxBTR*) located within the LTRs of the majority of HIV-1 isolates. Over 72% of infected individuals worldwide fall into the HIV-1 group M subtypes A, B or C³³, of which 90% are predicted to contain the exact *loxBTR* sequence. The *loxBTR* sequence is also conserved in most other HIV-1 subtypes with an overall representation of 82% (**Supplementary Table 1**), indicating that HIV-1 provirus in more than 28 million affected individuals would be a Brec1 target.

The fact that Brec1 can also recognize *loxBTR*s carrying point mutations (**Fig. 3**) might further increase the potential patient range. This feature might also lower concerns about the development of resistance, because Brec1 would still recognize at least some mutations arising in an infected individual. Moreover, within the experimental

Figure 6 Brec1-mediated antiviral effects in NSG mice engrafted with human CD34⁺ HSC. (a) Newborn NSG mice engrafted with LV-Ctr– (negative control) or LV-Brec1–transduced unselected human cord blood–derived CD34⁺ HSC pools were infected with HIV-1. Viral load (red dots) and percentage of human CD45⁺CD4⁺ cells (blue squares) were monitored over time. Human cells in PBMC, bone marrow (BM), the spleen and liver of the respective animal was determined at necropsy by FACS using single-cell suspensions (bar chart). (b) Analysis of antiviral Brec1 activity in NSG mice at week 24 after HIV-1 infection. Human CD4⁺ T cells were isolated from the spleen of animals that were engrafted with either LV-Ctr– or LV-Brec1–transduced HSC, and genomic DNA was prepared for subsequent HIV-1 *gag*- and human *hbg*-specific real-time PCR. The ratio of *gag/hbg* sequences in two animals from each engraftment, LV-Ctr and LV-Brec1, is shown. (c) Immunohistochemical analysis of spleen sections prepared from mice engrafted with either LV-Ctr– or LV-Brec1–transduced HSC and stained for human CD3 and HIV p24 antigen. Scale bars, 600 μ m.

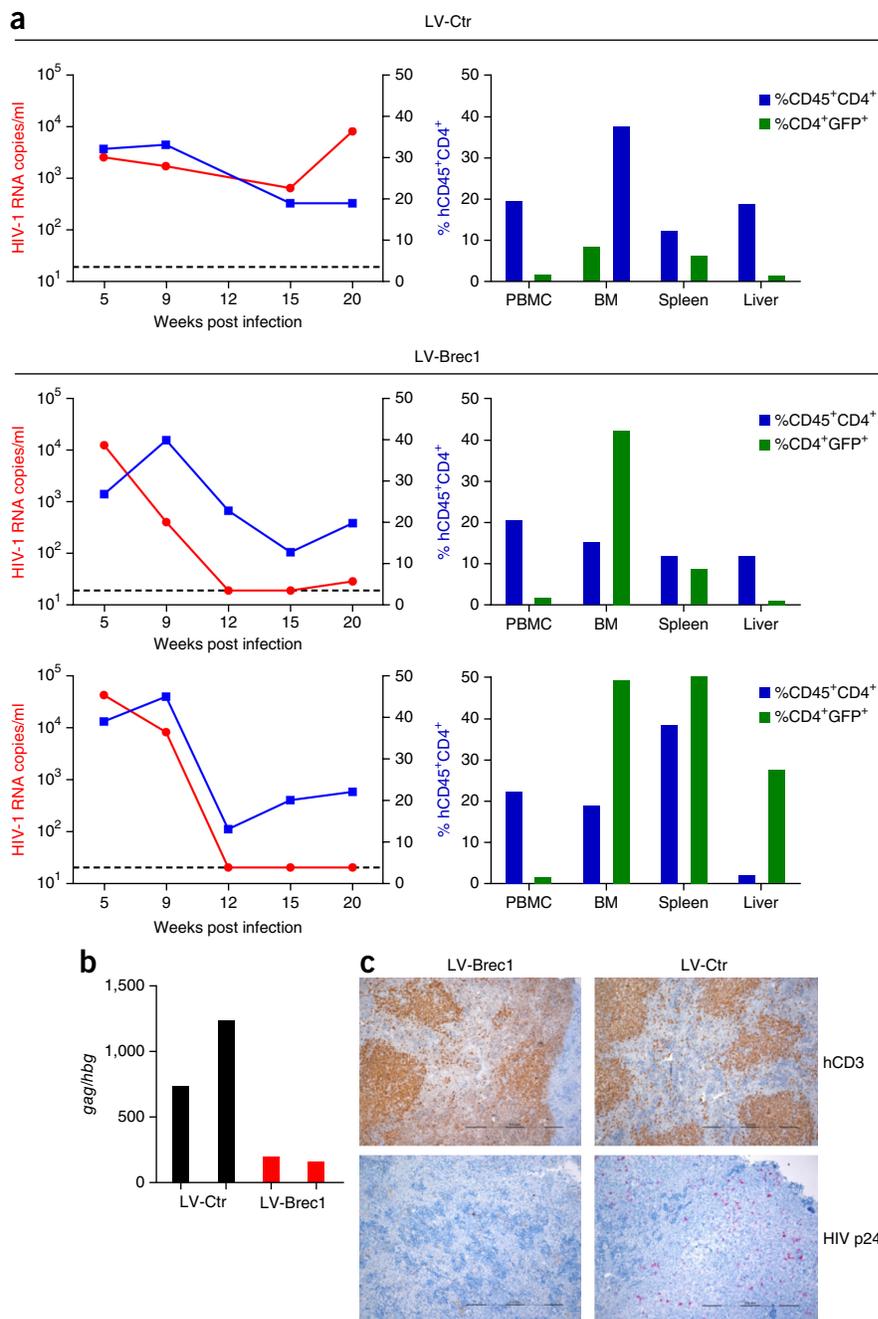
time window of up to 21 weeks, resistant HIV-1 did not emerge in patient-derived cells treated with Brec1 (Fig. 5), supporting the notion that rapid virus escape might not occur.

Brec1 was generated using substrate-linked directed evolution²⁰ starting from the well-established Cre/*loxP* system. 145 generation cycles were required to evolve an enzyme that efficiently and specifically recombines the asymmetric *loxBTR* site, slightly more cycles than were necessary to evolve Tre recombinase to recombine *loxLTR*¹⁵. Therefore, generating a new enzyme with desired properties is time-consuming. However, one evolution cycle can be completed within 2–3 d, streamlining the total development time for a new enzyme to approximately 1 to 2 years: a short time period compared to other drug development efforts.

Deep sequencing of the final Brec library allowed us to define regions and pinpoint amino acids that are crucial for recombinase specificity. This information together with the discovery of novel, naturally occurring Cre/*loxP*-like recombination systems^{22,23,35} should help accelerating the generation of future recombinases with new properties.

The *loxBTR* sequence has only remote similarity to *loxP* and other recombinase target sites of evolved Cre-based recombinases^{15,20,21}. The successful generation of Cre-like designer recombinases with very different target site specificities suggests that directed evolution is very versatile and a suitable approach to generate tailored custom enzymes.

Compared to the directed evolution approach, other recently developed technologies, such as zinc finger nucleases (ZFN), transcription activator–like effector nucleases (TALENs) or the clustered, regularly interspaced, short palindromic repeats (CRISPR) and its associated Cas9 protein, known as the CRISPR-Cas9 systems, offer



faster development of genome editing tools. In fact, several designer nucleases targeting the HIV-1 co-receptor CCR5 (refs. 30,31,36) or sequences in the LTRs of HIV-1 (refs. 16–18) have been described recently. Although these programmable nucleases are relatively rapidly generated, they also have drawbacks. Off-target cleavage has been reported for such nuclease systems^{21,37–40}, which may result in undesired genomic rearrangements and other genotoxic effects. We did not observe similar off-target activities for Brec1. Although impressive improvements have been made to reduce off-target cleavage by programmable nucleases⁴¹, another fundamental feature may hamper the development of these enzymes for clinical applications. Nuclease cleavage results in double-strand breaks (DSBs), which activate the cellular error-prone nonhomologous end-joining (NHEJ) pathway. Repair through this pathway leads to unpredictable sequence rearrangements after the lesions have been

repaired, with potential detrimental effects. By contrast, site-specific recombinases do not rely on cellular pathways during genome editing. The reactions are solely performed by the recombinase protein itself, with predictable precision and without activating DNA repair pathways. Hence, gene therapy employing site-specific recombinases represents a promising strategy for future clinical applications aiming at deleting proviruses in HIV-1-infected subjects. However, we emphasize that it is not clear whether one system will ultimately prove effective for all patients, and a combination of two or more genome editing approaches may prove expedient. It should also be noted that gene therapy employing site-specific recombinases or other genome editing technologies bear risks. These should be carefully balanced against benefits before treatment is considered.

A potentially precarious event would be a case in which a cell carries two (or more) proviruses. BreC1-mediated recombination of *loxBTR* sites from the different proviruses could delete or invert genomic sequences in the cell, potentially, for example, deleting a tumor suppressor gene. However, several rationales argue against an increased risk for patients. First, single-cell studies have revealed that the majority of infected cells contain only one copy of HIV-1 DNA^{42,43}. Second, recombination between *loxBTRs* from two different proviruses is highly unlikely. In cells that do carry two or more integrated HIV-1 proviruses, deletion of genetic material by BreC1-mediated recombination could theoretically occur. However, the recombination frequency between recombinase target sites negatively correlates with the distance of these sites in the genome⁴⁴. In the rare cases of multiple integrations the distance of the proviruses will almost certainly be bigger than the distance between the two *loxBTR* sites of the individual proviruses. Therefore, recombination between *loxBTR* sites of one integrated provirus will occur at an exceedingly higher rate than recombination between *loxBTR* sites of two separately integrated proviruses. Because recombinase expression in the described system is driven by Tat, BreC1 is not expressed any longer when recombination has deleted the individual proviruses, further reducing the possibility of recombination between *loxBTR* sites of multiple HIV-1 proviruses (**Supplementary Fig. 6**). Of note, this fact distinguishes the recombinase system from the nuclease approach (i.e., ZFN, TALEN or CRISPR-Cas), which will cleave the recognition target sites at equal rates, independent of the distance of the integrated proviruses. Third, even if BreC1-mediated recombination between multiple proviruses would occur and delete genetic material, the consequences are expected to be mild. Virtually all recombination events would cause haploinsufficiency, meaning that the cell would still carry another fully functional chromosome. Haploinsufficiency for most genes does not cause a change in phenotype. However, for some tumor suppressor genes, a haploid state could indeed increase the risk of tumorigenesis⁴⁵. On the other hand, activation of oncogenes through HIV-1 insertion would be much more likely, because activation of oncogenes is neither dependent on haploinsufficiency nor on multiple copies of integrated virus. Thus, BreC1-mediated proviral excision reduces the risk of transformation due to oncogene activation. Therefore, the benefit for a patient of receiving BreC1 most likely exceeds the risks for unwanted side effects due to recombination of multiple proviruses.

In recent years, HIV cure research has repeatedly demonstrated that long-term remission (i.e., durable control of the infection in the absence of cART) or even the complete eradication of HIV-1 is unachievable using a single antiviral strategy^{36,46,47}. Rather, it is believed that future cure strategies will require a combination of different antiviral approaches, including application of small-molecular-weight latency-reversing agents, immune modulating strategies

and possibly gene therapies^{1,48}. Gene therapies might employ the autologous transfer of either antiviral gene-modified CD4⁺ T cells or gene-modified CD34⁺ HSC into HIV-infected subjects⁴⁹. Indeed, both types of gene transfer approaches are currently being tested in clinical trials using various protein or RNA-based antiviral mechanisms⁵⁰. With respect to BreC1, an HSC-based phase 1 trial (i.e., phase 1b/2a) in a small number of HIV patients is conceivable. In such a scenario, CD34⁺ HSC would be mobilized with granulocyte-colony stimulatory factor (G-CSF) and isolated upon apheresis. *Ex vivo* gene transfer using advanced self-inactivating (SIN) LV-BreC1 technology would be followed by autologous reinfusion of the gene-modified HSC into (preconditioned) patients. Engraftment of these HSC is expected to result in HIV-1 host cell progeny that, upon HIV-1 infection, will remove the integrated proviral DNA. This removal process would occur independently of virus tropism due to high conservation of the BreC1 target sequence *loxBTR* (i.e., CCR5-, CXCR4- and dual-tropic viruses are recognized equally well). As the SIN vector used here expresses BreC1 only in Tat-producing cells, HSC-based BreC1 therapy may therefore be ideally combined with novel reservoir purging (e.g., shock and kill) approaches^{51,52}. This not only provides a high margin of biosafety by expressing the antiviral transgene only when needed (i.e., in provirus-containing cells), but, by lowering the proviral load, it may also improve the functional reconstitution of the patient's immune system. Indeed, drug-induced purging and subsequent killing of remaining latently infected cells may greatly benefit from BreC1 technology.

Besides latency-reversing agents, BreC1-based stem cell gene therapy could also be combined with immune-modulatory treatments or with strategies that, for example, simultaneously target the adult peripheral CD4⁺ T cells with antiviral entities. However, even with such advanced combinatory approaches not every latently infected reservoir cell may be reached. Therefore, with respect to eliminating latent infection, direct targeting of BreC1 to reservoir cells *in vivo* would be highly desirable. In fact, new vector developments may indeed allow this in the near future. For example, nonintegrating capsid engineered adeno-associated viral (AAV) vectors, displaying cell type-specific tailored ankyrin repeat proteins (DARPs)⁵³, could be directly injected into patients to deliver BreC1 transiently to CCR7⁺ central memory (T_{cm}) CD4⁺ T cells, a cell population that has been shown to represent a major cellular reservoir for HIV-1 (ref. 54).

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Sequencing data are available from the NCBI SRA under the following BioProject accession codes: [PRJNA288458](#) and [PRJNA287747](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors are indebted to M. Karimova (TU Dresden; TUD) for providing reagents, R. Kuhnert (TUD) for help with mouse work, A. Käßner-Frensel (TUD) for SKY analyses, R. Naumann (Max Planck Institute for Molecular Cell Biology and Genetics; MPI-CBG) for the generation of transgenic mice, S. Winkler (MPI-CBG) and A. Dahl (DFG-Forschungszentrum für Regenerative Therapien Dresden) for DNA sequencing, D. Dubrau (Heinrich Pette Institute) for initial ASLV-vector construction, G. Pilnitz-Stolze, B. Weseloh and B. Abel (Heinrich Pette Institute) for technical assistance, and D. Indenbirken and M. Alawi (Heinrich Pette Institute) for next-generation sequencing of nrLAM-PCR products. This work was supported by the DFG (BU 1400/7-1), the BMBF (GO-Bio FKZ 0315090), the Else

Kröner-Fresenius-Stiftung (2010_A82), and the “Viral Latency” program at the Heinrich Pette Institute–Leibniz Institute for Experimental Virology, Hamburg. The reporter plasmid pCAGGS-*loxP*-mCherry-*loxP*-EGFP, where a floxed mCherry gene was inserted between the CAG enhancer/promoter module and the EGFP gene, was a generous gift from E.M. Tanaka (DFG-Forschungszentrum für Regenerative Therapien Dresden). The mammalian expression vector pCAGGS-IRES-puro was a generous gift from K. Anastasiadis (TUD).

AUTHOR CONTRIBUTIONS

J.K., I.H. and J.C. designed experiments, performed experiments evaluated data and wrote the manuscript. C.S., D.C., N.B., H.H.-S., U.C.L. and K.H. designed experiments, performed experiments and evaluated data. A.G., M.P.-R. and V.S. designed experiments, performed computational analyses and evaluated data. E.S. designed experiments and evaluated data. J.A.-G. and M.T.P. performed structure-based analyses and evaluated data. A.S., C.L. and J.v.L. provided essential reagents. J.H. and F.B. designed the study, directed the study, evaluated data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Passaes, C.P. & Sáez-Cirión, A. HIV cure research: advances and prospects. *Virology* **454–455**, 340–352 (2014).
- Maartens, G., Celum, C. & Lewin, S.R. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet* **384**, 258–271 (2014).
- Fauci, A.S., Marston, H.D. & Folkers, G.K. An HIV cure: feasibility, discovery, and implementation. *J. Am. Med. Assoc.* **312**, 335–336 (2014).
- Mbonye, U. & Karn, J. Transcriptional control of HIV latency: cellular signaling pathways, epigenetics, happenstance and the hope for a cure. *Virology* **454–455**, 328–339 (2014).
- Battistini, A. & Sgarbanti, M. HIV-1 latency: an update of molecular mechanisms and therapeutic strategies. *Viruses* **6**, 1715–1758 (2014).
- Chun, T.W., Davey, R.T. Jr., Engel, D., Lane, H.C. & Fauci, A.S. Re-emergence of HIV after stopping therapy. *Nature* **401**, 874–875 (1999).
- Akay, C. *et al.* Antiretroviral drugs induce oxidative stress and neuronal damage in the central nervous system. *J. Neurovirol.* **20**, 39–53 (2014).
- Richman, D.D. Antiviral drug resistance. *Antiviral Res.* **71**, 117–121 (2006).
- Pinoges, L. *et al.* Risk factors and mortality associated with resistance to first-line antiretroviral therapy: multicentric cross-sectional and longitudinal analyses. *J. Acquir. Immune Defic. Syndr.* **68**, 527–535 (2015).
- Petoumenos, K. *et al.* D:A:D Study Group. Predicting the short-term risk of diabetes in HIV-positive patients: the Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) study. *J. Int. AIDS Soc.* **15**, 17426 (2012).
- Nachega, J.B. *et al.* HIV treatment adherence, drug resistance, virologic failure: evolving concepts. *Infect. Disord. Drug Targets* **11**, 167–174 (2011).
- Rappold, M. *et al.* Treatment modification in HIV-Infected individuals starting antiretroviral therapy between 2011 and 2014. *J. Int. AIDS Soc.* **17** (suppl. 3), 19768 (2014).
- Schackman, B.R. *et al.* The lifetime cost of current human immunodeficiency virus care in the United States. *Med. Care* **44**, 990–997 (2006).
- Beltrán, L.M. *et al.* Influence of immune activation and inflammatory response on cardiovascular risk associated with the human immunodeficiency virus. *Vasc. Health Risk Manag.* **11**, 35–48 (2015).
- Sarkar, I., Hauber, I., Hauber, J. & Buchholz, F. HIV-1 proviral DNA excision using an evolved recombinase. *Science* **316**, 1912–1915 (2007).
- Qu, X. *et al.* Zinc-finger-nucleases mediate specific and efficient excision of HIV-1 proviral DNA from infected and latently infected human T cells. *Nucleic Acids Res.* **41**, 7771–7782 (2013).
- Ebina, H., Misawa, N., Kanemura, Y. & Koyanagi, Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep.* **3**, 2510 (2013).
- Hu, W. *et al.* RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. USA* **111**, 11461–11466 (2014).
- Surendranath, V., Chusainow, J., Hauber, J., Buchholz, F. & Habermann, B.H. SeLOX—a locus of recombination site search tool for the detection and directed evolution of site-specific recombination systems. *Nucleic Acids Res.* **38**, W293–W298 (2010).
- Buchholz, F. & Stewart, A.F. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.* **19**, 1047–1052 (2001).
- Santoro, S.W. & Schultz, P.G. Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. USA* **99**, 4185–4190 (2002).
- Sauer, B. & McDermott, J. DNA recombination with a heterospecific Cre homolog identified from comparison of the *pac-c1* regions of P1-related phages. *Nucleic Acids Res.* **32**, 6086–6095 (2004).
- Karimova, M. *et al.* Vika/vox, a novel efficient and specific Cre/*loxP*-like site-specific recombination system. *Nucleic Acids Res.* **41**, e37 (2013).
- Buchholz, F. & Hauber, J. In vitro evolution and analysis of HIV-1 LTR-specific recombinases. *Methods* **53**, 102–109 (2011).
- 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 (2013).
- Van Duyn, G.D. A structural view of cre-*loxP* site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 87–104 (2001).
- Ennifar, E., Meyer, J.E., Buchholz, F., Stewart, A.F. & Suck, D. Crystal structure of a wild-type Cre recombinase-*loxP* synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation. *Nucleic Acids Res.* **31**, 5449–5460 (2003).
- Suerth, J.D., Maetzig, T., Galla, M., Baum, C. & Schambach, A. Self-inactivating alpharetroviral vectors with a split-packaging design. *J. Virol.* **84**, 6626–6635 (2010).
- Hauber, I. *et al.* Highly significant antiviral activity of HIV-1 LTR-specific tre-recombinase in humanized mice. *PLoS Pathog.* **9**, e1003587 (2013).
- Perez, E.E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808–816 (2008).
- Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 *in vivo*. *Nat. Biotechnol.* **28**, 839–847 (2010).
- Whitney, J.B. *et al.* Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–77 (2014).
- Taylor, B.S., Sobieszczyk, M.E., McCutchan, F.E. & Hammer, S.M. The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* **358**, 1590–1602 (2008).
- Crowell, T.A. *et al.* Hospitalization rates and reasons among HIV elite controllers and persons with medically controlled HIV infection. *J. Infect. Dis.* **211**, 1692–1702 (2015).
- Suzuki, E. & Nakayama, M. VCre/VloxP and SCre/SloxP: new site-specific recombination systems for genome engineering. *Nucleic Acids Res.* **39**, e49 (2011).
- Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).
- Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection. *Nat. Methods* **8**, 765–770 (2011).
- Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* **29**, 816–823 (2011).
- Tsai, S.Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
- Wang, X. *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **33**, 175–178 (2015).
- Kim, H. & Kim, J.S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).
- Josefsson, L. *et al.* Majority of CD4+ T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc. Natl. Acad. Sci. USA* **108**, 11199–11204 (2011).
- Josefsson, L. *et al.* Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS Pathog.* **9**, e1003432 (2013).
- Ringrose, L., Chabanis, S., Angrand, P.O., Woodroffe, C. & Stewart, A.F. Quantitative comparison of DNA looping *in vitro* and *in vivo*: chromatin increases effective DNA flexibility at short distances. *EMBO J.* **18**, 6630–6641 (1999).
- Santarosa, M. & Ashworth, A. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. *Biochim. Biophys. Acta* **1654**, 105–122 (2004).
- Archin, N.M. *et al.* Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* **487**, 482–485 (2012).
- Caskey, M. *et al.* Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* **522**, 487–491 (2015).
- Blankson, J.N., Siliciano, J.D. & Siliciano, R.F. Finding a cure for human immunodeficiency virus-1 infection. *Infect. Dis. Clin. North Am.* **28**, 633–650 (2014).
- Scherer, L.J. & Rossi, J.J. Ex vivo gene therapy for HIV-1 treatment. *Hum. Mol. Genet.* **20**, R100–R107 (2011).
- Herrera-Carrillo, E. & Berkhout, B. Bone Marrow Gene Therapy for HIV/AIDS. *Viruses* **7**, 3910–3936 (2015).
- Deeks, S.G. HIV: Shock and kill. *Nature* **487**, 439–440 (2012).
- Katlama, C. *et al.* Barriers to a cure for HIV: new ways to target and eradicate HIV-1 reservoirs. *Lancet* **381**, 2109–2117 (2013).
- Münch, R.C. *et al.* Off-target-free gene delivery by affinity-purified receptor-targeted viral vectors. *Nat. Commun.* **6**, 6246 (2015).
- Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009).

ONLINE METHODS

Plasmids for substrate-linked protein evolution, recombinase expression and reporter assays. The evolution vectors pEVO-*loxP*, pEVO-*loxH* (pEVO-6) and pEVO-*loxLTR* were described previously^{15,20}. pEVO vectors containing the described coding sequences of recombinases and target sites were generated using standard cloning methods as described previously¹⁵. pEVO-*loxBTR-loxP* and pEVO-*loxBTR-loxH* were each generated in two cloning steps from pEVO-*loxBTR* using the Cold Fusion Cloning technology (System Biosciences) following the manufacturer's instructions. The first cloning step used the primers lox1-F (5'-CTTAGATCCTCATATGGTGCAC-lox-TCTCAGTACAATCTGCTCTGATGC-3') and lox1-R (5'-TGCAGTGTCCGTCGACTGTTCTTA-3') and the restriction enzymes ApaI and AvrII. The second cloning step used the primers lox2-F (5'-CTAGCTTCCCGGCACAATTA-lox-TAGACTGGATGGAGGCGGATAA-3') and lox2-R (5'-ACTTCGGGTCATGAGCC TAGA-3') and the restriction enzyme AseI. The multi-host reporter plasmids pSVpAx and pSV*loxLTR* were described previously¹⁵. pSV*loxBTR* was generated from pSVpAx by replacing the *loxP* sites with *loxBTR* sequences using Cold Fusion Cloning with the primers *loxBTR*-F (5'-CCTACTTCACTAACAACCGGTACAG TTCGA-*loxBTR*-AGCTTGCATGCCTGCAGTGC-3') and *loxBTR*-R (5'-CCAAGCTACTCGCGACCATGGTGGCCCGG-*loxBTR*-GGATCCAGACATGATAAG ATAC-3') and the vector restriction sites AgeI and NruI. The mammalian expression vectors pIRESneo-Cre and pIRESneo-Tre were reported previously¹⁵. pIRESneo-Brec1 was generated in the same manner as described for pIRESneo-Cre and pIRESneo-Tre. The Brec1 coding sequence was codon-optimized for expression in mammalian cells and equipped with a nuclear localization signal (NLS) before ligating it into the vector.

The reporter plasmid pCAGGS-*loxP*-mCherry-*loxP*-EGFP, where a floxed mCherry gene was inserted between the CAG enhancer/promoter module and the EGFP gene, was a generous gift from Elly M. Tanaka (DFG-Forschungszentrum für Regenerative Therapien Dresden). pCAGGS-*loxLTR*-mCherry-*loxLTR*-EGFP and pCAGGS-*loxBTR*-mCherry-*loxBTR*-EGFP were generated from pCAGGS-*loxP*-mCherry-*loxP*-EGFP by replacing the *loxP* sites with either *loxLTR* sequences or *loxBTR* sequences using standard cloning methods.

pCAGGS-Brec1 was built by subcloning the Brec1 coding sequence into the mammalian expression vector pCAGGS-IRES-puro (a generous gift from Konstantinos Anastasiadis)⁵⁵. For expression in mammalian cells or mice, the Brec1 coding sequence was codon-optimized and equipped with a nuclear localization signal (NLS) before ligating it into the above described vectors. All restriction enzymes were purchased from New England BioLabs (NEB).

Generation of recombinase libraries and evolution strategy. A complex library of randomly mutated recombinases was generated by pooling and family shuffling Cre and several known Cre-like recombinases, namely a library of Cre mutants (Cre/Fre)²⁰, a previously generated Tre library¹⁵, Dre²², Zre (isolated from *Salmonella enterica*, GenBank accession number NZ_ABEW01000015), and Shew (*Shewanella* sp. strain ANA-3, GenBank accession number CP000470). To promote the efficiency of family shuffling, Dre and Shew were modified ("Cre-ed") by changing the nucleotide sequence in such a way that they best resembled Cre without resulting in an altered protein sequence. To perform substrate-linked protein evolution, the complex recombinase library was cloned into the evolution vector pEVO-recomb-target, a plasmid that also contained the recombinase target sequence, thereby linking the protein coding information to the substrate (Supplementary Fig. 1b).

The multicycle substrate-linked protein evolution (SLiPE) process was done as described previously^{15,24}. Briefly, the recombinase library was expressed in *E. coli* under the control of an inducible promoter (P_{BAD}) allowing regulation of expression levels. For the first evolution cycle recombinase expression was induced with 200 μ g/ml of L-arabinose (Sigma-Aldrich). By decreasing the L-arabinose concentration cycle by cycle (to 1 μ g/ml), the library was enriched for recombinases with activity at lower expression levels. For each cycle, plasmid DNA was isolated from an induced overnight culture and digested with NdeI to linearize the plasmids that had not been recombined, rendering them unsuitable for the following PCR amplification step. The recombinase coding sequences that had successfully recombined the vector were subsequently amplified by PCR using the primers P1 (5'-CTTACTGTTTCTCCATAC-3')

and P3 (5'-AGGAATAAGGGCGACA-3')^{15,24}. The error-prone Taq polymerase was used to generate more mutations and increase variability of the recombinase library in every cycle. The PCR product was digested with BsrGI and XbaI and subcloned back into the evolution vector for the next evolution cycle. After each third evolution cycle, DNA shuffling was carried out as described previously^{15,24} to rapidly propagate and combine beneficial mutations. Following DNA shuffling, the full-length recombinase coding sequences were amplified by PCR using the primers P1 (see above) and P2 (5'-AATCTTCTCTCATCCGC-3').

The evolution procedure was carried out in a stepwise manner and was first initialized on subset target sites (Fig. 2a) with each of them containing only a few differing nucleotides (Supplementary Fig. 1a). All subsites were generated as symmetric target sites, with one half site containing the required mutations and the other half site being the reverse complement. Asymmetry was applied when shuffling and combining the libraries from subsites 1 and 2 for the evolution cycling steps on the final target site *loxBTR*. After enriching the subset libraries for recombinases with high activity on the subset target sites, they were shuffled to combine advantageous mutations. Once a Brec library with high activity on the *loxBTR* site was obtained, additional evolution cycles selecting against recombination activity on *loxP* and *loxH* were performed alternately to enhance Brec specificity (Fig. 2a). For counter selection against *loxP* and *loxH*, the evolved Brec library was cloned into an evolution vector containing the two *loxBTR* sites intertwined with two *loxH* sites (pEVO-*loxBTR-loxH*) for the first cycle and with *loxP* sites (pEVO-*loxBTR-loxP*) for the second cycle, followed by more cycles alternating *loxH* and *loxP*. Upon induction of recombinase expression, recombination on *loxBTR* results in the removal of the single NdeI site, whereas recombination on *loxH* or *loxP* does not. L-arabinose was applied at 100 μ g/ml during the first cycles and reduced cycle by cycle down to 1 μ g/ml. Plasmid DNA isolated after each evolution cycle was digested with NdeI and the recombinase coding sequences that had successfully recombined *loxBTR* rather than *loxH* or *loxP* were amplified by PCR as described above and subcloned back into the evolution vector for the next evolution cycle. DNA shuffling was carried out after each sixth cycle of counter selection (i.e., three cycles against *loxH* alternating with three cycles against *loxP*). The Brec library obtained was subcloned back into pEVO-*loxBTR* for one more, final cycle.

Cellular screening for well-tolerated recombinases in human cells. To select for recombinases that are well tolerated even at high expression levels in mammalian cells, the active Brec library obtained after 145 cycles was subcloned into a lentiviral vector (described below) for constitutive co-expression of the recombinase and EGFP in human cells. Transduced Jurkat T cells were sorted for EGFP expression by sequential fluorescence-activated cell sorting (FACS) on days 19 and 29 after transduction with increased stringency for high EGFP expression. After another 4 d of culturing following the final sorting, cells apparently expressing well-tolerated recombinases were harvested and genomic DNA extracted using the QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's instructions. The final Brec library was retrieved by PCR and subcloned back into the pEVO-*loxBTR* vector for further analyses and selection of individual recombinase clones. More than a hundred individual clones were isolated and analyzed in *E. coli* for their recombination efficiency and target site specificity.

Next-generation sequencing. Evolved recombinase libraries were sequenced using the Ion Torrent semiconductor system. Whole library recombinase DNA was prepared by BsrGI and XbaI digestion of pEVO-recomb-target plasmid library and purifying the linear 1-kb-sized recombinase library from a preparative agarose gel. Subsequently, the DNA was fragmented enzymatically using the Ion Shear Plus Reagents Kit (Life Technologies) following the manufacturer's instructions, aiming for 200-bp fragments. The fragmented DNA was purified using the QIAquick PCR Purification Kit (Qiagen). After ligation of adapters and amplification by PCR using the Ion Plus Fragment Library Kit (Life Technologies), the library was size-selected with an E-Gel SizeSelect 2% agarose gel (Life Technologies). Following TaqMan assay quantification using the Ion Library Quantitation Kit (Life Technologies), the library was further processed for sequencing by preparing template-positive Ion OneTouch 200 Ion Sphere Particles (ISPs) using the Ion OneTouch 200 Template Kit v2 DL

with the Ion OneTouch System. Finally, the library was sequenced using the Ion PGM 200 Sequencing Kit with the Ion Personal Genome Machine (PGM) system and an Ion 314 chip.

Genomic DNA isolated from T cells of an HIV-1-infected patient was analyzed by deep sequencing using an Illumina HiSeq 2000 platform. The sequencing reads obtained from Brec1-treated and control-treated samples were initially aligned to the human reference genome (hg38 assembly) using bowtie2 version 2.1.0 (ref. 56). A local alignment mode was used with “-no-mixed,” “-no-discordant,” and “-fast” parameter settings. All read pairs that failed to align were then mapped (with a “-sensitive” option setting) to a database of HIV sequences (<http://www.hiv.lanl.gov/>, “Filtered web” alignments) and all reads that remained unmapped were then aligned to a sequence of the recombinase expression provirus. Inspection of the deep sequencing data did not reveal mutations in the *loxBRT* site for the Brec1-treated sample. Successfully mapped reads formed a set of fragments unambiguously originating from the recombinase sequence. To create a set of reads originating from HIV, the last two alignment steps were reversed. Bowtie2 output files processing was handled by samtools version 1.2.1 (ref. 57) and read count statistics were calculated in R version 3.1.2. Sequencing data are available from the NCBI SRA BioProject (<http://www.ncbi.nlm.nih.gov/bioproject/>) under the accession number PRJNA287747.

To detect potential genomic rearrangements caused by Brec1 expression, all full reads that failed to properly align to the human genome were exhaustively searched for the presence of sequences resembling sequences similar to *loxBTR*. This step was performed with patman aligner version 1.2.2 (ref. 58). For the search, the spacer part was replaced with ‘N’s, and additionally, up to ten mismatches were allowed.

Structure-based analyses of mutations. The high-resolution crystal structure of the Cre/*loxP* complex^{26,27} (PDB ID 1Q3U) was used to map in 3D the observed mutations²⁵. Jalview⁵⁹ was used to prepare the alignments of the sequences of Tre and the clones obtained in the Brec library with the sequence of Cre. Four mutational hotspot regions were found and analyzed in detail paying particular attention to those residues differently mutated in each of the three systems (Cre/*loxP*, Tre/*loxLTR*, Brec1/*loxBTR*).

Recombination assays and specificity tests in bacteria. Recombinase expression was induced with 0–1,000 µg/ml L-arabinose. To assay the recombination efficiency of recombinase libraries or individual clones, plasmid DNA (pEVO-recomb-target) from an induced overnight culture was digested with BsrGI and XbaI (NEB), resulting in different fragment sizes for recombined versus nonrecombined substrate on an agarose gel.

To test substrate specificity, Brec1 recombinase was tested for its activity on *loxP*, *loxLTR* and SINΔlox (*loxBTR*-like site present in a lentiviral expression vector designed for Brec1 delivery in mammalian cells) and compared to its activity on *loxBTR*.

The lacZ-based recombination reporter assay used XL1-Blue cells harboring the reporter plasmids pSV*loxBTR*, pSVpaX or pSV*loxLTR*, transfected with recombinase expression plasmids (pEVO-recomb-target) and grown overnight at 1 mg/ml L-arabinose, selecting for both the reporter (ampicillin) and the recombinase expression plasmid (chloramphenicol). To avoid competition between the target sites of the two plasmids, a nonspecific target site was chosen for the expression plasmid. Plasmid DNA was isolated and transfected again into XL1-Blue cells. Cells were subsequently plated on X-gal-containing plates selecting for the reporter plasmid.

Assay for Brec1 activity on *loxBTR*-like sites from viral strain variants. HIV-1 strains with *loxBTR*-like sites as potential targets for Brec1 were identified using the Los Alamos HIV sequence database (<http://www.hiv.lanl.gov/>). The respective *loxBTR*-like target sites were generated by site-directed mutagenesis of pEVO-Brec-*loxBTR* using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) following the manufacturer’s instructions. Recombination activity of Brec1 on the variant sites was tested in *E. coli* as described above. Recombinase expression was induced with 1 mg/ml L-arabinose and recombination assayed as described above. Recombination on *loxBTR* served as a positive control.

Cell culture conditions, transfection and reporter cell line generation. HeLa wild-type (ATCC, Catalog Number CCL-2) and HIV-1-infected cells²⁹ were cultured as monolayers at 37 °C and 5% CO₂ in Dulbecco’s Modified Eagle Medium (high glucose, GlutaMAX, pyruvate) supplemented with 10% FBS (FCS) and 100 U/ml penicillin-streptomycin (Invitrogen Life Technologies).

Jurkat (ATCC, clone E6-1, Catalog Number TIB-152) and PM1 (NIH AIDS Research & Reference Reagent Program, Catalog Number 3038) T cells were cultured at 37 °C and 5% CO₂ at a density of 1 × 10⁶ cells per ml in RPMI 1640 medium (Lonza) supplemented with 10% FCS, glutamate, pyruvate (Biochrom) and 100 U/ml penicillin-streptomycin (Invitrogen Life Technologies). Once a week dead PM1 T cells were separated from intact cells by hyperdensity gradient centrifugation. All cell lines were screened for absence of mycoplasma contamination by multiplex PCR. Cell lines were authenticated by multiplex PCR and STR analyses. Multiplex PCR was used to ensure mycoplasma-free culture conditions.

Patient-derived HIV-1-infected CD4⁺ T cells were negatively selected using magnetic beads as per the manufacturer’s instructions (Human CD4⁺ T cell enrichment kit, STEMCELL Technologies) and cultured in RPMI 1640 supplemented with 10% FCS, with 100 U/ml penicillin-streptomycin, 2 mM L-glutamine and 50 U/ml recombinant human IL-2 (Sigma-Aldrich). For transduction, cultures were pre-stimulated with CD3/CD28 magnetic beads (Invitrogen) for 24 h according to the manufacturer’s instructions. After pre-stimulation, viral particles were added in the presence of 2 µg/ml polybrene (Sigma-Aldrich) and the cells were spinoculated as previously described²⁹. Transduction rates for CD4⁺ T cells transplantation experiments were as follows: **Figure 5d**: LV-Ctr; 2 × 10⁶ CD4⁺ T cells (42% GFP⁺), LV-Brec1; 3 × 10⁶ CD4⁺ T cells (30% GFP⁺); **Supplementary Figure 7**: LV-Ctr; 1 × 10⁶ (25% GFP⁺) and 1 × 10⁶ CD4⁺ T cells (18% GFP⁺), respectively; LV-Brec1: 1 × 10⁶ (23% GFP⁺) and 1 × 10⁶ (25% GFP⁺) CD4⁺ T cells, respectively. T cells were isolated from a total of three patients with long-term control of HIV replication. All patients had chronic established HIV-1 infection (median duration 18 years, range 13–21 years) and were selected based on long-term control of viral replication (>5 years viral load below 50 RNA copies/ml, no confirmed viral blips) and stable immune reconstitution with CD4 counts >500/µl during long-term cART. All patients were male (age: 52–70 years) and had received stable cART for more than 5 years (range: 5–13 years: pt.1 TDF/FTC/EFV; pt.2 TDF/FTC/MVC; pt.3 TDF/FTC/ATZ/r). Pt. 3 participated in a clinical phase 3 study before donating blood in this study. Median nadir CD4 count before starting therapy was 330 cells/µl (range: 328–516 cells/µl) and peak viral loads were in the intermediate range (48,500–280,000 RNA copies/ml).

Preparation of CD34⁺ HSC from umbilical cord blood was performed with the EasySep human cord blood CD34⁺ selection kit (STEMCELL Technologies) and the RoboSep system. For transduction, HSC cultures were prestimulated with the cytokine cocktail StemSpan CC110 (STEMCELL Technologies) and the cultures were subjected to spinoculation as described in detail before²⁹.

HeLa cells were transfected using Effectene (QIAGEN) according to the manufacturer’s instructions. For stable transfection, cells were passaged into selective medium 24 h after transfection.

Reporter HeLa cell lines with stably integrated target sites *loxBTR*, *loxP* or *loxLTR* were generated by transfecting HeLa cells grown in 90 mm dishes with 2 µg of reporter plasmid pSV*loxBTR*, pSVpaX or pSV*loxLTR*, and selection with puromycin (Sigma-Aldrich) at 5 µg/ml.

Recombination assays and specificity tests in mammalian cell culture. Codon-optimized NLS-Brec1 coding sequence was subcloned into a mammalian expression vector, pIRESneo, for constitutive high-level expression of Brec1 from the human cytomegalovirus (CMV) major immediate early promoter/enhancer. For the fluorescence-based recombination reporter assay, HeLa cells were co-transfected with recombinase expression plasmids and double-fluorescent reporter plasmids with exclusively monomeric Cherry red fluorescent protein (mCherry) being constitutively expressed from the non-recombined, and enhanced green fluorescent protein (EGFP) expressed from the recombined plasmid, both driven by the same promoter. The mCherry coding sequence was sandwiched between the two recombination target sites and hence excised upon recombination, allowing the promoter to drive expression of EGFP that was silent before recombination (**Fig. 4a**). HeLa cells were



grown in 6-well plates and transfected with a total of 400 ng of DNA at a ratio of recombinase expression plasmid to fluorescent reporter plasmid of 1:4. Co-transfected cells were assayed by fluorescence imaging using the EVOS FL microscope at 48 h after transfection. In addition, expression of mCherry and EGFP was quantified by fluorescence-activated cell sorting (FACS) using a Becton Dickinson Canto II system equipped with 488, 561 and 640 nm lasers. For live/dead cell discrimination the samples were stained using the LIVE/DEAD Fixable Far Red Dead Cell Stain Kit for 640 nm excitation (Life Technologies). FACS measurements were performed on three technical replicates of each sample. Data were analyzed using BD FACSDiva 8.0 software.

To assay recombination activity in the genomic context, reporter HeLa cell lines with stably integrated target sites *loxBTR*, *loxP* or *loxLTR* were grown in 6-well plates and transfected with 400 ng of recombinase expression plasmid. Genomic DNA was isolated 48 h after transfection using the QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's instructions and assayed for recombination by PCR using the primers F (5'-GCCTCGGCCTAGGAACAGT-3') and R (5'-CCGCCACATATCCTGATCTT-3') (Supplementary Fig. 3).

Off-target recombination analysis. Potential off-target Brec1 recombination sites were identified by screening the human genome using SeLOX¹⁹. The respective genomic sites were generated by site-directed mutagenesis of pEVO-Brec1-*loxBTR* using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) following the manufacturer's instructions. Recombination activity of Brec1 on human genomic sites was tested in *E. coli*. Recombinase expression was induced with 1 mg/ml L-arabinose and recombination assayed as described above. Recombination on *loxBTR* served as positive control.

Viral vectors and production of viral particles. The self-inactivating (SIN) replication-incompetent lentiviral vector (LV) backbone used for delivery of Brec1 and GFP has been described in detail previously²⁹. In the respective vector (LV-Brec1), the gene sequence encoding Brec1 is placed under the control of an engineered Brec1-resistant tandem TAR repeat (2TAR), the *cis*-active target sequence of the HIV-1 Tat *trans*-activator, thereby limiting Brec1 expression to HIV infected cells. In addition, the human phosphoglycerate kinase (PGK) promoter directs constitutive expression of EGFP. For various toxicity analyses, the 2TAR promoter element in LV-Brec1 was substituted by the constitutive elongation factor 1 alpha (EF1 α) promoter, and the PGK promoter (regulating GFP expression) was replaced by the SFFV promoter, resulting in the construct LV-cBrec1. A negative lentiviral control vector, LV-Ctr, was constructed by deleting the 2TAR-*brec1* expression cassette in LV-Brec1. In RV-Brec1 and RV-Ctr, the respective *brec1* and/or *egfp* expression cassettes from LV-Brec1 and LV-Ctr were inserted into an alpharetroviral SIN vector backbone previously described in detail²⁸.

VSV-G pseudotyped lentiviral (LV) and retroviral (RV) particles were produced by transient cotransfection of 293T cells with split-packaging expression plasmids as described^{28,29}. Likewise, titers of the respective viral particles were determined as fluorescent forming units per ml (ffu/ml) exactly as described²⁹.

Transduction and infection of cell cultures. For transduction of PM1 T cells and HIV-infected HeLa cells LV or RV particles were added to the cells at various amounts in the presence of 5 μ g/ml protamine sulfate. Cells were spinoculated at 600g for 90 min at room temperature. Medium was changed 8 h after transduction. PM1 T cells (1×10^6) were infected with 500 ng of HIV-1_{Bal}Luc2 in the presence of 1 μ g/ml polybrene (Sigma-Aldrich) and cells were spinoculated at 600g for 90 min at room temperature. Medium was changed 8 h after infection.

For transduction of patient's HIV-1 CD4⁺ T cells, cell cultures were pre-stimulated with equilibrated anti-CD3/CD28 magnetic beads (Invitrogen) for 24 h at a 1:1 ratio. Following pre-stimulation, various amounts of LV-Ctr or LV-Brec1 particles were added in presence of 5 μ g/ml protamine sulfate and cells were spinoculated at 600g for 90 min at ambient temperature. After 24 h of incubation the transduction procedure was repeated. Prior to further analyses, transduced cells were cultured in the presence of 50 U/ml recombinant human IL-2 at 37 °C and 5% CO₂.

Safety analyses. To determine cell cycle distribution 1×10^6 LV-transduced Jurkat T cells were harvested, washed with PBS, suspended in 500 μ l PBS/1% EDTA and fixed drop-wise with 5 ml of 80% ice cold ethanol. After incubating for 20 min on ice, the cells were incubated for 24 h at -20 °C. Afterwards, the cells were pelleted and rehydrated in 450 μ l of PBS supplemented with 16.6 μ l RNase A (10 mg/ml; Sigma-Aldrich) and 33 μ l propidium iodide solution (0.5 mg/ml; Sigma-Aldrich). Incubation at 37 °C for 30 min, was followed by further incubation at ambient temperature in the dark for 2 h before flow cytometry using a BD FACSCanto system.

Analysis of apoptotic events was performed using the Annexin V FITC Kit (Invitrogen) together with the antibody Annexin V-APC conjugate (Becton Dickinson). For analysis, 5×10^5 LV-transduced primary human CD4⁺ T cells were harvested, stained according to the manufacturer's protocol, and analyzed using a BD FACSCanto (Becton Dickinson) system.

For analysis of immune activation, LV-transduced primary human CD4⁺ T cells were stimulated for 12–24 h with phorbol myristate acetate (PMA) (50 ng/ml final conc.) and ionomycin (0.67 μ M final conc.). Specific cytokine levels were monitored by ELISA, Elispot and intracellular cytokine staining (ICS). Human Th1, Th2, Th17 Cytokine Multi-Analyte ELISArray (Qiagen) was performed with supernatants from 1×10^6 stimulated cells or unstimulated controls, according to the manufacturer's instructions. ICS was essentially performed as previously described²⁹. Mouse anti-human CD3-APC H7 (clone SK7; BD Biosciences, #641397), mouse anti-human CD4-APC (clone SK2; Becton Dickinson, #345771) and mouse anti-human CD154-PE (clone TRAP1; BD Pharmingen, #555700) antibodies were used for surface staining according to the manufacturer's instructions, except that a fourfold excess of the CD154 antibody was added directly to the cells during stimulation. For intracellular staining, mouse anti-human IFN γ -PE Cy7 antibody (clone 4S.B3 BD Pharmingen, #560741) was used. In addition to the cytokine staining, a live/dead staining was also performed using the LIVE/DEAD Fixable Aqua Dead Cell Stain Kit for 405 nm excitation (Life Technologies).

Elispot analysis was performed by coating polyvinylidene plates (96-well; Millipore) with 50 ng of recombinant anti-human IL4 antibody (mouse IgG1; Mabtech Code: 3410-3-250) in phosphate-buffered saline at 4 °C for 12 h. Afterwards, 3×10^3 to 1×10^5 cells were seeded on the coated plates and stimulated with PMA/ionomycin as indicated above. Secreted IL4 was detected using the biotinylated detection antibodies anti-human IL4 (Mabtech; Code: 3410-2A). Antibodies used have been validated previously²⁹.

The differentiation potential of transduced HSC by colony forming unit (CFU) assay was performed using methocult H4435-enriched methylcellulose (STEMCELL Technologies). 100 transduced or mock-treated cells were suspended in 1 ml of methylcellulose and seeded into a 3.5-cm diameter cell culture dish. After incubation at 37 °C and 5% CO₂ for 14 d, various cell colonies were identified and counted.

Spectral karyotyping (SKY) assay. LV-cBrec1-transduced primary human CD4⁺ T cells were arrested in mitosis 7 d after transduction by treating the cells with 0.1 mg/ml colcemid for 4 h. Cells were then treated with 75 mM KCl, incubated at 37 °C for 15 min and fixed in 75% methanol/25% acetic acid. Cell suspension was dropped onto glass slides. Metaphase chromosomes were hybridized with the SKY probe mixture and analyzed using the SpectraCube system (Applied Spectral Imaging) coupled to an epifluorescence microscope (Leica).

Array-comparative genomic hybridization (array-CGH) analyses. LV-cBrec1-transduced primary human CD4⁺ T cells were harvested 7 d after transduction and genomic DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen) for array-CGH analysis. DNA was hybridized against DNA from mock-transduced cells on an Agilent SurePrint G3 Human CGH Microarray Kit 2X400K. The minimum number of probes affected to designate an aberration was set to 3. The median overall probe spacing was 5.3 kb (4.6 kb in RefSeq genes) on the array used.

Generation and characterization of Brec1-expressing transgenic mice. Codon-optimized NLS-Brec1 was subcloned into the mammalian expression vector, pCAGGS-IRES-puro, for constitutive expression of Brec1 from the cytomegalovirus (CMV) immediate early enhancer-chicken beta-actin hybrid

(CAG) promoter. For pronuclear injection of fertilized oocytes, the construct was digested with HindIII and the 6.1-kb fragment containing the CAG-BreC1 sequence purified from a preparative electrophoresis gel. Following injection with the DNA, surviving fertilized eggs were surgically transferred to the oviducts of pseudo-pregnant mice. After mouse pups were born, a small tail biopsy was obtained and genomic DNA extracted for PCR-based genotyping using the BreC1-specific primers mF (5'-AACAAACCGGAAGTGGTTC-3') and mR (5'-TCCTTGATTCTGATCCGGGC-3'). Founder mice were subsequently mated with wild-type mice (C57BL/6). Offspring carrying the transgene were identified by genotyping as described above and analyzed for BreC1 expression by reverse transcription PCR (RT-PCR) from tail biopsy RNA; BreC1 complementary DNA (cDNA) was detected using the BreC1-specific primers mF and mR. For recombination tests, PBMC from 500 µl peripheral blood from C57/Bl6 or BreC1 mice were isolated. Plasma was depleted by centrifugation (1,500g, 2 min, 4 °C) and erythrocytes were lysed with red blood cell lysis buffer (SIGMA) within 15 min on ice according to manufacturer's instruction. PBMC were separated by centrifugation and cells were washed twice with PBS. After seeding of 300,000 cells per well into 96-well plates, Dynabeads^R Mouse T-Activator CD3/CD28 beads were added (30 µl original suspension per well). 6 h later, VSV-G pseudo-typed lentiviral SIN particles carrying the reporter (**Supplementary Fig. 6d**) were added at a multiplicity of infection (MOI) of 20 and cultures were supplemented with 12.5 µg/ml protamine sulfate (SIGMA) and 125 IU/ml rhuIL2 (Biomol). Cells were spinoculated for 10 min at 650g. Afterwards, cells were incubated for 7 d at 37 °C, 5% CO₂, 95% humid. Genomic DNA was isolated from the cells (Qiagen blood DNA kit) and was subjected to PCR analysis to analyze the presence of lentiviral sequences in the genome of the mice cells (primers P3/P4) and to provide evidence for the occurrence of the circularized excision product due to BreC1 enzymatic activity (primers P1/P2).

DNA and RNA extraction from mouse tail biopsy and cDNA synthesis.

Genomic DNA was isolated from tail biopsy by adding 500 µl lysis buffer (100 mM Tris-HCl (pH 8.5), 5 mM EDTA (pH 8.0), 0.2% SDS, 200 mM NaCl) and proteinase K to a final concentration of 100 µg/ml to the sample and incubating overnight at 55 °C with gentle shaking. After pelleting hair and other cellular debris by centrifugation in a microcentrifuge at maximum speed and 4 °C for 15 min, the supernatant was transferred to a new tube and DNA precipitated by adding 500 µl isopropanol and inverting the tube ten times. After storing the DNA precipitate at -80 °C for 30 min, it was pelleted by centrifugation at maximum speed and 4 °C for 20 min. The supernatant was removed and the DNA pellet dried and resuspended in 100 µl nuclease-free water. To help the DNA dissolve it was incubated at 55 °C for 2 h.

Total RNA was purified from tail biopsy using the RNeasy Protect Mini Kit (Qiagen) following the manufacturer's instructions.

First-strand cDNA was synthesized by RT-PCR from mRNA using SuperScript III RT (Invitrogen) following the manufacturer's instructions.

Generation and analysis of humanized mice. To generate "personalized" Rag2^{-/-}γc^{-/-} (hu-Rag2) mice engrafted with lentiviral vector (LV-Ctr or LV-BreC1)-transduced patient-derived CD4⁺ T cells, 6-week-old male or female animals were preconditioned by intra-peritoneal (i.p.) injection of 60–80 µl of Clophosome-A-Clodronate Liposomes (FormuMax Scientific Inc.). 48 h later, animals were irradiated using a dose of 4 Gy (at 4 h before transplantation) from a cesium 137 source at 3.75 Gy/min (CSL-12; Conservatome). Subsequently, 1 × 10⁶ lentiviral vector (LV-BreC1 or LV-Ctr)-transduced cells in 150 µl PBS containing 0.1% human AB serum (PAN Biotech GmbH) were transplanted into mice by i.p. injection. Animals that showed engraftment with >5% human CD4⁺ cells were included, whereas animals displaying >20% weight loss were excluded from the study. Analysis of human cell engraftment was verified by FACS analysis of peripheral blood samples at 4–21 weeks after transplantation, using retro-orbital sampling. Likewise, viremia was assayed by diluting cell-free mouse plasma with human serum (PAN Biotech GmbH) using the ultrasensitive (<20 HIV-1 RNA copies/ml) Cobas AmpliPrep/Cobas TaqMan HIV-1 Test version 2.0 (Hoffmann-La Roche Ltd.). Randomization and blinding were not performed for these and the following animal experiments.

For analysis of peripheral cells, 50 µl to 100 µl of blood was collected from the retro-orbital venous sinus (r.o.) into 100 µl bleeding-buffer (PBS plus 10 mM EDTA) and red blood cells were lysed by treatment with Red Blood Cell Lysing Buffer (Sigma-Aldrich). The white blood cell pellet was resuspended in FACS-buffer (PBS containing 2% FCS and 2 mM EDTA) and stained with monoclonal antibodies. Single-cell suspensions of spleen and liver were prepared at necropsy by manual tissue dissection and filtering through a sterile 70-µm nylon mesh (BD Biosciences) for antibody staining and FACS analysis. Stained cells were analyzed in FACS-buffer plus 1% paraformaldehyde using a FACSCanto (Becton Dickinson) system with BD FACSDiva Software v5.0.3 and FlowJo software v7/9 for PC (Treestar). To monitor human cell engraftment, r.o. collected cells were stained with monoclonal antibodies raised against mouse CD45.2 (clone 104; eBioscience, #47-0454), human CD45 (clone H130; eBioscience, #12-0459), and human CD3 (clone UCHT1; Beckman Coulter, #A94680) (eBioscience Inc.). HIV-1 infected mice were analyzed by staining with monoclonal antibodies directed against the human antigens CD45 (H130) and CD4 (clone RPA-T4; eBioscience, #25-0049-42). Isotype antibodies and cells obtained from mice without transplants served as negative staining controls.

Animals into which human CD34⁺ HSC were transplanted were generated by injecting newborn NOD.Cg-Prkdc^{scid}IL2rg^{tm1Wjl}/SzJ (NSG) mice intra-hepatically (i.h.) with 3 × 10⁵ lentiviral vector (LV-Ctr or LV-BreC1)-transduced CD34⁺ cells in 30 µl PBS containing 0.1% human AB serum. Prior to i.h. injection (i.e., at 4 h before transplantation), the newborns were irradiated with 1 Gy as before. Engraftment was verified by FACS analysis of peripheral blood samples at week 16 after transplantation. Viremia was assayed by TaqMan analysis as described before. Antibodies used have been validated previously²⁹.

Immunohistochemical analysis. Formalin-fixed and wax-embedded sections were analyzed. Deparaffinized sections were incubated in citrate buffer in an 85 °C waterbath overnight for human CD3 antigen detection. Monoclonal anti-human CD3 (Dako M7254, clone F7.2.38) was used in a 1:1,000 dilution. Biotinylated anti-mouse monoclonal antibody in combination with horseradish peroxidase streptavidin was used for visualization. The TNB-Amplification Kit (Dako) and diaminobenzidine were used as substrates. Sections were counterstained with haemalumn.

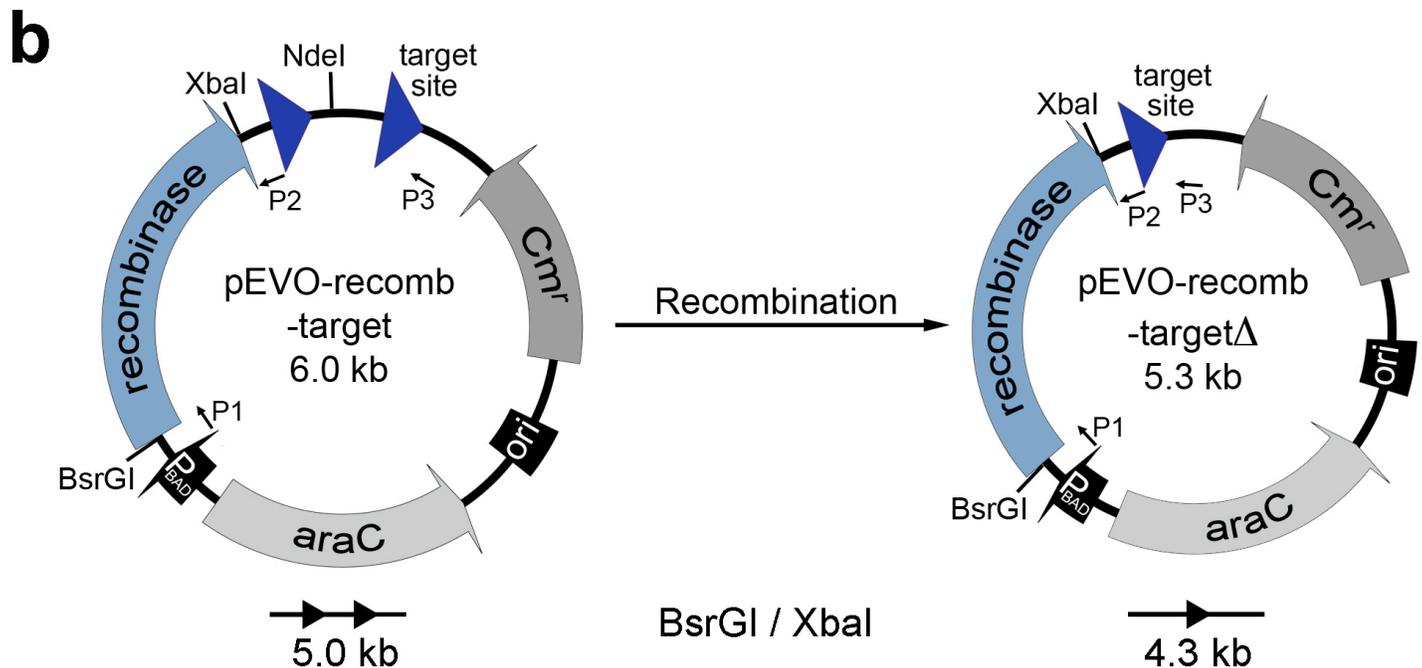
To visualize p24 antigen, the monoclonal antibody clone Kal-1 (Dako M0857) was used in a 1:10 dilution. Deparaffinized sections were boiled for 20 min in retrieval buffer S 1699 (Dako) using a pressure cooker set at 100 °C. Streptavidin alkaline phosphatase and the TNB-Amplification Kit with Permanent Red were used for visualization. Antibodies used have been validated previously²⁹.

Ethics statement. The animal experiments were performed according to the guidelines of the German Animal Protection Law. The experimental protocols were reviewed and approved by the relevant German authority, the local ethics commission (Ärztchamber Hamburg; OB-050/07, WF-010/2011 and PV4666) and the Freie und Hansestadt Hamburg, Behörde für Gesundheit und Verbraucherschutz (Nr.: 63/09, 23/11 and 29/14). Patient material was obtained after written informed consent, which was approved by the local ethics commission (Ärztchamber Hamburg).

55. Anastassiadis, K. *et al.* Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *Dis. Model. Mech.* **2**, 508–515 (2009).
56. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Li, H. *et al.* 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Prüfer, K. *et al.* PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics* **24**, 1530–1531 (2008).
59. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

a

	1	Halfsite	13	Spacer	22	Halfsite	34	
loxBTR	A	ACCCACTGCTT	TAAGCCTCAATA		AAGCTTGCCTT			Brec
subsite 1	A	ACCCACTGCTT	TAAGCCTCAATA		AGCAGTGGGTT			pre-Brec intermediates
subsite 1a	A	ACCATTGTATA	TAAGCCTCAATA		TACAATGGGTT			
subsite 1b	A	AACA	CTGCTT	TAAGCCTCAATA	AGCAGT	GTTAT		
subsite 1A	A	ACACATTGTATA	TAAGCCTCAATA		TACAATGTGTT			
subsite 1B	A	TACCACTGTATA	TAAGCCTCAATA		TACAAGTGGTAT			
subsite 1C	A	TAACATTGCTT	TAAGCCTCAATA		AGCAATGTTAT			
subsite 2	A	AGGCAAGCTT	TAAGCCTCAATA		AAGCTTGCCTT			
subsite 2a	A	AGGCTAGGTATA	TAAGCCTCAATA		TACCTAGCCTT			
subsite 2b	A	TACAAGCTT	TAAGCCTCAATA		AAGCTTGTTAT			
subsite 2A	A	AGACTAGGTATA	TAAGCCTCAATA		TACCTAGTCTT			
subsite 2B	A	TAGCAAGGTATA	TAAGCCTCAATA		TACCTTGCTAT			
subsite 2C	A	TAACTAGCTT	TAAGCCTCAATA		AAGCTAGTTAT			



Supplementary Figure 1

Substrate-linked protein evolution components.

(a) loxBTR subsites used for the combinational directed-evolution process to generate Brec1. Sequence differences (compared to the

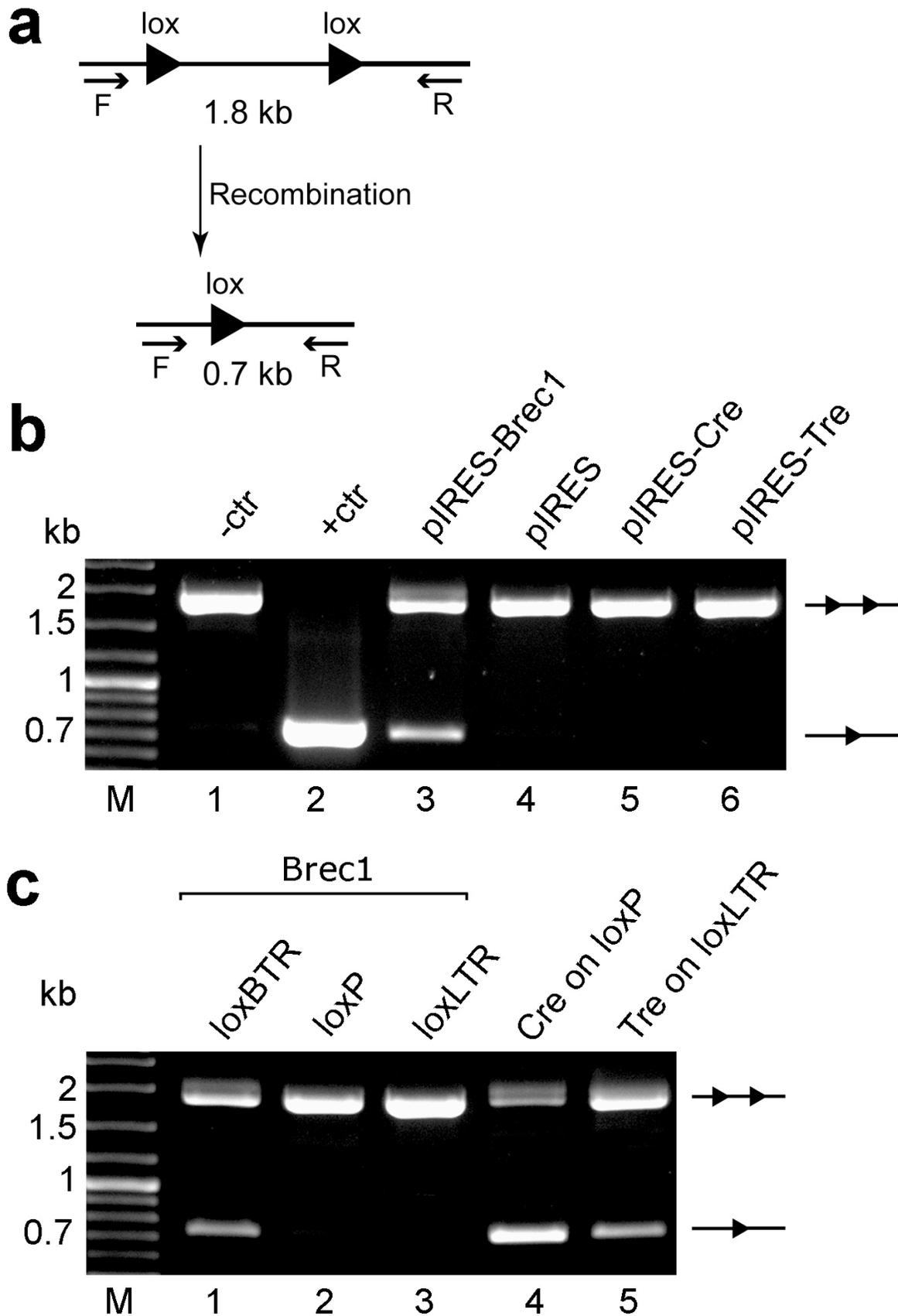
pool of employed related recombinase target sites) in the enzyme bound half sites of the Brec1 target site loxBTR and loxBTR subsites are marked in red or green for the reverse complementary half site. The spacer sequences are shown in grey. **(b)** Evolution vector pEVO-recomb-target. In *E. coli*, the recombinase is expressed from the pBAD promoter upon induction with L-arabinose. The vector also contains the regulatory gene *araC* and a chloramphenicol resistance marker (Cm^r). Recombination at the target sites leads to deletion of the intervening region containing a unique NdeI restriction site; following DNA cleavage with NdeI, successful candidates were retrieved by PCR using primers P1 and P3. Primer P2 was employed for PCR amplification of the full-length recombinase coding sequence after DNA shuffling. BsrGI and XbaI restriction sites flanking the recombinase coding sequence facilitate convenient cloning of the recombinases and simple assaying of recombination efficiency in bacteria. The sizes of the unrecombined and recombined (Δ) forms of the vector are indicated.

M V P^{NLS} K K R K V S I³ L L T L⁷ H Q S¹⁰ L S¹² A L
¹⁵L V D A T S D E A²³ R K N L M D V³⁰ L³¹ R D R Q A
F S E R⁴⁰ T W K V⁴⁴ L L S V C R T⁵¹ W A A W C K L
N N R K W F P A E P E D V R D Y L L H⁷⁷ L Q A
R G L A V N⁸⁶ T I L⁸⁹ Q H L A⁹³ Q L N M L H R R¹⁰² F
G L P R P G¹⁰⁸ D S D¹¹¹ A V S L V M R R I R R¹²² E N
V D A G E R T¹³¹ K Q A L A F E R T D F D Q V R
¹⁴⁷A L M E N S E¹⁵³ R G¹⁵⁵ Q D I R T¹⁶⁰ L A L¹⁶³ L G V¹⁶⁶ A Y
N T L L R V¹⁷⁴ S¹⁷⁵ E I A R I R I¹⁸² K D I S R T D G
G R M L I H I S¹⁹⁸ R T K T L V S T A G V E K A
L S L G V T K L V E R W I S V S G V A S²³² D P
N N Y L F C Q²⁴¹ V R I²⁴⁴ N G V A V²⁴⁹ P S A T S R²⁵⁵ L
S T D²⁵⁹ V²⁶⁰ L R²⁶² K²⁶³ I F E A A²⁶⁸ H R L I Y G A K D G²⁷⁸
S G Q R Y L A W S G H S A R V G A A R D M A
R A G V S I A³⁰⁷ E I M Q A G G W T T³¹⁷ V E³¹⁹ S³²⁰ V M
N Y I R N L D S E T G A M V R L L E D G D *

Supplementary Figure 2

Amino acid sequence of Brec1.

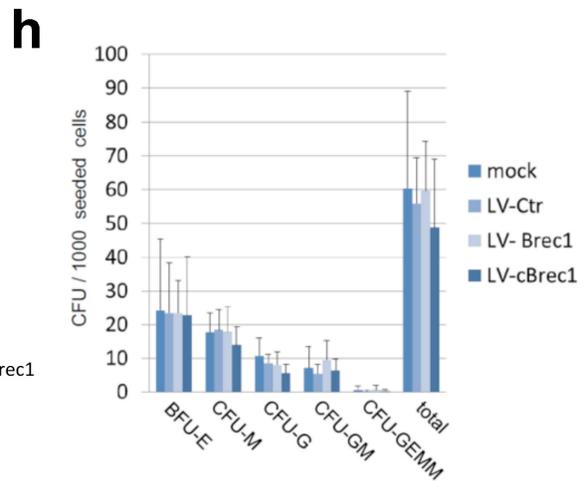
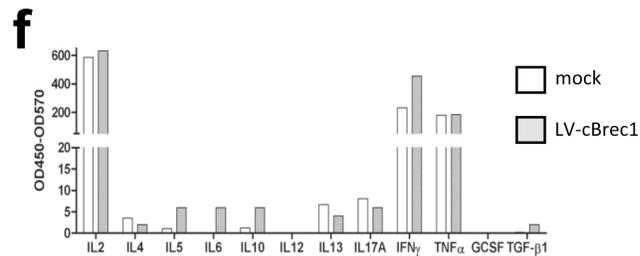
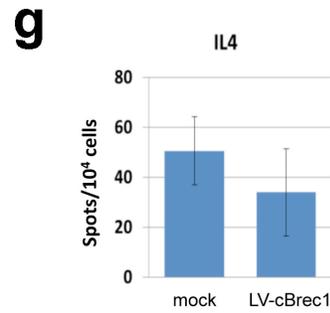
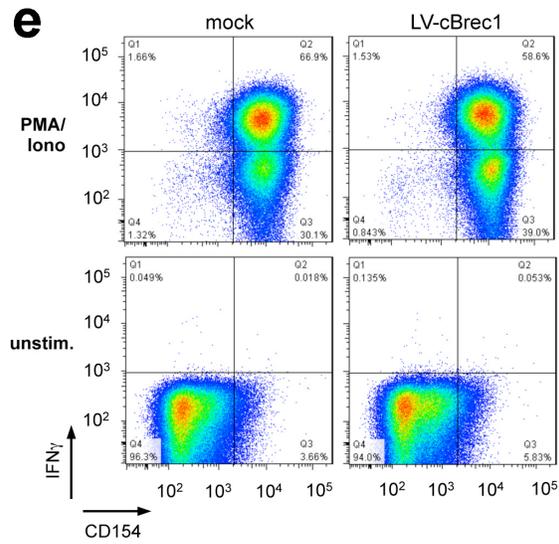
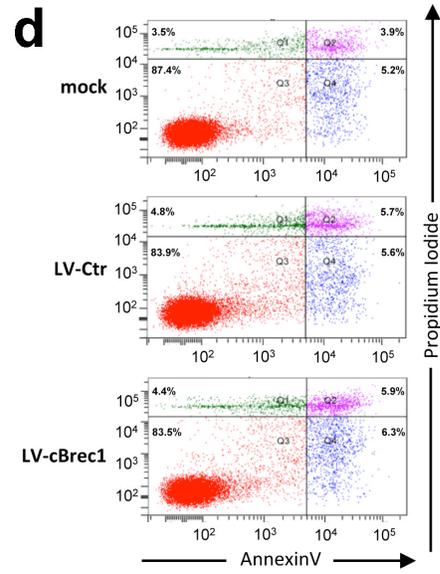
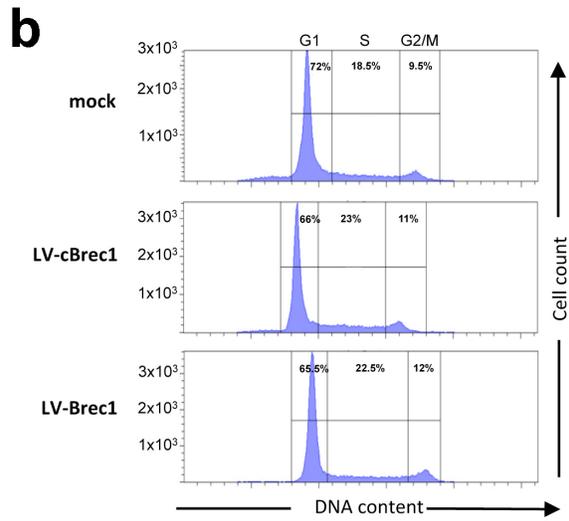
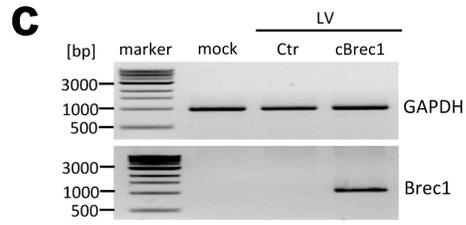
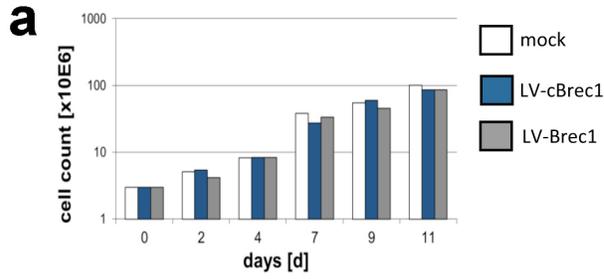
The NLS sequence, added for nuclear localization of Brec1 in mammalian cells, is shown in blue. Mutations and their corresponding amino acid number in the Cre sequence are shown in red, evolutionary conserved mutations are shown in bold. DNA contact residues based on the Cre crystal structure are underlined.



Supplementary Figure 3

Brec1 efficiency and specificity recombines loxBTR in a genomic context.

(a) Schematic representation of recombination of reporter constructs stably integrated in the genome. After recombination, PCR results in a 0.7 kb product (one triangle), while non-recombined template produces a 1.8 kb PCR product (two triangles). (b) Reporter HeLa cell lines with stably integrated loxBTR target sites were transfected with recombinase expression plasmid. Genomic DNA was isolated and assayed for recombination by PCR using primers F and R that anneal to the integrated DNA regions: -ctr, transfection of wild type cells (lane 1); +ctr, transfection with recombined reporter (lane 2); lanes 3-6 show results from cells carrying genomic loxBTR sites transfected with the indicated expression plasmids; M, DNA marker lane. (c) Reporter HeLa cell lines with stably integrated target sites loxBTR, loxP or loxLTR were transfected with recombinase Brec1 expression plasmid (lanes 1-3). Genomic DNA was isolated and assayed for recombination by PCR using primers F and R. Controls for cells harboring the loxP or LoxLTR reporters, transfected with Cre or Tre recombinase expression vectors are shown in lanes 4 and 5.



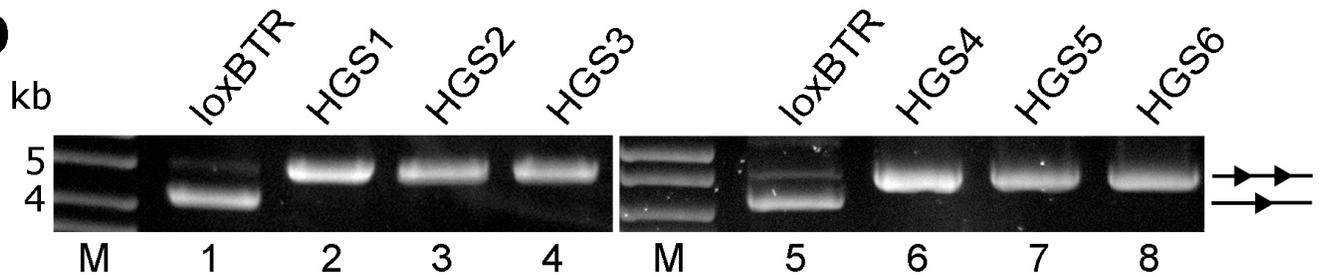
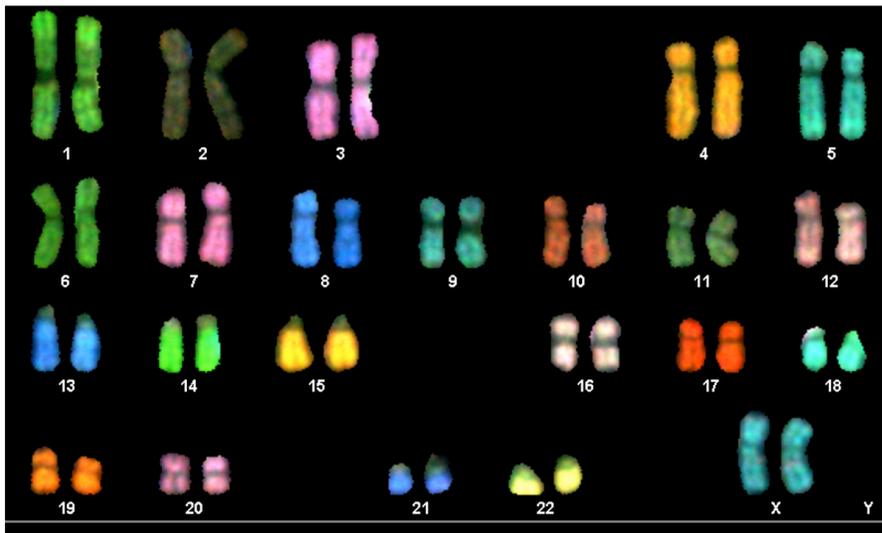
Supplementary Figure 4

Safety evaluations.

Exponentially growing Jurkat T cells were transduced with LV-cBrec1, encoding constitutively expressed Brec1 recombinase, with LV-Brec1 (Tat-inducible promoter configuration), or mock transduced, respectively. Transduced GFP⁺ cells were enriched by FACS. **(a)** Transduced cells (3×10^6) were seeded into dishes and were grown for 11 days. At indicated time point cells were counted and total cell numbers were calculated. **(b)** Cell cycle progression monitored by DNA staining at week 6 of Brec1 expression. Jurkat cells were harvested, fixed, and DNA content was stained with propidium iodide. Signals were measured by analytic flow cytometry. **(c)** Primary human CD4⁺ T cells were transduced with LV-cBrec1 (constitutive promoter configuration), with LV-Ctr (encoding GFP only), or were mock transduced. Transduced GFP⁺ cells were enriched by FACS and cultured for two weeks. Subsequently, expression of Brec1 or GAPDH (control) was analyzed by RT-PCR. **(d)** Analysis of apoptosis in primary human CD4⁺ T cells at week two post transduction by Annexin V/propidium iodide staining. **(e)** Functionality of LV-cBrec1 transduced primary CD4⁺ T cells at two weeks post transduction tested by flow cytometric analysis of CD154 and IFN γ expression after 3 hours of PMA/ionomycin stimulation. **(f)** Secretion pattern of Th1-, Th2- and Th17-specific cytokines in primary transduced CD4⁺ T cells after PMA/ionomycin stimulation two weeks after transduction as determined by multiplex ELISA. The pattern matches that of non-transduced cells (mock). **(g)** IL4-specific Elispot analysis of human primary mock or LV-cBrec1 transduced CD4⁺ T cells after PMA/ionomycin stimulation two weeks after transduction. **(h)** CD34⁺ HSC were mock-transduced or transduced with LV-Ctr, LV-Brec1, or LV-cBrec and 100 cells were seeded into cytokine-containing methyl-cellulose. Culture dishes were incubated for 14 days before counting colonies.

a

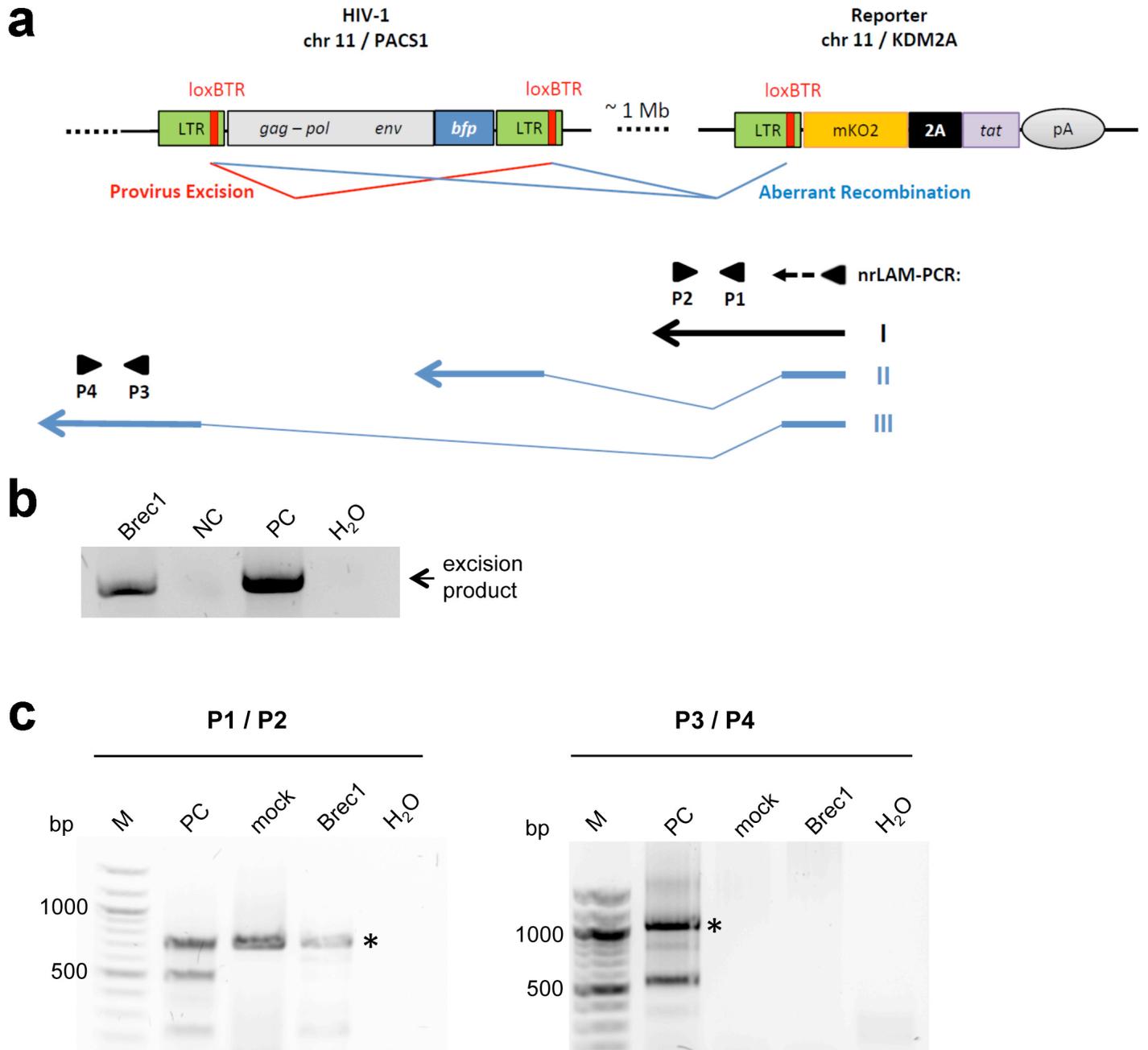
	Halfsite	Spacer	Halfsite
loxBTR	AACCCACTGCTT	AAGCCTCAATAA	AGCTTGCCTT
Human genomic sites:			
HGS1	AA G CCCTT G CTT	TAAAAGGATT	TAAAG A AT G TTTA
	(chr1, 2 sites)		
HGS2	AA A TTA ^C T	TGCTTATGAAGAA	TAAAGC C AGC A TT
	(chr4/chr8, 2 sites)		
HGS3	A T CC ^C G A T AGCTT	ATTTAATAATAA	AG T TT G T A T A
	(chr17, 2 sites)		
HGS4	A T CCCACTGCT G AATATCCTCT	TAAAGCTT	T C T G T
	(chr6, 2 sites)		
HGS5	G AC G CA T T C CTT	ATTCTTGAA	AAAAGCTT G C A T A
	(chr2/chrX, 2 sites)		
HGS6	C ACA A T C T T CTT	ACTACTGTAGT	TAAAGCTT G C T T G
	(chr9, 4 sites)		

b**c**

Supplementary Figure 5

Brec1 specificity assays.

(a) Nucleotide sequences of human genomic sites that most closely resemble the loxBTR sequence and their locations in the human genome. Sequences are aligned to loxBTR. Nucleotides that differ from loxBTR are shown in red. (b) Agarose gel showing the activity of Brec1 on loxBTR (lanes 1 and 5, positive control) and the lack of activity for the six loxBTR-like human genomic sites HGS1 – HGS6 (lanes 2-4 and 6-8). *E. coli* cells harboring the pEVO vector containing the Brec1 coding sequence and indicated target sites were grown at 1 mg/ml L-arabinose. Recombination was assayed by restriction enzyme digest, resulting in a smaller fragment for recombined (one triangle) and a larger fragment for non-recombined substrate (two triangles). M, DNA marker. (c) SKY (spectral karyotyping) analysis of metaphase spreads isolated from primary human CD4⁺ T cells constitutively expressing Brec1 (LV-cBrec1). An RGB display of the 24-colour SKY hybridization of a representative normal metaphase is shown. Number of analyzed chromosome spreads: 27.

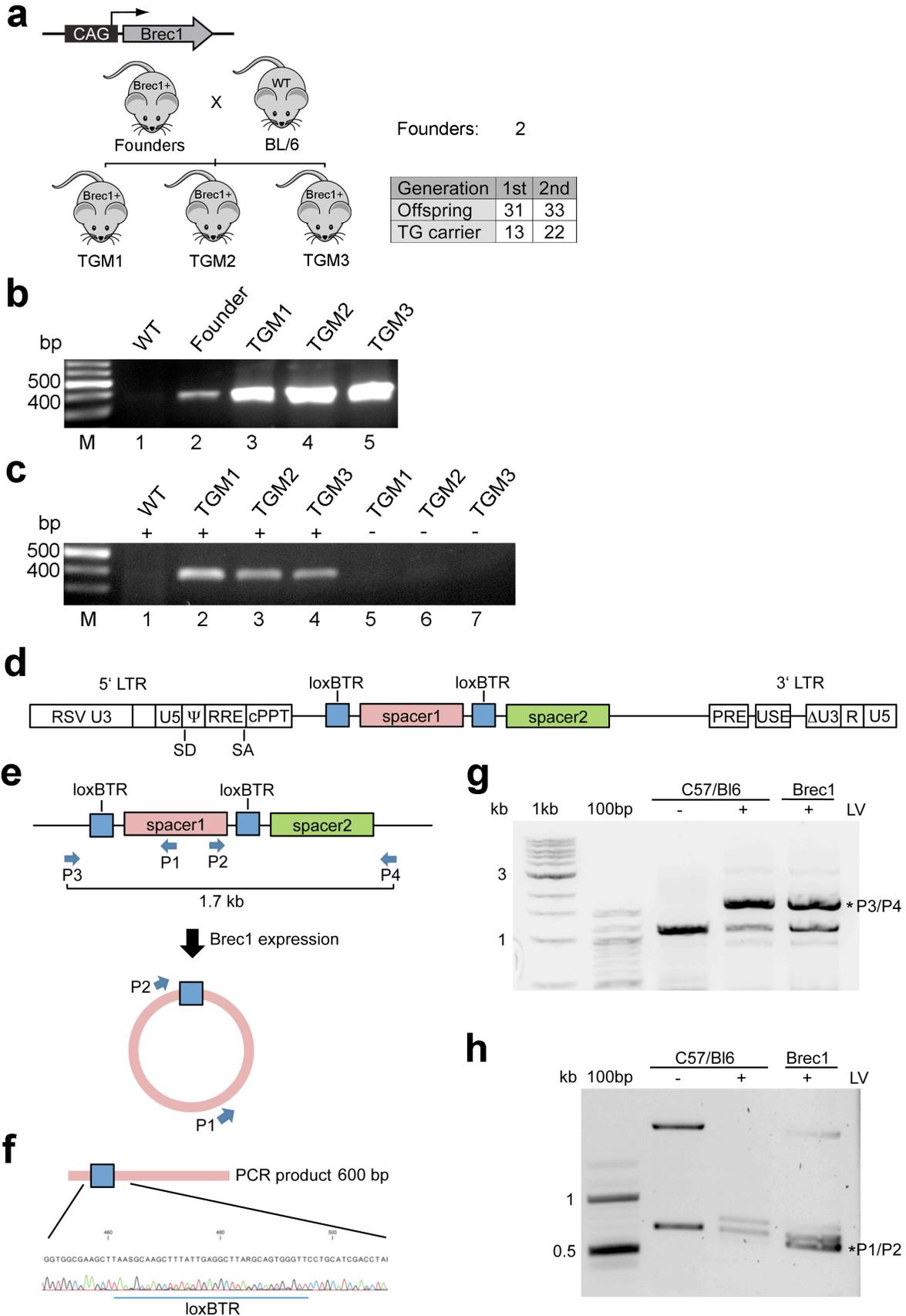


Supplementary Figure 6

Targeting of multiple chromosomal loxBTR sequences by Brec1.

(a) HeLa-smurf cells, containing at the PACS1 locus on chromosome 11 a full-length HIV-1 proviral DNA that expresses blue fluorescent protein (BFP), have been previously described²⁹. At a distance of ~ 1 Mb, an additional self-contained reporter construct was inserted at the KDMA2 locus via CRISPR/Cas9 technology. This reporter simultaneously expresses from an HIV-1 LTR promoter the fluorescent protein mKO2 and the HIV *trans*-activator Tat by use of an ERAV 2A-like sequence. (b) Brec1-mediated provirus recombination results in an excision product that can be detected by PCR (depicted in Fig. 4c). At day 3 post transduction with LV-Brec1, this circular excision product was observed in chromosomal DNA isolated from HeLa-smurf reporter cells. Brec1, LV-Brec1 transduced cells; NC, negative control (non-transduced cells); PC, positive PCR control; H₂O, negative PCR control (omitting DNA template). (c) Analysis of potential aberrant Brec1-mediated loxBTR recombination. To monitor LTR/genomic DNA junctions, and LTR/proviral genome junctions, non-restrictive linear amplification-mediated PCR (nrLAM-PCR) was performed as described

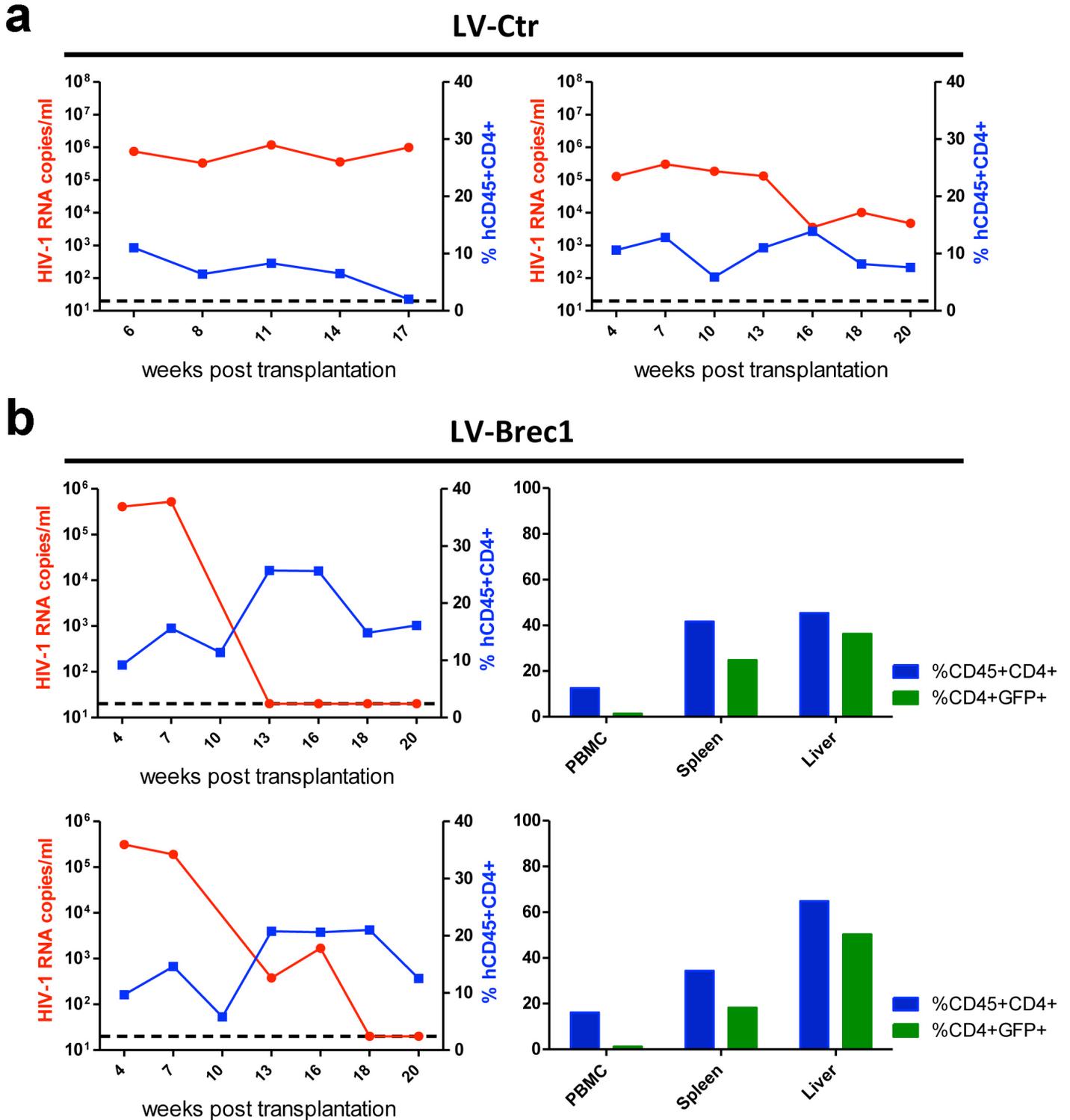
previously²⁹, using an mKO2-specific anchor primer (indicated in panel a). In addition to extension products from the reporter construct integrated at the KDM2A locus (product I in panel a), aberrant recombination between distant loxBTR sequences would be expected to yield nrLAM extension products which contain either proviral or 5'-proximal host sequences from the PACS1 locus (products II and III in panel a, respectively). As shown in panel c, PCR analysis with primer pairs designed to amplify upstream integration site sequences from extension products I and III (indicated in panel a) revealed the expected amplification product (asterisk) for primer pair P1/P2, but not for primer pair P3/4, indicating that aberrant recombination (i.e. recombination of the proviral 5'LTR and the LTR of the reporter construct) did not occur. M, DNA marker; PC positive PCR control; mock, nrLAM-PCR template from non-transduced cells; Brec1, nrLAM-PCR template from LV-Brec1 transduced cells; H₂O, negative PCR control (omitting DNA template). Additionally, next generation bulk sequencing of nrLAM amplicons (2x250 bp paired end, Illumina MiSeq platform) yielded 1698 reads with 5'-fragments that spanned the host/loxBTR junction of the reporter construct but none that would indicate aberrant recombination, again indicating that Brec1 expression resulted exclusively in provirus excision.



Supplementary Figure 7

Generation and analysis of transgenic Brec1 mice.

(a) Schematic representation of the Brec1 expression cassette. Brec1 is expressed from the cytomegalovirus (CMV) immediate early enhancer-chicken beta-actin hybrid (CAG) promoter. Breeding strategy following generation of Brec1 founder mice by pronuclear injection: WT, wild type mouse; TGM, transgenic mouse. The number of offspring and the TG carriers indicated for the first and second generation show that the transgene is inherited with a Mendelian ratio. **(b)** Agarose gel showing PCR-based genotyping of one founder (lane 2) and three transgenic offspring (lanes 3-5) using Brec1-specific primers. WT, wild type control (lane 1); M, DNA marker lane. **(c)** Agarose gel showing PCR-based analysis of cDNA generated by reverse transcription of mRNA isolated from WT animal (control, lane 1) and three transgenic Brec1 mice (TGM1-TGM3, lanes 2-7): +, reverse transcriptase added to RT reaction; -, no-reverse transcriptase control. **(d)** Schematic representation of the lentiviral reporter construct. Important elements are indicated. The vector backbone has been previously described in detail²⁹ **(e)** Schematic drawing of the primer binding sites within the lentiviral reporter construct. Important elements are indicated and the size of the expected amplicon for primers P3 and P4 is shown. Brec1-mediated recombination leads to the formation of a circular excision product that can be detected employing primer P1 and P2. **(f)** Schematic of the 600 bp amplification product for primers P1 and P2. A sequencing read is shown confirming the expected sequence. The loxBTR sequence is marked. **(g)** Confirmation of lentiviral integration of the reporter construct. PCR assays for indicated animals is provided. The 1.7 kb fragment resulting from amplification with primers P3 and P4 is marked by an asterisk. **(h)** Confirmation of Brec-1 mediated recombination. Gel showing PCR analysis using DNA isolated from cells of indicated animals. The product showing the recombined band is indicated with P1/P2.



protection in hu-Rag2 mice engrafted with LV-Brec1 transduced patient-derived CD4⁺ T lymphocytes. Human cells in PBMC, spleen and liver of the respective Brec1-expressing animal was determined at necropsy by FACS using single cell suspensions (bar chart).

Supplementary Table 3 Human sequences with the highest similarity to loxBTR

Sequence (mismatches in red, spacer in gray)	Genomic coordinates (GRCh38/hg38)	
	Upper sequence	Lower sequence
ATCCCACTGCTGAATATCCTCTAAAGCTTCTGT	chr6:57983492-57983526(-) chr6:60734963-60734997(-)	
AAATT ^C _T GTCTTATGAAGAAATAAAGCCAGCATT	chr8:65603300-65603334(-)	chr4:138478068-138478102(-)
GACGCATTCCCTTATTCTTGAAA ^A AAAGCTTG ^A CATA	chr2:87894679-87894713(-) chrX:144143848-144143882(+)	
CACAATCTTCTTACTACTGTAGTAAAGCTTGCTTG	chr9:39631897-39631931(-) chr9:42385618-42385652(-) chr9:62710515-62710549(+) chr9:66745043-66745077(+)	
ATCC ^G _C ATAGCTTATTTAATAATAAAGTTGTATA	chr17:20833853-20833887(-)	chr17:22537726-22537760(-)
AAGCCCTTGCTTAAAAGGATTTAAAGAAATGTTTA	chr1:121169348-121169382(-) chr1:143956994-143957028(-) chr1:145110955-145110989(+) chr1:206187886-206187920(-)	
GATCCACAGTTAATCAAAGATTAACTTGCCTT	chr5:176120177-176120211(+) chr5:177717935-177717969(-) chr5:178053166-178053200(+)	
CCCCAGTCTTTAT ^G _A TATTATAAAGCTGACTT	chr6:58335745-58335779(+)	chr6:61114838-61114872(-)
AACCCAA ^A _T GTTCAACTTTTATTGCATCTTGCCAT	chr12:131321961-131321995(-) chr17:51314110-51314144(+)	chr2:123470747-123470781(+)
AACCCAGTGCTGCTGTTTGTAGTAAAG ^G _T TGGGGT	chr6:168612586-168612620(-)	chr6:168612348-168612382(-)
AAGGA ^C _T ACTGTTATACTGATTCTAAGTATGCCTT	chr3:110968380-110968414(-)	chr1:173639110-173639144(+)
ATC ^C _A CAGTGCTTGTGTTCAAGTAA ^C _C CTTGTTTT	chr7:137745627-137745661(+)	chr3:17299119-17299153(-)
AACACT ^C _T GTCTATATTGTTATAAAGTTGCTTT	chr6:144097474-144097508(+)	chr18:2211029-2211063(+)
AAC ^C _T ACTGATCACAGCCAGTTACAGATTTCTTT	chr3:23273835-23273869(+)	chr2:121040649-121040683(-)
AACCCTCTTTTATAGTGTCTGGAAGCGGC ^A _C TT	chr8:43967414-43967448(+) chr8:44033822-44033856(+) chr8:44036209-44036243(+) chr8:44038758-44038792(+) chr8:44039951-44039985(+) ...and 1096 more.	chr8:45824413-45824447(+)
ATCCCACTGAGTTGCAGACTTTAAAGAGTGA ^C _A TT	chrY:17254195-17254229(-)	chrX:4398866-4398900(-)
AATCTGCT ^A _G ATTAGCAAAAAGTAAAGATTTCTT	chr15:20461155-20461189(+)	chr15:22539249-22539283(-)
AACCCTCTT ^T _C TTATAGTGTCTGGAAGCGGCATT	chr8:43967414-43967448(+) chr8:44033822-44033856(+) chr8:44036209-44036243(+) chr8:44038758-44038792(+) chr8:44039951-44039985(+) ...and 1096 more.	chr8:45090207-45090241(+)
ATCCCTCTCCTTACGCCAGGCTAAGCTGGC ^A _T	chr11:67711993-67712027(-)	chr3:75619785-75619819(-)
AATCCACTTTTATGGCCATCTAAATCCTTTATT	chr9:39755058-39755092(+) chr9:62587319-62587353(-) chr9:66621841-66621875(-)	

Sequence (mismatches in red, spacer in gray)	Genomic coordinates (GRCh38/hg38)	
	Upper sequence	Lower sequence
AACACTTGGTTTACTGTATCATAATGCTAGCTTT	chr5:69699842-69699876(+) chr5:69855349-69855383(-) chr5:70279580-70279614(+) chr5:70575154-70575188(+) chr5:70730497-70730531(-) ...and 1 more.	
AACACTTGGTTTACTGTATTATAATGCTAGCTTT	chr5:21352522-21352556(-) chr5:34234152-34234186(+) chr6:26945871-26945905(+)	
AAAGCACTATTAAAGATTTTGGTAAAAGTTGCATT	chr10:131369961-131369995(+) chr10:131371212-131371246(+)	
ACAGCACTGCTTTACAGAATGTAAACCGTGCCT	chr1:199194-199228(+) chr12:30732-30766(+) chr2:113584744-113584778(-)	
ACAGCACTGCTTTGAGAATGTAAACCGTGCCT	chr1:28678-28712(+) chr19:70286-70320(+) chr9:28792-28826(+)	
CACTCAATTCTTACCTTCTGTAATGTTGACTC	chr13:19127896-19127930(-) chrX:101128816-101128850(-)	
AACAATTGCCTAATCCAAGCTAAGTATTGTCTT	chr7:150088243-150088277(+) chr7:154012381-154012415(-)	
AGGACACTGCTTCGAGTGTCTGAATGTTGTCTT	chr14:18497748-18497782(+) chr22:15424704-15424738(+)	
GATGCAATGCTTATGTGG ^G _T CAGAAGGCTTGATT	chr1:206358497-206358531(-)	chr1:121340165-121340199(-)
AACACTTGGTTTACTGTAT ^C _T CATAATGCTAGCTTT	chr5:69699842-69699876(+) chr5:69855349-69855383(-) chr5:70279580-70279614(+) chr5:70575154-70575188(+) chr5:70730497-70730531(-) ...and 1 more.	chr5:21352522-21352556(-) chr5:34234152-34234186(+) chr6:26945871-26945905(+)
TAC ^A _G CACTCCTTATGAGAATCTAACTAATGCCTG	chr11:77605581-77605615(+) chr6:39709341-39709375(-)	chr10:115337911-115337945(-)
ATCACA ^A _G TGCTTGTTCAAGTAAACCTTGTTT	chr3:168850196-168850230(-)	chr3:17299119-17299153(-)
AACCCACTGGTCTAGTTCCTAAAACACTCACTT	chrY:18598352-18598386(+) chrY:18722015-18722049(-)	
AAGACAATACTTCTCTGCACAATGCTTACCTT	chr17:21632234-21632268(+) chr17:22070996-22071030(-) chr20:26114437-26114471(-) chr21:9548399-9548433(+)	
AAGCCCTTACTTAAATCATCCCAATGCCTGGCTT	chr17:36313123-36313157(+) chr17:38117226-38117260(+) chr17:38203260-38203294(-)	
ATCACAGTGCTTATGTTCAAGTAAACCTTATTTT	chr1:44107701-44107735(+) chr5:153998283-153998317(+) chr9:115529723-115529757(-)	
AGCACAGTGCTGGGTCTTGCCCAAGGCCTGCCTT	chr13:35037799-35037833(+) chr14:19273501-19273535(-) chr22:15848833-15848867(+)	
ATTCCAGGGCTTTATTTTCCCTAAAGCCTGGCTC	chr5:175988474-175988508(-) chr5:177855291-177855325(+)	
AAGCCA ^G _A GACCTAAATGGCCTATCAGTTGCCTT	chr13:52334488-52334522(-) chr13:52488747-52488781(+)	
TACACACTCCTTATGAGAATCTAACTAATGCCTG	chr11:77605581-77605615(+) chr6:39709341-39709375(-)	
AACTGAGTGCTTTTCAAATATTAACCTTCACTA	chrX:90482508-90482542(+) chrY:4060233-4060267(+)	

Sequence (mismatches in red, spacer in gray)	Genomic coordinates (GRCh38/hg38)	
	Upper sequence	Lower sequence
AAGACATTGCTACCTTATCTTCAACCCTTGCCTT	chr13:19457905-19457939(+) chr13:24401473-24401507(-)	
AATCCATTGCAATAATTCCTTCTAAACACTGTCAT	chrY:6377996-6378030(-) chrY:9786690-9786724(+)	
AACCCAAGTTCAACTTTTATTGCATCTTGCCAT	chr12:131321961-131321995(-) chr17:51314110-51314144(+)	
CACACACGGCTAAAACAGTCCTAAATCTTCCGTG	chr4:68996924-68996958(-) chr4:69126427-69126461(+)	
ATAGCTCTACTTATAAGATTCTAAAGATTCCTTT	chr9:33568338-33568372(-) chr9:38575996-38576030(+)	
ATCCAAATATTTAAACACTAAAAGCTTACCAT	chrX:135130670-135130704(-) chrX:135241831-135241865(+)	
AGCCCACTGTTTTGCTTTATAAAATCCTGATTT	chr3:123965759-123965793(+) chr3:125986360-125986394(-)	
AACTCACCCATTTTGTGTAGAGAAAGTTGCTTT	chr5:21552481-21552515(+) chr6:57729548-57729582(-)	
AACCCCTTTTATAGTGTCTGGAAAGCGGCATT	chr8:43971154-43971188(+) chr8:43973022-43973056(+) chr8:43974890-43974924(+) chr8:43976757-43976791(+) chr8:43978625-43978659(+) ...and 37 more.	
AAGCAACTATTATAATTCTGAAATTTTCCTT	chrX:91777845-91777879(+) chrY:4998770-4998804(+)	
AACCCCTTTTATACTGTCTGGAAAGCGGCATT	chr8:44204762-44204796(+) chr8:45280745-45280779(+) chr8:45501367-45501401(+) chr8:45528384-45528418(+) chr8:45690584-45690618(+) ...and 1 more.	
AACACAAGCTTAC _T TTTATAATAAAGTTTGAAT	chr5:61097379-61097413(+)	chr9:26817426-26817460(+)
TGCCCACTCCTTAT _G AGAATCTAATGCTTGATGA	chr1:188434761-188434795(-)	chr14:72778379-72778413(+)
ATCACAGTGCTTATGTTCAA _G TAAACCCTTATTTT	chr2:174995860-174995894(+)	chr1:44107701-44107735(+) chr5:153998283-153998317(+) chr9:115529723-115529757(-)
ATCACAGTGCTTATGTTCAAATAA _C CCTTATTTT	chr2:174995860-174995894(+)	chr13:107114119-107114153(+)
TACACACTCCTTATGAGAA _C TAACTAATGCCTG	chr11:115964959-115964993(-)	chr11:77605581-77605615(+) chr6:39709341-39709375(-)
AAGACAATACTTCTCTG _C ACAAATGCTTACCTT	chr17:21632234-21632268(+) chr17:22070996-22071030(-) chr20:26114437-26114471(-) chr21:9548399-9548433(+)	chr7:57599304-57599338(+)
ATCACAGTGCTTGT _G TTC AAGTAAACCCTTGTTTT	chr9:71491707-71491741(+)	chr3:17299119-17299153(-)
ATCACAGTGCTTATGTT _A AAGTAAACCCTTATTTT	chr1:44107701-44107735(+) chr5:153998283-153998317(+) chr9:115529723-115529757(-)	chr5:163894650-163894684(-)
AATGCATTGC _G TCTATCATGTTAAACGTGCATT	chr4:173615355-173615389(+)	chr15:78897618-78897652(-)
AACCACTGTATAGCT _G TACAAAAATATTTCTT	chr4:102773999-102774033(-)	chr10:106145859-106145893(-)
AACCCAAGTTCAACTTTTAT _C TGCATCTTGCCAT	chr12:131321961-131321995(-) chr17:51314110-51314144(+)	chr20:25639037-25639071(-)

Sequence (mismatches in red, spacer in gray)	Genomic coordinates (GRCh38/hg38)	
	Upper sequence	Lower sequence
AACCCCTTTTATA ^G _C TGTCTGGAAAGCGGCATT	chr8:43971154-43971188(+) chr8:43973022-43973056(+) chr8:43974890-43974924(+) chr8:43976757-43976791(+) chr8:43978625-43978659(+) ...and 37 more.	chr8:44204762-44204796(+) chr8:45280745-45280779(+) chr8:45501367-45501401(+) chr8:45528384-45528418(+) chr8:45690584-45690618(+) ...and 1 more.
ATCACAGTGCTTATGTTCAAGTAAC ^T _C CTTATTTT	chr4:78280240-78280274(-)	chr1:44107701-44107735(+) chr5:153998283-153998317(+) chr9:115529723-115529757(-)
AACCC ^T _C CTTTTATACTGTCTGGAAAGCGGCATT	chr8:44550961-44550995(+)	chr8:44204762-44204796(+) chr8:45280745-45280779(+) chr8:45501367-45501401(+) chr8:45528384-45528418(+) chr8:45690584-45690618(+) ...and 1 more.
AACCC ^T _C CTTTTATAGTGTCTGGAAAGCGGCATT	chr8:43967414-43967448(+) chr8:44033822-44033856(+) chr8:44036209-44036243(+) chr8:44038758-44038792(+) chr8:44039951-44039985(+) ...and 1096 more.	chr8:43971154-43971188(+) chr8:43973022-43973056(+) chr8:43974890-43974924(+) chr8:43976757-43976791(+) chr8:43978625-43978659(+) ...and 37 more.
AAACA ^G _A TGCTTCATATCAGTGAAGCTCTTCTT	chr6:122643984-122644018(+)	chr7:48081359-48081393(-)
AAAGCACTACTGAAACCAAGATAAACT ^G _C TGGCTT	chr20:30401213-30401247(+)	chr20:28578724-28578758(-) chr20:29478161-29478195(-) chr22:10937412-10937446(-) chr22:12628589-12628623(+)
AAACACTGCTATACCTATTA ^A _G AAACCCTGCATT	chr1:202426328-202426362(-)	chrY:26338197-26338231(+)
CACACACGGCTAAACAGTCCTAAATCTTC ^G _A TG	chr4:68996924-68996958(-) chr4:69126427-69126461(+)	chr4:68883874-68883908(-)
AAAGCACTACTGAAACCAAGATAAA ^T _C TCTGGCTT	chr4:189971259-189971293(+)	chr20:28578724-28578758(-) chr20:29478161-29478195(-) chr22:10937412-10937446(-) chr22:12628589-12628623(+)