Supplemental Materials and Methods

Plasmid Constructs

The previous backbone of the pACE-GFP plasmid [25,31] was modified by PCR-mediated, oligonucleotide-directed mutagenesis [36] in 6 ways as described in Results and shown in Fig.1 to yield pAC3-GFP (T5.0006). This plasmid, pAC3-GFP (T5.0006) was used as a basis for the vectors encoding cytosine deaminase and variants. The vector T5.0002 was subsequently sequenced to confirm these changes. In addition to the changes 1 through 6 described in the main text, further sequences observations were made as follows. Two silent base difference were found in the gag region (positions 1197 and 1979, in reference to pAC3-yCD2) compared to the canonical MoMLV sequence in NCBI (NC_001501.1, http://www.ncbi.nlm.nih.gov/nuccore/9626958). In addition a single C to T change at position 3201 in the pol gene that substitutes a tyrosine for a histidine residue in comparison to NC_001501.1 was observed. The Friend and Rauscher strains of MLV have the same nucleotide and amino acid substitution in the pol gene compared to NC 001501.1. These differences are probably natural variants of MLV as our sequencing of pZAP2 [31] a derivative of the original infectious MoMLV provirus isolate [37] showed the same polymorphisms. Two methods of inserting the coding sequence cassettes were used. The first method resulted in the sequence 5'TTATAAT3', and the second in the sequence 5'TTATAA3' immediately upstream of the ATG start codon in the IRES. The second method was simpler, as it involved simple PstI1 and Not1 enzyme cuts in the vector and the synthetic cytosine deaminase genes, followed by insertion of the transgene and religation. The vector pAC3-yCD (T5.0007) was made by reconstructing the TTATAAT sequence and so has an unmodified IRES. The vector pACE-CD (T5.0000) is the vector used for mouse studies in Tai et al. [15], and has the sequence TAATACC in the IRES immediately upstream of the CD ATG start codon. The other CD plasmids,

including T5.0002 were made using the simpler Psi1/Not1 excision of GFP from pAC3-GFP and insertion of the corresponding CD or CD-hybrid genes and so have 5'TTATAA3' upstream of the ATG. The modified CD and CD hybrid genes were designed and synthesized with Psi1 and Not1 sites and inserted in the corresponding sites in plasmid pUC57 (Biobasic, Ontario, Canada). Vectors with cytosine deaminase inserts CDopt (yCD1) and the CDopt+3pt (yCD2) (see Table1) were made in both ways with either TTATAAT or TTATAA upstream of the ATG. Infectious virus preparations were made by transient transfection of 293T cells. U87cells were then infected in culture, at a MOI of 0.1, and the cells grown until 100% infected. Cell extracts of 100% infected cells were assayed for cytosine deaminase activity as described in Materials and Methods (see also Figure 3B), and the specific activity of the enzyme was found to be equivalent for constructs with either upstream sequence, that were otherwise identical. Therefore in Table 1, Fig.1, and Fig.S1, pAC3-GFP (T5.0006) and pAC3-yCD (T5.0007) have the first upstream sequence (5'TTATAAT3'), while all other constructs that were further tested have the second (5'TTATAA3'). The sequences of all the modified CD genes are available (PCT publication WO2010045002A2).

Infectious vector production

293T cells were seeded in 6 T-25's, each containing 5 mL of media, at a cell density of 1.8×10^6 cells/T-25 (or 7.2×10^4 cells/cm^2). One day after seeding, six flasks of cells were each transfected with 20μg vector plasmid using the Calcium Phosphate Transfection Kit from Promega (Cat No. E1200). Eighteen hours following transfection, the medium in the flasks was replaced with fresh medium containing 10 mM sodium butyrate. Eight hours post-butyrate treatment, the medium was replaced with fresh medium containing no butyrate. The expression

was allowed to continue until the next day (22 hours duration). The supernatants were harvested after 20-24 h and assayed for their titers by qPCR expressed in Transducing Units (TU)/mL (see below).

Determination of titers of vector preparations

Titers of vector preparations were determined using PC3 cells as target for infection, adding AZT 24h after infection to stop vector replication and counting the number of integrated proviruses in the target cell population by qPCR. PC3 cells were seeded on 12-well plates on Day 0, (12-18 hours prior to transduction) in one mL complete DMEM medium at a concentration of 1.0E5/mL to 2.0E5/mL. On day 1 the PC3 cells were transduced with 10-20 µl of the vector preparations and 10 fold dilutions of those in complete DMEM medium plus 4µg/mL polybrene. Plates were returned to the incubator for 24 hours and on Day 2, 4µL per mL of medium of a stock 10mM solution of AZT was added to the cultures to arrest viral replication. Cells were harvested on day 3, and DNA prepared from the cells in each well. Genomic DNA for use in qPCR was prepared from the cells, using the Promega Maxwell 16 Instrument and associated cartridges, and eluted in 150-180ul. 2 µl was used for concentration determination on the Nanodrop ND-1000. The integrity of the DNA was checked by electrophoresis of 50 – 500ng on a 0.8% ethidium bromide agarose gel (Invitrogen e-gel). qPCR was carried out in triplicate using the following primers and probe (Integrated DNA Technologies): Forward primer 5--MLV-U3-B:5' AGCCCACAACCCCTCACTC; Reverse primer 3-MLV-PSI: 5'TCTCCCGATCCCGGACGA; Probe FAM-5'CCCCAAATGAAAGACCCCCGCTGACG-BHQ, on a Bio-Rad C100 Thermal Cycler equipped with the Bio-Rad CFX96 Real Time System. The amplicon size is 192 bp. These primers and detection probe will only detect the integrated provirus with the 3'UTR transposed to the 5'end of the provirus, and will not detect

any contaminating plasmid DNA used in the transient transfection step to make the infectious virus genome. Standards (8 serial log dilution of the plasmid pAZE-GFP [25,31] and DNA from cells infected with a control vector were included for quantitation and standard no-template controls were run to determine contamination or background. Typically titers were around 5E6 TU/mL in the media supernatants after transient transfection.

Preparation of RNA and cell lysates from infected U87 cells, RT-PCR and Western blot measurements.

U87 cells were harvested from T25 flasks after infection at about 20% confluence and at a MOI of 0.1 on day 0. At days 1, 2, 3, 4, 5, 6, and 7 growth media of T25 flasks containing U87cells transduced with the different vectors (T5.0000, T5.0001, T5.0002,T5.0007) and non-infected U87 cells were harvested. Cell suspension were spun, washed in PBS, divided in 2 vials (1 mL for RNA and 1 mL for Western Blot), and centrifuged 10 min at 4˚C 1000 rpm in a microcentrifuge. The cells were washed again in PBS and cell pellets for RNA samples were frozen in LN_2 and those for Western Blot (WB) were lysed with 20 µL of RIPA buffer (VWR, Cat# 95029-284 Lot# 083005) containing Protease inhibitor. WB samples were incubated on ice for 20 min, then cell pellets and lysates were storage at -80 ˚C until completed collection of all time points to be run in parallel for the RT-PCR and WB assays. RNA was prepared from frozen cell pellets using the Absolutely RNA Miniprep Kit (Stratagene, Cat# 400800, Lot# 0006036834) and was quantified after the column elution step in water, by optical density readings on the Nanodrop 8000. RT-PCR quantitation of relative levels of CD RNA was carried out using the β-actin Housekeeping Gene Primers (Qiagen, Cat# QT01680476) as the relative standard. The PCR reactions were carried out using the iScript One-Step RT-PCR Kit with SYBR Green (Bio-Rad, Cat# 170-8893, Lot# 20119A) on a Bio-Rad C100 Thermal Cycler

equipped with the Bio-Rad CFX96 Real Time System. In order to measure actual CD RNA sequences, the non-humanized genes (T5.0000 and T5.0007) used one set of primers (Forward: 5'-TGGTGTTCCTATTGGCGGATGTCT-3' Reverse: 5'-ATCTCACCATGTAGTGTGGCGGAT-3'**)** and the humanized genes (T5.0001 and T5.0002) another (forward: 5'-TGTGTACCGGCGCTATTCATCATGT-3' reverse: 5'-TCTCATCGTCAACAACCACCACCT-3'). Both these primers give rise to a 108 bp. amplicon. As the relative efficiencies of these primers is unknown, it is not possible to make quantitative comparisons of RNA levels between the humanized and non-humanized samples and the RT-PCR assays can only be qualitatively compared between vectors. For Western Blots, protein concentrations were measured using the Pierce BCA Protein Assay Kit, Pierce (Cat. No. PI23227) and 40 µg of total protein from each sample was loaded onto a 4%-12% discontinuous polyacrylamide gel (Invitrogen, Cat# NP0321BOX), 0.3 µg of purified yeast CD Recombinant Yeast Cytosine Deaminase (Calzyme, Lot# 6-1-53) was run as a positive control and Plus 2 Pre-stained standards (Invitrogen, Cat# C-33250) were run as molecular weight markers. Gels were electroblotted with the iBlot Dry Blotting System (Invitrogen, Cat# IB1001). Non-specific binding by the antibodies was blocked by bathing the membrane in 1X TBST buffer containing 5% nonfat dry milk for 1 h with gentle agitation at room temperature (RT). The primary antibody (sheep polyclonal anti-yCD, ABCAM, Cat# ab35251) was diluted 1/1500 in 10 mL of 1X TBST buffer + 0.5% nonfat dry milk. The membrane was incubated in this fluid overnight at 4°C with constant shaking, then was washed with agitation for 5 min, 5 times with 1X TBST buffer. The secondary antibody (AP-linked rabbit anti-sheep polyclonal antibody ABCAM, Cat# ab6748) was diluted 1/10,000 in 10 mL of 1X TBST buffer + 0.5% nonfat dry milk and the membrane was incubated 1h at RT in constant shaking then was washed with agitation for 5 min, 5 times

with cold 1X TBST buffer and the reaction developed using 3 mL of 1-Step NBT/BCIP Plus Suppressor (Thermo, Cat# 34070). After approximately 5 min the membrane was washed with distilled water and photographed.