**YMTHE, Volume *31***

## Supplemental Information

# Precise excision of HTLV-1 provirus with a designer-recombinase

**Teresa Rojo-Romanos, Janet Karpinski, Sebastian Millen, Niklas Beschorner, Florian Simon, Maciej Paszkowski-Rogacz, Felix Lansing, Paul Martin Schneider, Jan Sonntag, Joachim Hauber, Andrea K. Thoma-Kress, and Frank Buchholz**

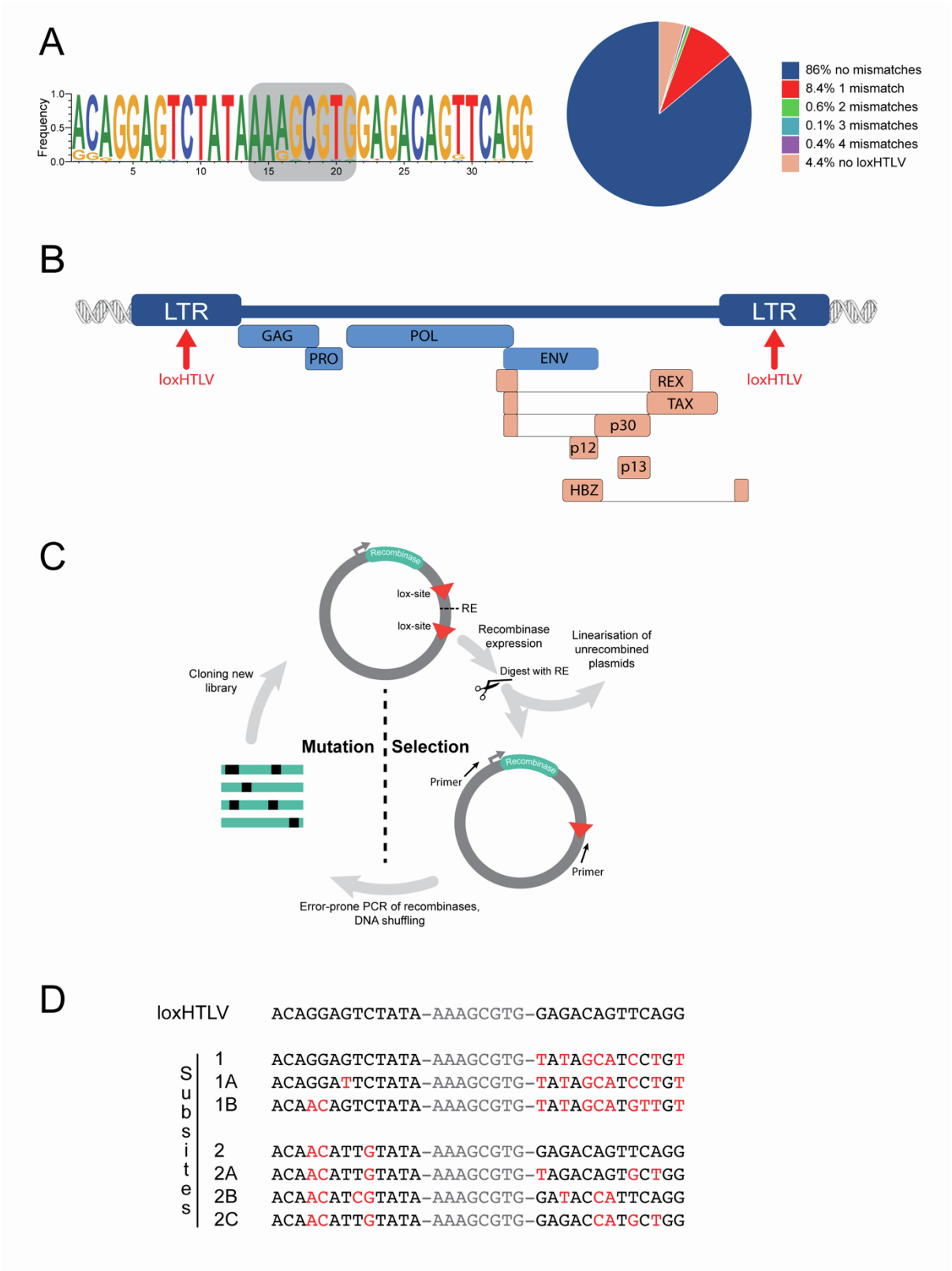



**Figure S1. loxHTLV is conserved within HTLV-1 isolates and located in the LTRs of the HTLV-1 virus**
A) loxHTLV is conserved within sequenced HTLV-1 isolates. Left: Graph showing frequency of conservation of the loxHTLV sequence in the sequenced isolates by nucleotide position; Right: Conservation of the loxHTLV sequence in the sequenced isolates with no mismatches and 1-4 mismatches shown as percentage of the total. B) Schematic representation of HTLV-1 proviral genome and the location of RecHTLV indicated with red arrows (Adapted from Boxus and Willems, 2009).[33] C) Overview of substrate-linked directed evolution (SLiDE). Upon expression of the recombinase library in the pEVO vector, the plasmid is extracted by miniprep and digested with restriction enzymes (RE) whose sites are located between the two lox-sites. The non-recombined plasmids will be linearized and therefore will not be amplified in the subsequent PCR (primers indicated as arrows). By the

mutations introduced by PCR or shuffling of the libraries, variability is introduced to the recombinases and the resultant library is cloned back in the pEVO vector to start another cycle of evolution. D) Alignment of the loxHTLV (RecHTLV) target sequence to the subsites of evolution (1, 1A, 1B, 2, 2A, 2B, 2C). Mismatches compared to loxHTLV are marked in red and spacers are shown in grey.

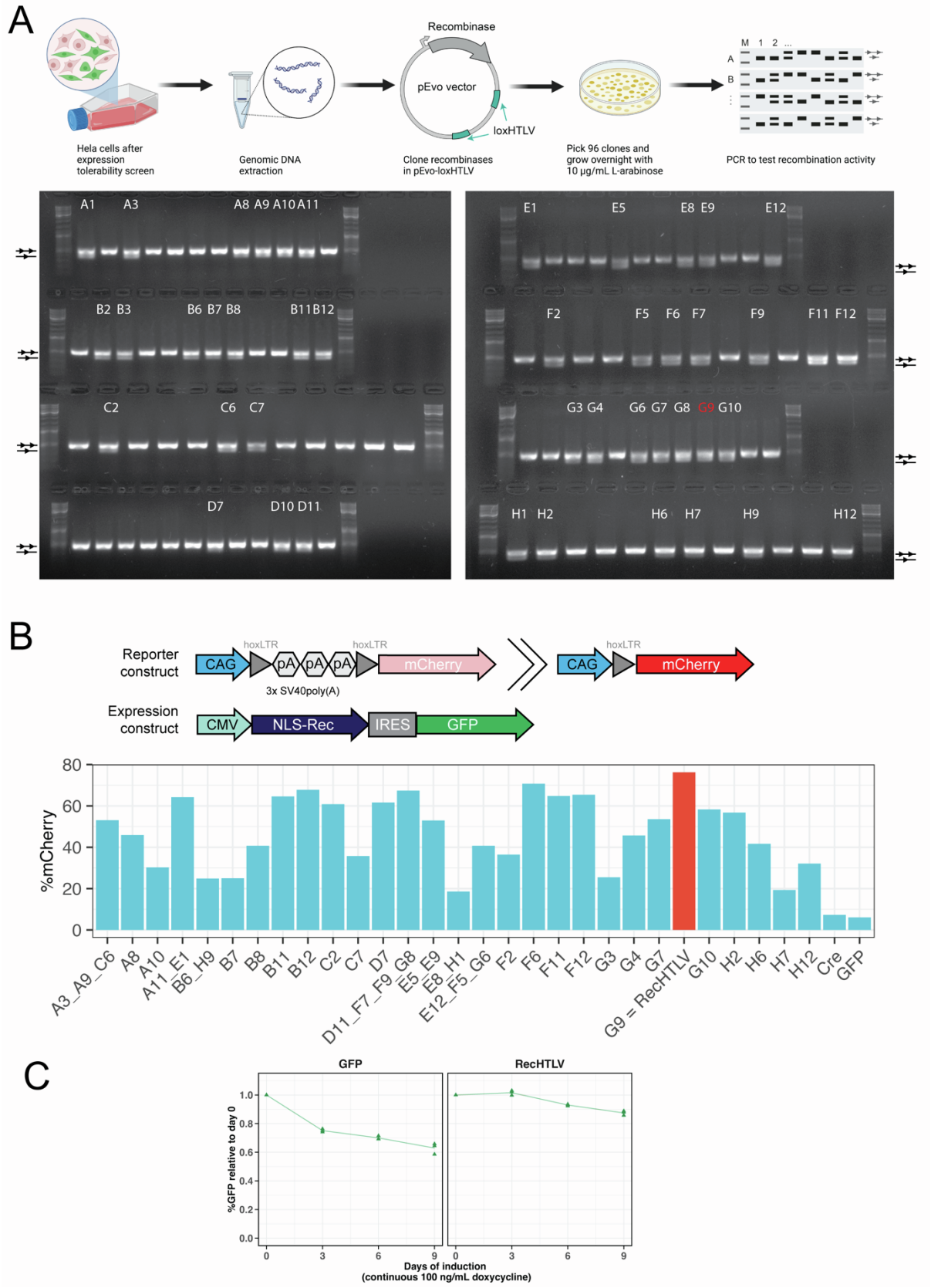



**Figure S2. Screening clones from the library from the library for activity and specificity**

A) Top: Schematic representation of the workflow for testing and selection of clones. After expressing the recombinase library in HeLa cells, genomic DNA was extracted and the library of recombinases was amplified by PCR and cloned in the pEVO vector. After transformation in bacteria, 96 colonies were picked and grown at 10 µg/mL of L-arabinose and activity was assessed with a PCR based test. Bottom: Agarose gel showing the PCR for the analysis of activity of the 96 clones. The upper band shows the PCR product from unrecombined plasmid

(illustrated with a line with two triangles) and the lower band shows the PCR product from recombined plasmid (illustrated with one triangle). The clones which show activity are marked with the given name of the clone and in red the G9 clone, RecHTLV, is indicated (n=1). B) Top: Schematic representation of the reporter and expression constructs used for testing the HTLV-1 clones for activity in HeLa cells. Bottom: Quantification of the recombination activity based on flow cytometry analysis and quantitation of the frequency of mCherry positive cells among the total number of transfected cells (GFP+). In red, the RecHTLV clone G9, which has the highest recombination activity is indicated. As negative controls for recombination, Cre and GFP vectors were used. C) Tolerability assay in HeLa cells. Quantification of the percentage of RecHTLV expressing cells in the course of 9 days indicated by the percentage of GFP positive cells relative to day 0. The same vector, only carrying GFP, was used as a negative control. Parts of the figure were created with BioRender.com.

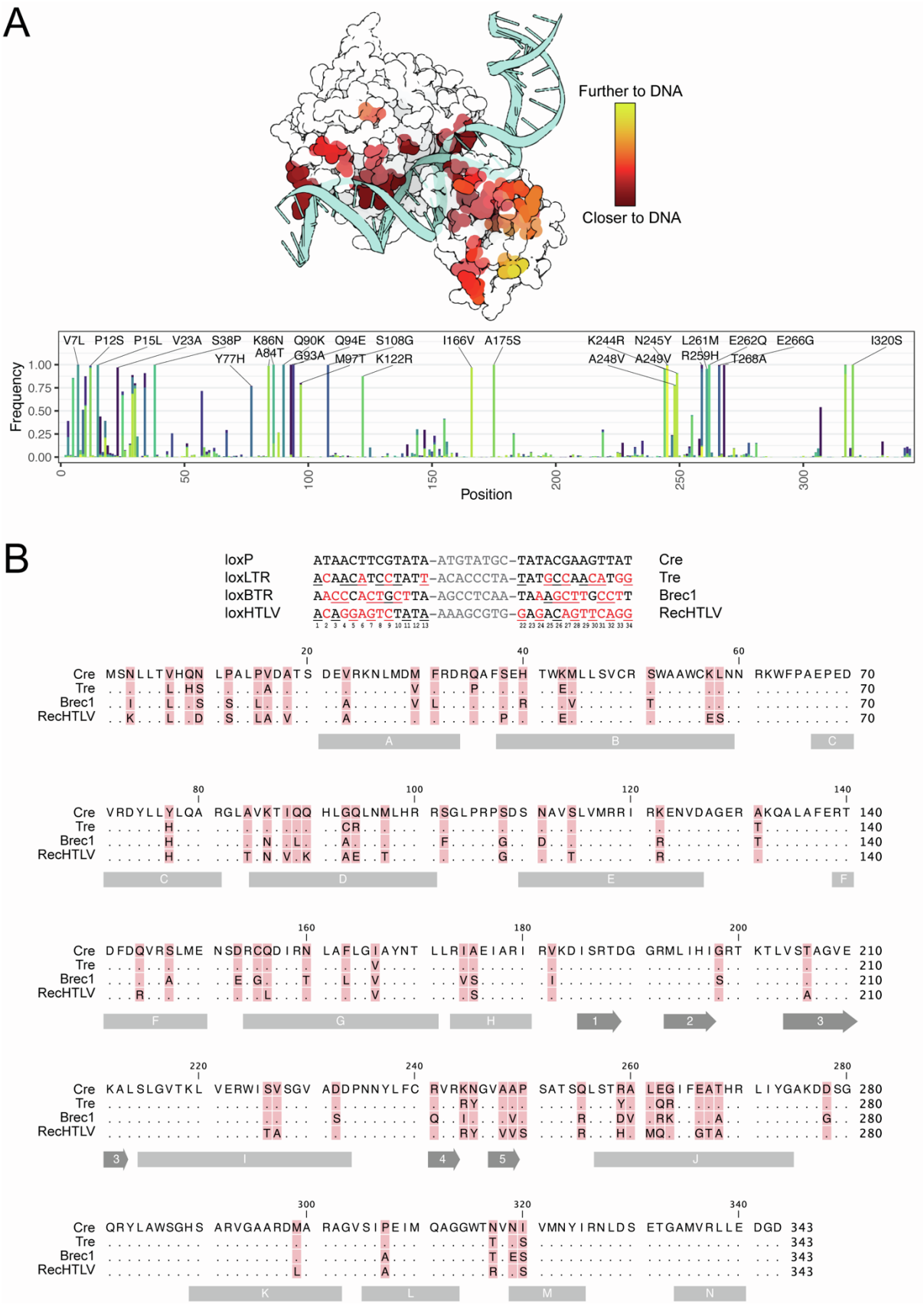



**Figure S3. Mutations acquired in the library and RecHTLV during evolution**

A) Deep sequencing of the RecHTLV library reveals mutations acquired during evolution. Top: Representation of the 3D structure of Cre bound to loxP (1Q3U) using the Protein imager software (https://3dproteinimaging.com/protein-imager/).[1,2] Coloured in a gradient from yellow to dark red are the residues that are mutated in the RecHTLV library when compared to the sequence of Cre. Bottom: Graph showing the

frequency of mutated sequences in the library compared to Cre. Each bar represents an amino acid position. The indicated changes are present in more than 75% of the sequenced clones and also present in RecHTLV.
B) Comparison of RecHTLV with the other retrovirus-targeting recombinases, Tre and Brec1. Top: alignment of the loxP, loxLTR, loxBTR and loxHTLV target sequences. The mismatches compared to loxP are coloured in red and the asymmetric positions are underlined. Bottom: amino acid alignment of Cre, Tre, Brec1 and RecHTLV. The alignment was performed using the CLC Genomics Workbench and residues marked in pink indicated non-conserved positions compared to Cre.

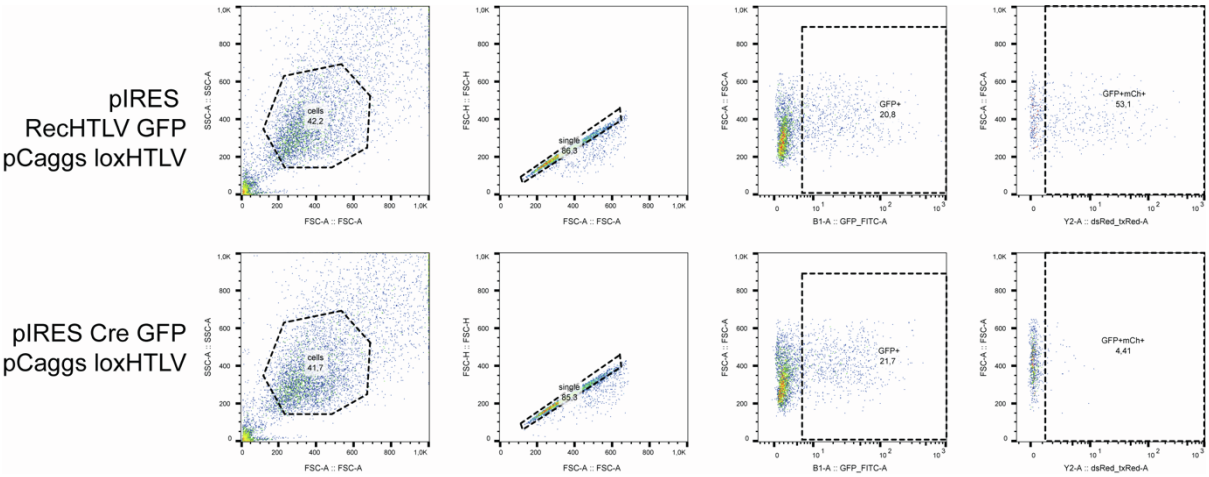
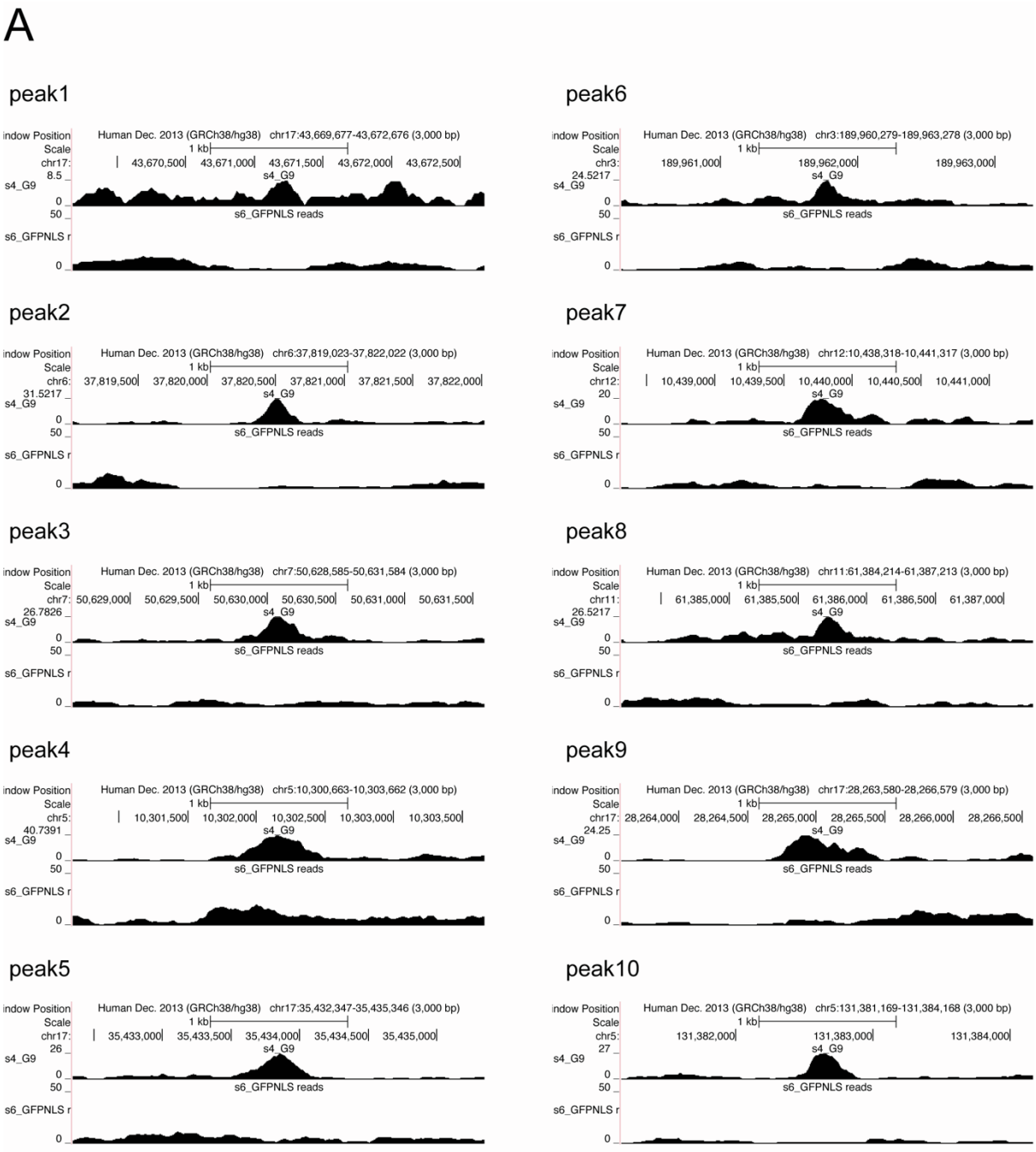



**Figure S4. Gating strategy of the reporter plasmid assay**
Gating strategy used for quantification shown in Figure 3. First, cells were gated for doublet discrimination and the single cells were then gated based on transfection efficiency (GFP+). From the transfected population mCherry+ cells were quantified.

A




B

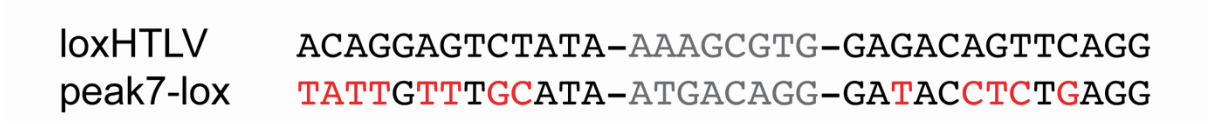



**Figure S5. Analysis of the Chip-seq data**
A) ChIP-seq pileups at the loci of the selected 1 to 10 peaks for the RecHTLV (G9) sample and GFP control. B) Alignment of the loxHTLV sequence to peak7-lox. Mismatches to the loxHTLV sequence are marked in red and the spacers are shown in grey.

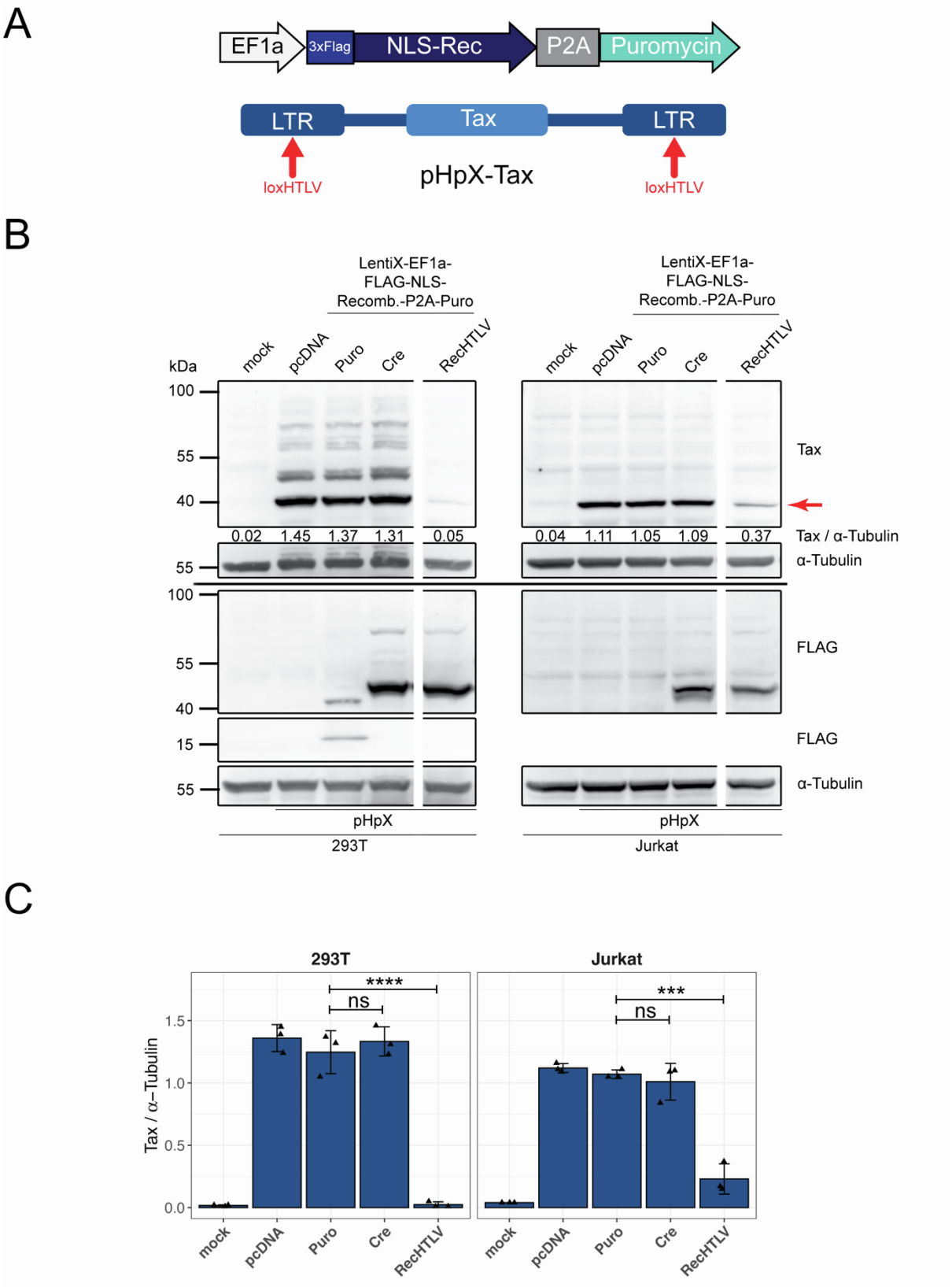



**Figure S6. RecHTLV recombines loxHTLV in the context of the full HTLV-1 LTR sequence**
A) Scheme of the recombinase expression vector and the Tax expression vector pHpX, which is flanked by two LTRs. B) Detection of Tax protein and FLAG-tagged recombinases at 48 hours after transient transfection of 293T cells (left) and Jurkat cells (right) with recombinase expression constructs together with the pHpX-Tax plasmid. Numbers indicate densitometric analysis of Tax protein, normalized to α-Tubulin. Marked with a red arrow is the expected sizes of Tax. C) Densitometric analysis of Tax protein normalized to α-Tubulin. The bar shows the mean ± s.d. of three independent experiments (*** $p<0.001$, **** $p<0.000$, ns: not significant, unpaired two-sided t-test using the logarithm of the normalized values)

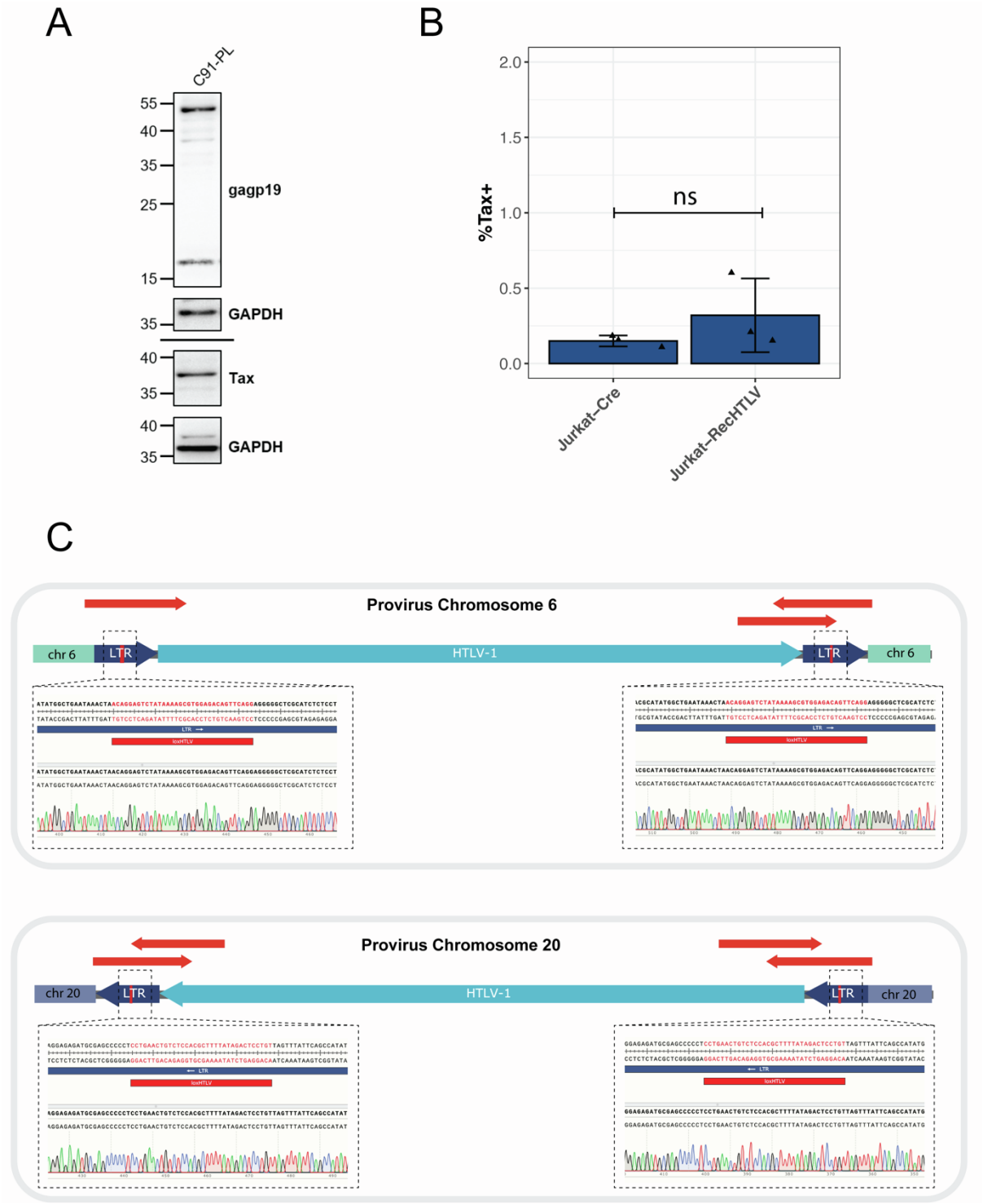



**Figure S7. C91-PL donor cells produce Gag p55/p19 protein and Jurkat acceptor cells do not express Tax protein. loxHTLV is present in SP cells**

A) Detection of Gag p55/p19 protein and Tax protein in C91-PL donor cells. GAPDH served as control. Blots were cut due to technical reasons. B) Background staining on negativity infected control cells shown as the quantification of the frequency of Tax+ cell in Jurkat-acceptor cells Jurkat-Cre and Jurkat-RecHTLV C) Representation of the sequencing strategy and results from the proviral integration in SP cells on chromosomes 6 and 20. The red arrows indicate the extent of the Sanger sequencing reads, which confirm the position of the integrant.

**Table S1. Primer list**
Table containing the list of primers used

**Supplemental References**


1. Ennifar E, Meyer JEW, Buchholz F, Stewart AF, 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*. 2003;31(18):5449-5460. doi:10.1093/nar/gkg732

2. Tomasello G, Armenia I, Molla G. The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics*. 2020;36(9):2909-2911. doi:10.1093/bioinformatics/btaa009