

A CRISPR/Cas9 library to map the HIV-1 provirus genetic fitness

K. E. YODER

The Ohio State University College of Medicine, Cancer Biology and Genetics, 460 West 12th Ave. Room 1006, Columbus, OH 43210 USA

Received June 27, 2018; accepted September 27, 2018

Summary. – The integrated proviral genome is the major barrier to a cure for HIV-1 infection. Genome editing technologies, such as CRISPR/Cas9, may disable or remove the HIV-1 provirus by introducing DNA double strand breaks at sequence specific sites in the viral genome. Host DNA repair by the error-prone non-homologous end joining pathway generates mutagenic insertions or deletions at the break. CRISPR/Cas9 editing has been shown to reduce replication competent viral genomes in cell culture, but only a minority of possible genome editing targets have been assayed. Currently there is no map of double strand break genetic fitness for HIV-1 to inform the choice of editing targets. However, CRISPR/Cas9 genome editing makes it possible to target double strand breaks along the length of the provirus to generate a double strand break genetic fitness map. We identified all possible HIV-1 targets with different bacterial species of CRISPR/Cas9. This library of guide RNAs was evaluated for GC content and potential off-target sites in the human genome. Complexity of the library was reduced by eliminating duplicate guide RNA targets in the HIV-1 long terminal repeats and targets in the *env* gene. Although the HIV-1 genome is AT-rich, the *S. pyogenes* CRISPR/Cas9 with the proto-spacer adjacent motif NGG offers the most HIV-1 guide RNAs. This library of HIV-1 guide RNAs may be used to generate a double strand break genetic fragility map to be further applied to any genome editing technology designed for the HIV-1 provirus.

Keywords: HIV-1; genome editing; CRISPR; genetic fitness; guide RNAs

Introduction

As a retrovirus, human immunodeficiency virus 1 (HIV-1) integrates a copy of the viral genome into the host DNA of CD4+ T cells (Coffin *et al.*, 1997). This integrated provirus may become transcriptionally silent, or latent (Finzi *et al.*, 1999). HIV-1 infection can be controlled by treatment with a combination therapy of anti-retroviral drugs. In spite of persistently undetectable viral loads, the HIV-1 latent reservoir

persists for many years and quickly re-establishes infection when anti-retroviral therapy is stopped (Archin *et al.*, 2012; Palmer *et al.*, 2011). Latently infected cells prevent the cure of HIV-1 infection (Siliciano *et al.*, 2003).

Genome editing technologies have been proposed as one strategy for a cure by deleting or disabling the latent HIV-1 provirus (Aubert *et al.*, 2011; Ebina *et al.*, 2013; Hu *et al.*, 2014; Liao *et al.*, 2015; Qu *et al.*, 2013; Zhu *et al.*, 2015). The proposed strategies have included homing endonucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspersed short palindromic repeat (CRISPR/Cas9) (Aubert *et al.*, 2011; Ebina *et al.*, 2013; Manjunath *et al.*, 2013; Qu *et al.*, 2013). All of these genome editors target a double strand break (DSB) to a sequence specific location in target DNA. DSBs are most commonly repaired by error-prone non-homologous end joining (NHEJ), which introduces insertions or deletions (indels) at the repair junction (Lieber,

E-mail: yoder.176@osu.edu; phone: +1-(614)-688-2106.

Abbreviations: CRISPR = clustered regularly interspersed short palindromic repeat; DSB = double strand break; FIV = feline immunodeficiency virus; gRNA = guide RNA; HIV-1 = human immunodeficiency virus 1; LTR = long terminal repeat; NHEJ = non-homologous end joining; NM = *N. meningitidis*; PAM = protospacer adjacent motif; SA = *S. aureus*; SP = *S. pyogenes*; ST1 = *S. thermophilus*

2010). The generation of indels is often used to disrupt the reading frame of a gene of interest. Except for CRISPR/Cas9, all of these genome editing technologies are sequence specific DNA binding proteins, requiring engineering of a protein to alter the target specificity. CRISPR/Cas9 sequence specificity is mediated by a guide RNA (gRNA) that has direct sequence homology to the DNA target. The relative ease of altering the target site of CRISPR/Cas9 editing and the ability to multiplex gRNAs make this technology highly attractive (Cong *et al.*, 2013; Wang *et al.*, 2013).

Clustered regularly interspersed short palindromic repeat is a rudimentary immune system in bacteria that disables foreign genetic elements (Makarova *et al.*, 2011). The *S. pyogenes* (SP) Type II CRISPR has become a common research tool (Shalem *et al.*, 2014; Wang *et al.*, 2013). In this CRISPR system a single protein, Cas9, binds a constant sequence scaffold RNA linked to a variable 20 base gRNA that is homologous to a DNA target site. CRISPR/Cas9 genomic targeting requires a protospacer adjacent motif (PAM) located in the target DNA at the 3' end of the gRNA target site sequence (Sternberg *et al.*, 2014). The PAM is recognized by the Cas9 protein during the target search but is not included in the gRNA sequence. Cas9 contains two endonuclease domains that introduce a blunt end DSB into the genomic DNA 3 bp 5' to the PAM (Gasiunas *et al.*, 2012; Nishimasu *et al.*, 2014). The PAM varies between bacterial species; for example, the SP PAM is NGG, the *N. meningitidis* (NM) PAM is NNNGMTT, the *S. thermophilus* (ST1) Cas9 PAM is NNAGAAW, and the *S. aureus* (SA) Cas9 PAM is NNGRRT (Cong *et al.*, 2013; Esvelt *et al.*, 2013; Garneau *et al.*, 2010; Ran *et al.*, 2015). The HIV-1 genome is AT-rich with <43% GC content, suggesting a Cas9 with a more AT rich PAM might have more targets in the HIV-1 genome.

CRISPR/Cas9 genome editing strategies are intriguing approaches to combat HIV-1 infection. CRISPR/Cas9 targeted to the host HIV-1 co-receptor gene *CCR5* has shown clinical promise, but is still inefficient for HIV-1 eradication (Tebas *et al.*, 2014). In the case of targeting the HIV-1 provirus, NHEJ associated indels could alter the reading frame of the viral genes (Wang *et al.*, 2016a,b,c; Yin *et al.*, 2016). Indels have also been proposed to alter critical RNA secondary structures by altering their coding DNA regions (Ebina *et al.*, 2013; Lebbink *et al.*, 2017; Yin *et al.*, 2016; Yoder and Bundschuh, 2016). In addition, targeting of the long terminal repeat (LTR) sequences that flank the HIV-1 proviral genome has been suggested as a mechanism to delete the viral genome (Ebina *et al.*, 2013; Hu *et al.*, 2014; Kaminski *et al.*, 2016a; Kaminski *et al.*, 2016b). This provirus deletion strategy is likely to be inefficient compared to generation of indels at the two targets independently (Canver *et al.*, 2014). CRISPR/Cas9 has been shown to disable both latent and replicating HIV-1 (Ebina *et al.*, 2013; Hu *et al.*, 2014; Liao *et al.*, 2015; Zhu *et al.*, 2015). Importantly, genome editing by CRISPR/

Cas9 does not necessarily require reactivation of latent HIV-1 proviruses (Zhu *et al.*, 2015). Studies of CRISPR/Cas9 editing of the HIV-1 proviral genome have used sets of 2 to 26 targeting gRNAs and SP Cas9; one study utilized SA Cas9 with 2 gRNAs (Ebina *et al.*, 2013; Hu *et al.*, 2014; Kaminski *et al.*, 2016a; Kaminski *et al.*, 2016b; Lebbink *et al.*, 2017; Liao *et al.*, 2015; Ueda *et al.*, 2016; Wang *et al.*, 2016a,b,c; Yin *et al.*, 2016; Yin *et al.*, 2017; Yoder and Bundschuh, 2016; Zhu *et al.*, 2015). As with any monotherapy, HIV-1 strains resistant to a single CRISPR gRNA arose (Wang *et al.*, 2016b,c; Yoder and Bundschuh, 2016). Some of the observed resistance mutations developed during the mutagenic process of reverse transcription (Wang *et al.*, 2016b; Yoder and Bundschuh, 2016). In addition, CRISPR gRNA resistant strains were generated by the error-prone NHEJ DNA repair of the Cas9 induced DSB (Wang *et al.*, 2016b,c; Yoder and Bundschuh, 2016). Indels at the repair junction conferred resistance to further cleavage and thus resistant strains. In non-coding regions the indels were typically a single bp, while coding regions included 3 bp indels (Yoder and Bundschuh, 2016).

The efficiency of any genome editing strategy of the HIV-1 provirus is inherently dependent on the genetic fragility of the target site. Genetic fitness may vary from fragile to robust throughout the viral genome. Fragile regions of the genome are not tolerant of mutations and reduce viral replication; genetically robust sites tolerate mutations and have little or no effects on viral replication. The measure of genetic fitness is not limited to the type of mutation and may include both substitution mutations and indel mutations (de Visser *et al.*, 2003). However, the current map of the HIV-1 genome fitness landscape is derived from sequencing studies of patient isolates (Zanini *et al.*, 2017). This map of genetic fitness is largely limited to substitution mutations. Previous studies mapped the genetic robustness of HIV-1 *capsid* and *integrase* genes by single amino acid substitutions (Rihn *et al.*, 2015; Rihn *et al.*, 2013). Genetic robustness was measured by the replication capacity of mutant viruses. Although both capsid and integrase proteins form multimer complexes, capsid was found to be more genetically fragile and intolerant of mutations. In capsid 70% of amino acid substitutions inactivated the virus, compared to only 35% of integrase mutations (Rihn *et al.*, 2013, 2015). Thus, integrase is more genetically robust than capsid. The selective pressures of patient infection or engineered substitution mutations do not necessarily reflect the pressures of DSBs and/or indels. Importantly, the map of HIV-1 genetic fitness based on substitution mutations will likely be different than the map based on DSB induced indel mutations.

The concept of DSB genetic fragility has not been applied to the choice of CRISPR gRNAs targeting the HIV-1 genome. The choice of gRNAs has largely been based on predicted Cas9 cleavage efficiency or sequence conservation among HIV-1 subtype B strains (Hu *et al.*, 2014; Kaminski *et al.*, 2016b; Liao *et al.*, 2015; Wang *et al.*, 2016a,b; Yin *et al.*, 2016;

Yoder and Bundschuh, 2016; Zhu *et al.*, 2015). One previous study compared the sequences of gRNA resistant strains to the HIV-1 subtype B patient isolates in the Los Alamos National Laboratory HIV Databases (Yoder and Bundschuh, 2016). Many of the resulting resistant strains displayed single base indels and are not present in the HIV-1 databases of patient isolates (Yoder and Bundschuh, 2016). In contrast, the mutations in HIV-1 strains that developed resistance to siRNA inhibition are present in the HIV-1 patient isolate databases (ter Brake *et al.*, 2008; von Eije *et al.*, 2008). These observations suggest that while a substitution mutation genome fragility map of HIV-1 may successfully be applied to the choice of inhibitory siRNAs, it will have limited utility with CRISPR/Cas9 genome editing. Instead a new DSB associated genetic fragility map of HIV-1 is required as a foundation for genome editing target predictions.

A DSB genetic fitness map of the HIV-1 genome may be derived from a CRISPR/Cas9 gRNA library targeting the provirus. In contrast to scanning alanine mutagenesis which is necessarily limited to protein coding regions, a CRISPR gRNA library has the ability to assay the fragility of non-coding viral sequences. The gRNAs that do not affect HIV-1 replication will be considered to target genetically robust sites of the viral genome. The gRNAs that delay HIV-1 replication will indicate that the targets are genetically fragile sites. We have computationally derived a comprehensive set of CRISPR gRNAs that may be assayed in a 96 well format to model DSB genetic fitness throughout the HIV-1 genome. This gRNA library assay may also be applied to off-target editing. Thorough quantitation and stratification of targets throughout the HIV-1 genome for genetic fragility and potential off-target editing will inform future strategies based on any genome editing technology.

Materials and Methods

The sequences of HIV-1 strains LAI (GenBank K02013.1), NL4-3 (GenBank AF324493.2), BK132 (GenBank AY173951), 671 (GenBank AY423387), and 1058 (GenBank AY331295) were input as fasta sequences to the web-based tool chopchop.cbu.uib.no (Leitner *et al.*, 2005; Montague *et al.*, 2014; Wain-Hobson *et al.*, 1985). HIV-1 strain LAI was analyzed for SP, NM, and ST1 editing with the CRISPR/Cas9 option and for Cpf1 editing with the CRISPR/Cpf1 option. Strains NL4-3, BK132, 671, and 1058 were evaluated for SP targets. The targets were evaluated in the background of the *H. sapiens* (hg38/GRCh38) genome. The PAM sequence options were NGG for SP, NNNGTTT for NM, NNAGAAW for ST1, and NNGRRT for SA. The method for identifying off-target sites in the human genome was set to no more than 0 mismatches in the 15 bases adjacent to the PAM, the protospacer seed region (Cong *et al.*, 2013). The gRNAs were mapped to the respective HIV-1 proviral features based on the GenBank annotations.

Results

Generating a library of gRNAs targeting the HIV-1 genome

LAI was one of the first HIV-1 strains cultured from a patient and subsequently sequenced (Wain-Hobson *et al.*, 1991). The proviral sequence was analyzed for gRNAs with the web-based tool chopchop.cbu.uib.no (Montague *et al.*, 2014; Wain-Hobson *et al.*, 1985). This website allows the analysis of a target sequence for genome editing by Cas9 proteins from multiple bacterial species with different PAM sequence requirements. SP CRISPR/Cas9 with an NGG PAM signal was the first described for use in mammalian cells (Cho *et al.*, 2013; Cong *et al.*, 2013). However, the HIV-1 genome is AT-rich with ~40% GC content. We considered that Cas9 proteins with a more AT rich PAM sequence might generate better coverage of the HIV-1 genome. The ST1 Cas9 PAM is NNAGAAW and the NM Cas9 PAM is NNNGTTT (Esvelt *et al.*, 2013; Garneau *et al.*, 2010). The SA Cas9 with PAM NNGRRT can be AT rich (Ran *et al.*, 2015). These Cas9 proteins function in the same way as the SP Cas9, digesting the target DNA 3 bp 5' of the PAM sequence. The probability of off-target editing was reduced by excluding gRNAs that perfectly matched a genomic sequence in the 3' 15 bp of the gRNA (Cong *et al.*, 2013; Montague *et al.*, 2014). The total gRNAs identified were 1046 targets with SP Cas9, 160 targets with NM Cas9, 83 targets with ST1 Cas9, and 283 targets with SA Cas9 (Table 1). Although the HIV-1 genome is AT-rich, the shorter SP PAM sequence is more abundant.

We also considered an alternative genome editing protein Cpf1 from *Francisella tularensis subsp. novicida* (Zetsche *et al.*, 2015). The Cpf1 PAM TTTN is on the 5' side of the gRNA target sequence. The enzyme generates a staggered cleavage 18–23 base pairs on the 3' side of the PAM. There are 554 Cpf1 gRNA target sites in the HIV-1 LAI genome (Table 1). Surprisingly, the total number of SP Cas9 target gRNAs was greater than any genome editor with an AT rich PAM.

Table 1. CRISPR/Cas9 gRNAs from 5 different bacterial species targeting the HIV-1 LAI genome

	SP	NM	ST1	SA	Cpf1
Total gRNAs	1046	160	83	283	554
GC < 40%	323	60	34	96	255
GC > 70%	26	1	0	6	7
Mismatches	40	0	0	1	7
Duplicates in LTR	71	12	0	16	12
env gene	112	11	13	34	47
Pared gRNAs	474	76	36	130	226
Accessory genes	88	10	11	26	36
Pared gRNAs (HeLa assay)	386	66	25	104	190

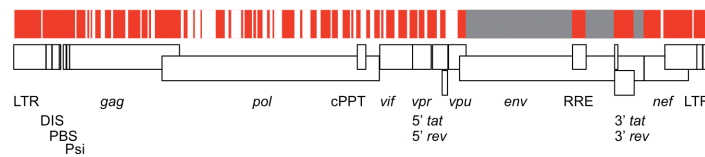


Fig. 1

Scale map of SP gRNAs targeting the HIV-1 genome

A cartoon of the HIV-1 LAI proviral genome is shown to scale in white boxes. Below the genome are the genomic elements, such as the LTRs, dimerization initiation site (DIS), primer binding site (PBS), genomic RNA packaging signal (Psi), central polypurine tract (cPPT), and Rev response element (RRE), as well as the viral genes in italics. *S. pyogenes* (SP gRNA) binding sites are shown in red. Gaps in gRNA coverage ≥ 20 bp are white spaces. Regions of the *env* gene that are excluded from the gRNA library are grey.

The gRNA targets were further analysed, and several criteria were used to exclude gRNAs (Table 1). The lists were pared by removing gRNAs with G/C content $<40\%$ or $>70\%$ since they are likely to be inefficient. All gRNAs with ≥ 1 mismatch with a genomic off-target site were excluded. The gRNA list includes targets in both LTRs; duplicate gRNAs were discarded. The HIV-1 *env* gene has been shown to be genetically divergent in patients and tolerant of mutations; gRNAs targeting the *env* gene were excluded (Hahn *et al.*, 1986). However, regions of *env* that also encode *tat*, *rev*, or the Rev response element (RRE) RNA stem-loop structure were included. According to these criteria, the final numbers of gRNAs were 474 SP Cas9 gRNAs, 76 NM Cas9 gRNAs, 36 ST1 Cas9 gRNAs, 130 SA Cas9 gRNAs, and 226 Cpf1 gRNAs (Table 1, Supplementary Table 1).

While the SP Cas9 library of gRNAs appears to offer the best coverage of the HIV-1 genome, some of the gRNA sequences overlap. In order to better visualize coverage of the LAI genome, SP gRNAs were mapped to scale (Fig. 1). There were 89 gaps in the SP gRNA coverage of the HIV-1 LAI genome. Most of the gaps were < 20 bp. The LTR sequences that flank the HIV-1 genome had only one 21 bp gap. The *pol* gene displayed the most gaps in SP gRNA coverage with only 48% of the gene targeted by 145 gRNAs. Since Cpf1 targets were more abundant than other Cas9 targets, we considered whether the Cpf1 gRNA coverage of *pol* might be more complete than SP gRNAs. There are 103 Cpf1 gRNAs

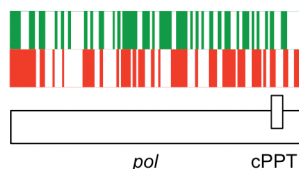


Fig. 2

Scale map of SP and Cpf1 gRNAs targeting the HIV-1 LAI *pol* gene

A cartoon of the LAI *pol* gene is shown as a box with the region also encoding the cPPT genetic element indicated by a smaller box. The SP gRNA recognition sites are shown to scale in red. Cpf1 gRNA binding sites are in green. Gaps between gRNAs ≥ 20 bp are white space.

targeting 47% of the *pol* gene. Mapping the Cpf1 gRNAs to the LAI *pol* gene reveals that many of the gaps in coverage are common to SP Cas9 and Cpf1 (Fig. 2). Combined libraries of Cpf1 and SP gRNAs would only cover 67.5% of the HIV-1 LAI *pol* gene. The SP gRNA library affords the most comprehensive editing of the LAI genome providing for the determination of a DSB associated genetic fitness map of the HIV-1 genome.

Application of the LAI gRNA library to other HIV-1 strains

As a screening tool, this HIV-1 gRNA library could conceivably be used to analyze other HIV-1 strains. We compared the gRNA library to the commonly used HIV-1 laboratory strain NL4-3 (Table 2). In addition, sequences of patient-derived reference strains BK132 from Thailand, 671 from Netherlands, and 1058 from USA were analyzed for complementarity with the LAI gRNA library (Table 2). Laboratory strain NL4-3 has 98% sequence identity to LAI, the highest of the strains evaluated. However, only 288 of 474 LAI gRNAs also targeted NL4-3. The patient derived strains display less sequence identity to LAI. Only 194, 141, or 113 LAI gRNAs also target BK132, 671, or 1058, respectively. As the LAI sequence identity of each strain decreased, the number of complementary LAI gRNAs also decreased. In contrast, an inverse correlation was observed between LAI sequence identity and the number of gRNAs unique to each strain. Further analysis of the patient strains and LAI revealed that only 53 gRNA target sequences are common

Table 2. Comparison of LAI to HIV-1 strains NL4-3, BK132, 671, and 1058

Strain	LAI identity	LAI gRNAs	Unique gRNAs
NL4-3	98%	288	193
BK132	94%	194	256
671	92%	141	337
1058	92%	113	339

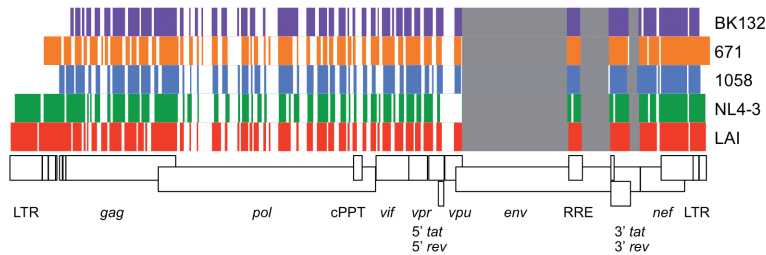


Fig. 3

Comparison of SP gRNAs targeting multiple HIV-1 strains

A scale map of the HIV-1 genome is shown in white boxes. SP Cas9 gRNA libraries targeting five HIV-1 strains are shown. Colored bars indicate gRNA binding sites. White spaces indicate gaps in coverage ≥ 20 bp. Regions of the *env* gene that are excluded from the gRNA library are grey. Libraries of gRNAs targeting HIV-1 strains LAI (red), NL4-3 (green), and patient strains BK132 (purple), 671 (orange), and 1058 (blue) are shown.

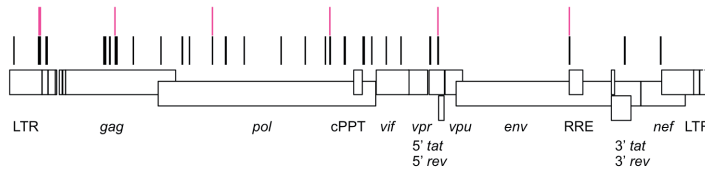


Fig. 4

SP gRNAs at conserved HIV-1 sequences

A scale map of the HIV-1 genome is shown in white boxes. SP Cas9 gRNAs that target five HIV-1 strains, shown in Fig. 3, are indicated by black lines. Previously characterized SP gRNAs are shown in pink.

to these 4 strains. These results suggest that the LAI gRNA library would not be an effective screening tool for other viral strains or patient quasi-species.

The patient strains and NL4-3 were evaluated for SP gRNA targets with the same methodology as LAI (Table 3). In all cases, the total number of possible gRNAs for each strain was less than LAI, ranging from 925 to 1029 total gRNAs identified. This is partially due to the incomplete LTR sequences for the patient strains. The pared lists of gRNAs ranging from 450 to 481 gRNAs were nearly equal or slightly greater than 474 LAI gRNAs. However, the total number of gRNAs is not necessarily indicative of correlation in the viral genomes. While the LAI gRNA library is likely not suitable for screening other viral strains, none of the gRNA libraries

from these other strains displayed any superiority to the LAI gRNA library.

The gRNA libraries were mapped to scale for each genome (Fig. 3). Comparison of the gRNA maps for all of the strains shows impressive correlation of coverage. All of the strains show significant gaps in coverage of the *pol* and *vpu* genes. In contrast, there are few gaps in coverage at several regions. For example, the LTRs offer many gRNA targets. We have previously targeted the LTRs with five different gRNAs and found that resistance mutations readily occur at these non-coding regions (Yoder and Bundschuh, 2016). The sequence surrounding the TATA box and TAR RNA stem loop coding regions are highly conserved suggesting they are genetically fragile and intolerant of mutations. However, when targeted by CRISPR/Cas9 these sequences proved to be genetically robust and tolerant of indels. More intriguing targets of the HIV-1 genome are those that have overlapping reading frames or RNA elements. For example, the regions of *env* that also encode the RRE RNA stem loop or that also encode the 3' exons of *tat* and *rev*. Although the specific gRNA sequences are not conserved between HIV-1 strains, comparison of the gRNA maps suggests the DSB genetic fragility map derived from LAI may be applied to other HIV-1 strains.

While the gRNA coverage maps appear similar, there are only 41 gRNA sequences in common between all strains

Table 3. SP gRNAs in NL4-3, BK132, 671, and 1058

	NL4-3	BK132	HIV 671	HIV 1058
Total gRNAs	1029	939	996	925
GC < 40%	331	313	365	321
GC > 70%	24	23	15	16
Mismatches	47	33	45	38
Duplicates in LTR	36	0	4	0
<i>env</i> gene	110	120	89	100
Pared gRNAs	481	450	478	450

analyzed. When mapped to the HIV-1 genome, it is readily apparent that these gRNAs occur throughout the provirus; they are not clustered in one gene or region (Fig. 4). Several of these gRNAs of highly conserved sequences have been previously characterized (Wang *et al.*, 2016b,c; Yin *et al.*, 2016; Yoder and Bundschuh, 2016; Zhu *et al.*, 2015). Although the sequence conservation suggests genetic robustness, these sites proved in all cases to not be the strongest inhibitors of HIV-1 replication. Thus, the DSB genetic fragility appears to be demonstrably different from the currently available genetic fitness maps based on sequence conservation and substitution mutations.

Discussion

CRISPR/Cas9 genome editing is a relatively new technology that creates DSBs in a genome of interest (Cong *et al.*, 2013; Esvelt *et al.*, 2013; Hsu *et al.*, 2013; Kim and Kim, 2014). The advantage of CRISPR/Cas9 genome editing compared to zinc finger nucleases or TALENs is the inherent ease of redirecting the DSB targeting via a gRNA rather than protein engineering. Generating a library of gRNAs is facile and offers the possibility of comparing multiple target sites in a high throughput screen. Our computational modeling identified 474 SP CRISPR/Cas9 gRNA targets throughout the HIV-1 LAI reference genome. This CRISPR/Cas9 gRNA library can induce DSBs throughout the length of the HIV-1 genome exerting novel pressures to select for resistant viral strains. The HIV-1 proviral genome does not experience this kind of DSB assault during infection of a patient. Thus, the HIV-1 genetic fitness map generated by patient samples may not faithfully recapitulate the fitness associated with induced DSBs. Direct comparison of these gRNAs can generate a new HIV-1 DSB genetic fitness map that will guide targeting of future genome editing strategies.

The HIV-1 LAI gRNA library proposed here may be implemented in a 96 well format. To evaluate the DSB genetic fitness of the HIV-1 genome, the CRISPR/Cas9 genome editing should occur in cells with replicating virus. Genetic fitness under pressure of genome editing DSBs is measured by HIV-1 replication dynamics in the presence of the gRNA compared to its absence. The gRNA library oligonucleotide synthesis and subcloning may be commercially subcontracted and distributed in a 96 well plate format. The cell type for screening the gRNA library should permit viral replication while also being amenable to transfection. Human CD4+ T cell lines readily allow HIV-1 replication, but these suspension cell lines are typically difficult to transfect. HeLa cells have been engineered to express the human *CD4* gene, the receptor for HIV-1 entry (Chesebro *et al.*, 1990; Chesebro and Wehrly, 1988; Maddon *et al.*, 1986). Adherent HeLa-CD4 cells allow HIV-1 replication and may be

efficiently transfected with the CRISPR/Cas9 gRNA library with lipid-based reagents. One important consideration of commercially available CRISPR/Cas9 delivery systems is that the lentiviral backbone may also be targeted by the gRNA library. The CRISPR/Cas9 system should be in a vector that does not include HIV-1 lentiviral sequences, which could be a similar feline immunodeficiency virus (FIV) based lentiviral vector (Poeschla *et al.*, 1998). This alternative lentiviral vector has similar delivery properties as an HIV-1 based vector but will not be targeted by the HIV-1 gRNAs. FIV vectors may also encode a puromycin resistance gene allowing selection of successfully transduced cells. Transduced HeLa-CD4 cells may be selected with puromycin generating stable HeLa-CD4-Cas9 cells.

One limitation of performing this screen in a HeLa derived cell line is that the accessory genes of HIV-1 – *vif*, *vpr*, *vpu* or *nef* – will not be functional in these cells. Although the gRNA library will target these genes, there will be no true selective pressure at these sites. However, HeLa cell lines are highly amenable to the development of library screening tools due to their ease of transduction and transfection (Brass *et al.*, 2008; Espeseth *et al.*, 2011; Zhou *et al.*, 2008). Removal of gRNAs that will be non-functional in HeLa cells reduces the total number of SP gRNAs targeting LAI to 389. Once this CRISPR gRNA library screen has been developed in HeLa cells, it may be adapted to different cell types, such as human CD4+ T cells, that will provide a selective test of the accessory genes.

The quantitation of HIV-1 replication in HeLa-CD4-Cas9 cells may be performed in a 96 well plate format. Virus replication dynamics over time may be measured in two ways. First, the relative amount of infectious virus may be determined by a cellular assay, such as TZM-bl indicator cells, a HeLa derived cell line that expresses beta-galactosidase and luciferase when infected with HIV-1 (Derdeyn *et al.*, 2000; Platt *et al.*, 1998, 2009; Platt *et al.*, 1998; Takeuchi *et al.*, 2008; Wei *et al.*, 2002). Infectious HIV-1 virions released by the HeLa-CD4-Cas9 cells may be measured by transferring media to TZM-bl indicator cells in a 96 well plate. Quantitation of luciferase or beta-galactosidase activity is a relative measure of infectious virus. Second, HIV-1 virus particles in the media may be directly measured by an ELISA for the capsid protein p24. Together these two assays allow for the quantitation of infectious activity of virions and of virus particles, which may be different. For example, successful genome editing of the HIV-1 *integrase* gene may allow secretion of virus particles to the media, but the defective integrase protein will render them noninfectious, and will not produce a signal in TZM-bl cells. Alternatively, genome editing of the HIV-1 promoter may prevent production of viral proteins and thus viral particles, ultimately resulting in reduced measurements in both the ELISA and TZM-bl assays. To stratify the gRNA library, primary weight is given

to data from cellular infections. For gRNAs at genetically robust sites, viral replication will be equal to untreated cells. In contrast, cultures that continue to suppress viral replication for extended times will have gRNAs targeting genetically fragile sites that are intolerant of mutations. In these cases, ELISA quantitation of virus particles could stratify these gRNA targets. In some cases, there will be no detectable virus particles. However, it is possible that virus particles are produced, but are not infectious, such as an integrase or reverse transcriptase mutant virus. Together these values will rank the DSB genetic fitness of the HIV-1 genome.

One significant complication of all genome editing technologies is their propensity to target additional spurious sites (Cong *et al.*, 2013). The human genome encodes endogenous retrovirus elements that are potential off-target editing sites for HIV-1 gRNAs. Although the HIV-1 gRNA library has been specifically designed to reduce off-target editing, the library should be empirically tested. Several previous reports of HIV-1 genome editing have relied on the Surveyor nuclease assay to test off-target editing. In this assay, potential off-target sites are predicted *in silico* (Hu *et al.*, 2014; Liao *et al.*, 2015; Yin *et al.*, 2017; Yoder and Bundschuh, 2016). This assay is inherently biased by the prediction of off-target sites *in silico*. Alternative methods to assay for off-target editing depend on deep sequencing or whole genome sequencing. These methods are not feasible for a library of gRNAs. Instead, an unbiased assay adaptable to a multi-well format is preferred. An unbiased high throughput assay for off-target genome editing takes advantage of the predilection for DSB repair to insert exogenous fragments of DNA non-specifically (Gabriel *et al.*, 2011; Wang *et al.*, 2015b). First, cells are transiently transfected with the *Cas9* gene and the gRNA. After 6 and 24 hours, the cells are transduced with an integrase-defective lentiviral vector encoding a puromycin selection gene. Importantly, this lentiviral vector should not be derived from HIV-1; instead an FIV based vector is more appropriate to avoid editing of the puromycin delivery genome. If off-target editing occurs, then the DSB may be repaired with the lentiviral genome encoding puromycin. Two days following transduction, the cells are put into drug selection. Following selection, the remaining viable cells may be measured by crystal violet staining. Only HIV-1 gRNAs that also edit the human genome at off-target sites will produce puromycin resistant cells. This assay is not biased by *in silico* prediction of off-target sites and is easily performed in multiwell plates. HIV-1 gRNAs that produce puromycin resistant cells will indicate regions of the HIV-1 genome that are significantly associated with off-target editing in the human genome. While these data do not inform the map of genetic fitness, it does indicate regions of the HIV-1 genome that should not be explored for future genome editing.

HIV-1 readily develops resistance to any mono-therapy, including single CRISPR/Cas9 gRNAs (Wang *et al.*, 2016b,c;

Yoder and Bundschuh, 2016). Typically, mutations that confer resistance to a single anti-retroviral drug are generated during the error-prone reverse transcription of the viral single strand RNA genome to a double strand cDNA. Resistance to CRISPR/Cas9 genome editing is due to mutations induced by both reverse transcription and the error-prone DSB repair NHEJ pathway (Wang *et al.*, 2016b,c; Yoder and Bundschuh, 2016). Thus, any single gRNA is likely to be inefficient in patient therapy. Two gRNAs capable of long-term suppression of replication of an HIV-1 lab strain in a cell line have been reported (Wang *et al.*, 2016a,b). These studies identified highly fragile DSB sites in the *gag* and *pol* genes. These gRNA targets do not display the highest sequence conservation among HIV-1 subtype B isolates, underscoring the concept that sequence conservation and DSB associated fragility are independent. While resistance to these gRNAs did not develop under constrained laboratory conditions, a single gRNA is not likely to be effective for patient treatment. Instead, multiple CRISPR/Cas9 gRNAs that target multiple fragile sites could be multiplexed to effectively disable the HIV-1 quasispecies present in a patient. The sequence divergence within patients will likely necessitate a multiplexed genome editing system (Panfil *et al.*, 2018). Currently the CRISPR/Cas9 system is the only genome editor capable of efficient multiplexing.

The major obstacle to an HIV-1 cure is the latent reservoir of long-lived cells with an integrated provirus (Palmer *et al.*, 2011; Siliciano *et al.*, 2003). The latent reservoir is not well understood, but these cells are known to be rare. Studies in cell lines suggest that CRISPR/Cas9 may effectively edit the latent provirus (Ebina *et al.*, 2013; Hu *et al.*, 2014; Zhu *et al.*, 2015). Unfortunately, the use of CRISPR/Cas9 genome editing is limited by effective cellular targeting technologies. It is not clear how an effective CRISPR/Cas9 editor could be delivered to latently infected cells in a patient. However, novel CRISPR/Cas9 delivery schemes continue to be developed (Chiou *et al.*, 2015; Cottle *et al.*, 2015; Han *et al.*, 2015; Wang *et al.*, 2015a). Whether CRISPR/Cas9 genome editing of the HIV-1 provirus will prove to be clinically useful is unknown. Any alternative genome editing technology targeting the integrated provirus will benefit from a DSB genetic fragility map of the HIV-1 genome.

Supplementary information is available in the online version of the paper.

References

- Archin NM, Vaidya NK, Kuruc JD, Liberty AL, Wiegand A, Kearney ME, Cohen MS, Coffin JM, Bosch RJ, Gay CL, Eron JJ, Margolis DM, Perelson AS (2012): Immediate antiviral therapy appears to restrict resting CD4+ cell HIV-1 in-

- fection without accelerating the decay of latent infection. *Proc. Natl. Acad. Sci. USA* 109, 9523–9528. <https://doi.org/10.1073/pnas.1120248109>
- Aubert M, Ryu BY, Banks L, Rawlings DJ, Scharenberg AM, Jerome KR (2011): Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI. *PLoS One* 6, e16825. <https://doi.org/10.1371/journal.pone.0016825>
- Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ (2008): Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319, 921–926. <https://doi.org/10.1126/science.1152725>
- Canver MC, Bauer DE, Dass A, Yien YY, Chung J, Masuda T, Maeda T, Paw BH, Orkin SH (2014): Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* 289, 21312–21324. <https://doi.org/10.1074/jbc.M114.564625>
- Chesebro B, Buller R, Portis J, Wehrly K (1990): Failure of human immunodeficiency virus entry and infection in CD4-positive human brain and skin cells. *J. Virol.* 64, 215–221. <https://doi.org/10.1093/infdis/163.1.64>
- Chesebro B, Wehrly K (1988): Development of a sensitive quantitative focal assay for human immunodeficiency virus infectivity. *J. Virol.* 62, 3779–3788.
- Chiou SH, Winters IP, Wang J, Naranjo S, Dudgeon C, Tamburini FB, Brady JJ, Yang D, Gruner BM, Chuang CH, Caswell DR, Zeng H, Chu P, Kim GE, Carpizo DR, Kim SK, Winslow MM (2015): Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes Dev.* 29, 1576–1585. <https://doi.org/10.1101/gad.264861.115>
- Cho SW, Kim S, Kim JM, Kim JS (2013): Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232. <https://doi.org/10.1038/nbt.2507>
- Coffin JM, Hughes SH, Varmus HE (1997): *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013): Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. <https://doi.org/10.1126/science.1231143>
- Cottle RN, Lee CM, Archer D, Bao G (2015): Controlled delivery of beta-globin-targeting TALENs and CRISPR/Cas9 into mammalian cells for genome editing using microinjection. *Sci. Rep.* 5, 16031. <https://doi.org/10.1038/srep16031>
- de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, Gibson G, Hansen TF, Krakauer D, Lewontin RC, Ofria C, Rice SH, von Dassow G, Wagner A, Whitlock MC (2003): Perspective: Evolution and detection of genetic robustness. *Evolution* 57, 1959–1972. <https://doi.org/10.1111/j.0014-3820.2003.tb00377.x>
- Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, Kappes JC, Shaw GM, Hunter E (2000): Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. *J. Virol.* 74, 8358–8367. <https://doi.org/10.1128/JVI.74.18.8358-8367.2000>
- Ebina H, Misawa N, Kanemura Y, Koyanagi Y (2013): Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep.* 3, 2510. <https://doi.org/10.1038/srep02510>
- Espeseth AS, Fishel R, Hazuda D, Huang Q, Xu M, Yoder K, and Zhou H (2011): siRNA screening of a targeted library of DNA repair factors in HIV infection reveals a role for base excision repair in HIV integration. *PLoS One* 6, e17612. <https://doi.org/10.1371/journal.pone.0017612>
- Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM (2013): Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10, 1116–1121. <https://doi.org/10.1038/nmeth.2681>
- Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, Smith K, Lisziewicz J, Lori F, Flexner C, Quinn TC, Chaisson RE, Rosenberg E, Walker B, Gange S, Gallant J, Siliciano RF (1999): Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* 5, 512–517. <https://doi.org/10.1038/8394>
- Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppe C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G, Holmes, MC, Gregory PD, Glimm H, Schmidt M, Naldini L, von Kalle C (2011): An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 29, 816–823. <https://doi.org/10.1038/nbt.1948>
- Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S (2010): The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71. <https://doi.org/10.1038/nature09523>
- Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012): Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* 109, E2579–2586. <https://doi.org/10.1073/pnas.1208507109>
- Hahn BH, Shaw GM, Taylor ME, Redfield RR, Markham PD, Salahuddin SZ, Wong-Staal F, Gallo RC, Parks ES, Parks WP (1986): Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* 232, 1548–1553. <https://doi.org/10.1126/science.3012778>
- Han X, Liu Z, Jo MC, Zhang K, Li Y, Zeng Z, Li N, Zu Y, Qin L (2015): CRISPR-Cas9 delivery to hard-to-transfect cells via membrane deformation. *Sci. Adv.* 1, e1500454. <https://doi.org/10.1126/sciadv.1500454>
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F (2013): DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. <https://doi.org/10.1038/nbt.2647>
- Hu W, Kaminski R, Yang F, Zhang Y, Cosentino L, Li F, Luo B, Alvarez-Carbonell D, Garcia-Mesa Y, Karn J, Mo X, Khalili K (2014): RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 111, 11461–11466. <https://doi.org/10.1073/pnas.1405186111>

- Kaminski R, Bella R, Yin C, Otte J, Ferrante P, Gendelman HE, Li H, Booze R, Gordon J, Hu W, Khalili K (2016a): Excision of HIV-1 DNA by gene editing: a proof-of-concept in vivo study. *Gene Ther.* 23, 690–695. <https://doi.org/10.1038/gt.2016.41>
- Kaminski R, Chen Y, Fischer T, Tedaldi E, Napoli A, Zhang Y, Karn J, Hu W, Khalili K (2016b): Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing. *Sci. Rep.* 6, 22555. <https://doi.org/10.1038/srep28213>
- Kim H, Kim JS (2014): A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 15, 321–334. <https://doi.org/10.1038/nrg3686>
- Lebbink RJ, de Jong DC, Wolters F, Kruse EM, van Ham PM, Wiertz EJ, Nijhuis M (2017): A combinational CRISPR/Cas9 gene-editing approach can halt HIV replication and prevent viral escape. *Sci. Rep.* 7, 41968. <https://doi.org/10.1038/srep41968>
- Leitner T, Korber B, Daniels M, Calef C, Foley B (2005): HIV-1 Subtype and Circulating Recombinant Form (CRF) Reference Sequences, 2005, 41–48. <https://doi.org/10.1111/j.0031-5990.2005.00003.x>
- Liao HK, Gu Y, Diaz A, Marlett J, Takahashi Y, Li M, Suzuki K, Xu R, Hishida T, Chang CJ, Esteban CR, Young J, Izpisua Belmonte JC (2015): Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells. *Nat. Commun.* 6, 6413. <https://doi.org/10.1038/ncomms7413>
- Lieber MR (2010): The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211. <https://doi.org/10.1146/annurev.biochem.052308.093131>
- Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R (1986): The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 47, 333–348. [https://doi.org/10.1016/0092-8674\(86\)90590-8](https://doi.org/10.1016/0092-8674(86)90590-8)
- Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011): Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. <https://doi.org/10.1038/nrmicro2577>
- Manjunath N, Yi G, Dang Y, Shankar P (2013): Newer gene editing technologies toward HIV gene therapy. *Viruses* 5, 2748–2766. <https://doi.org/10.3390/v5112748>
- Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E (2014): CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* 42, W401–407. <https://doi.org/10.1093/nar/gku410>
- Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O (2014): Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>
- Palmer S, Josefsson L, Coffin JM (2011): HIV reservoirs and the possibility of a cure for HIV infection. *J. Intern. Med.* 270, 550–560. <https://doi.org/10.1111/j.1365-2796.2011.02457.x>
- Panfil AR, London JA, Green PL, Yoder KE (2018): CRISPR/Cas9 Genome Editing to Disable the Latent HIV-1 Provirus. *Front Microbiol.* 9, 3107. <https://doi.org/10.3389/fmicb.2018.03107>
- Platt EJ, Bilska M, Kozak SL, Kabat D, Montefiori DC (2009): Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1. *J. Virol.* 83, 8289–8292. <https://doi.org/10.1128/JVI.00709-09>
- Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D (1998): Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J. Virol.* 72, 2855–2864.
- Poeschla EM, Wong-Staal F, Looney DJ (1998): Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors. *Nat. Med.* 4, 354–357. <https://doi.org/10.1038/nm0398-354>
- Qu X, Wang P, Ding D, Li L, Wang H, Ma L, Zhou X, Liu S, Lin S, Wang X, Zhang G, Liu S, Liu L, Wang J, Zhang F, Lu D, Zhu H (2013): 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. <https://doi.org/10.1093/nar/gkt571>
- Ran FA, Cong L, Yan WX, Scott DA, Gootenberg, JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F (2015): In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520, 186–191. <https://doi.org/10.1038/nature14299>
- Rihn SJ, Hughes J, Wilson SJ, Bieniasz PD (2015): Uneven genetic robustness of HIV-1 integrase. *J. Virol.* 89, 552–567. <https://doi.org/10.1128/JVI.02451-14>
- Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD (2013): Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog.* 9, e1003461. <https://doi.org/10.1371/journal.ppat.1003461>
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014): Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87. <https://doi.org/10.1126/science.1247005>
- Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ, Siliciano RF (2003): Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med.* 9, 727–728. <https://doi.org/10.1038/nm880>
- Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014): DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67. <https://doi.org/10.1038/nature13011>
- Takeuchi Y, McClure MO, Pizzato M (2008): Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. *J. Virol.* 82, 12585–12588. <https://doi.org/10.1128/JVI.01726-08>
- Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G, Holmes MC, Gregory PD, Ando DG, Kalos M, Collman RG, Binder-Scholl G, Plesa G, Hwang WT, Levine BL, June CH (2014): Gene editing of CCR5 in autologous CD4 T cells of persons

- infected with HIV. *N. Engl. J. Med.* 370, 901–910. <https://doi.org/10.1056/NEJMoa1300662>
- ter Brake O, von Eije KJ, Berkhout B (2008): Probing the sequence space available for HIV-1 evolution. *AIDS* 22, 1875–1877. <https://doi.org/10.1097/QAD.0b013e328309efe3>
- Ueda S, Ebina H, Kanemura Y, Misawa N, Koyanagi Y (2016): Anti-HIV-1 potency of the CRISPR/Cas9 system insufficient to fully inhibit viral replication. *Microbiol. Immunol.* 60, 483–496. <https://doi.org/10.1111/1348-0421.12395>
- von Eije KJ, ter Brake O, Berkhout B (2008): Human immunodeficiency virus type 1 escape is restricted when conserved genome sequences are targeted by RNA interference. *J. Virol.* 82, 2895–2903. <https://doi.org/10.1128/JVI.02035-07>
- Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M (1985): Nucleotide sequence of the AIDS virus, LAV. *Cell* 40, 9–17. [https://doi.org/10.1016/0092-8674\(85\)90303-4](https://doi.org/10.1016/0092-8674(85)90303-4)
- Wain-Hobson S, Vartanian JP, Henry M, Chenciner N, Cheynier R, Delassus S, Martins LP, Sala M, Nugeyre MT, Guetard D et al. (1991): LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur. *Science* 252, 961–965. <https://doi.org/10.1126/science.2035026>
- Wang D, Mou H, Li S, Li Y, Hough S, Tran K, Li J, Yin H, Anderson DG, Sontheimer EJ, Weng Z, Gao G, Xue W (2015a): Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum. Gene Ther.* 26, 432–442. <https://doi.org/10.1089/hum.2015.087>
- Wang G, Zhao N, Berkhout B, Das AT (2016a): A Combinatorial CRISPR-Cas9 Attack on HIV-1 DNA Extinguishes All Infectious Provirus in Infected T Cell Cultures. *Cell. Rep.* 17, 2819–2826. <https://doi.org/10.1016/j.celrep.2016.11.057>
- Wang G, Zhao N, Berkhout B, Das AT (2016b): CRISPR-Cas9 Can Inhibit HIV-1 Replication but NHEJ Repair Facilitates Virus Escape. *Mol. Ther.* 24, 522–526. <https://doi.org/10.1038/mt.2016.24>
- Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013): One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918. <https://doi.org/10.1016/j.cell.2013.04.025>
- Wang X, Wang Y, Wu X, Wang J, Wang Y, Qiu Z, Chang T, Huang H, Lin RJ, Yee JK (2015b): Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* 33, 175–178. <https://doi.org/10.1038/nbt.3127>
- Wang Z, Pan Q, Gendron P, Zhu W, Guo F, Cen S, Wainberg MA, Liang C (2016c): CRISPR/Cas9-Derived Mutations Both Inhibit HIV-1 Replication and Accelerate Viral Escape. *Cell. Rep.* 15, 481–489. <https://doi.org/10.1016/j.celrep.2016.03.042>
- Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, Saag MS, Wu X, Shaw GM, Kappes JC (2002): Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* 46, 1896–1905. <https://doi.org/10.1128/AAC.46.6.1896-1905.2002>
- Yin C, Zhang T, Li F, Yang F, Putatunda R, Young WB, Khalili K, Hu W, Zhang Y (2016): Functional screening of guide RNAs targeting the regulatory and structural HIV-1 viral genome for a cure of AIDS. *AIDS* 30, 1163–1174. <https://doi.org/10.1097/QAD.0000000000001079>
- Yin C, Zhang T, Qu X, Zhang Y, Putatunda R, Xiao X, Li F, Xiao W, Zhao H, Dai S, Qin X, Mo X, Young WB, Khalili K, Hu W (2017): In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models. *Mol. Ther.* 25, 1168–1186. <https://doi.org/10.1016/j.ymthe.2017.03.012>
- Yoder KE, Bundschuh R (2016): Host Double Strand Break Repair Generates HIV-1 Strains Resistant to CRISPR/Cas9. *Sci. Rep.* 6, 29530. <https://doi.org/10.1038/srep29530>
- Zanini F, Puller V, Brodin J, Albert J, Neher RA (2017): In vivo mutation rates and the landscape of fitness costs of HIV-1. *Virus Evol.* 3, vex003. <https://doi.org/10.1093/ve/vex003>
- Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Markarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2015): Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>
- Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS (2008): Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 4, 495–504. <https://doi.org/10.1016/j.chom.2008.10.004>
- Zhu W, Lei R, Le Duff Y, Li J, Guo F, Wainberg MA, Liang C (2015): The CRISPR/Cas9 system inactivates latent HIV-1 proviral DNA. *Retrovirology* 12, 22. <https://doi.org/10.1186/s12977-015-0150-z>

Supplementary information

A CRISPR/Cas9 library to map the HIV-1 provirus genetic fitness

K. E. YODER

The Ohio State University College of Medicine, Cancer Biology and Genetics, 460 West 12th Ave. Room 1006,
Columbus, OH 43210 USA

Received June 27, 2018; accepted September 27, 2018

LAI Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
CAAGGATATCTTGTCTTCGTTGG	sequence:20, 9152	-	40	LTR U3 NEF
GACAAGATATCCTTGATCTGTGG	sequence:28, 9160	+	40	LTR U3 NEF
TGTGGTAGATCCACAGATCAAGG	sequence:38, 9170	-	45	LTR U3 NEF
CTGTGGATCTACCACACACAAGG	sequence:45, 9177	+	50	LTR U3 NEF
CAGGGAAGTAGCCTTGTGTGTGG	sequence:56, 9188	-	55	LTR U3 NEF
CACAAGGCTACTTCCCTGATTGG	sequence:61, 9193	+	50	LTR U3 NEF
GTGTGTAGTTCTGCCAATCAGGG	sequence:74, 9206	-	45	LTR U3 NEF
GGTGTGTAGTTCTGCCAATCAGG	sequence:75, 9207	-	50	LTR U3 NEF
GATTGGCAGAACTACACACCAGG	sequence:78, 9210	+	50	LTR U3 NEF
ATTGGCAGAACTACACACCAGGG	sequence:79, 9211	+	45	LTR U3 NEF
CAGAACTACACACCAGGGCCAGG	sequence:84, 9216	+	60	LTR U3 NEF
AGAACTACACACCAGGGCCAGGG	sequence:85, 9217	+	55	LTR U3 NEF
GAACTACACACCAGGGCCAGGGG	sequence:86, 9218	+	60	LTR U3 NEF
GGATATCTGACCCCTGGCCCTGG	sequence:96, 9228	-	65	LTR U3 NEF
GTCAGTGGATATCTGACCCCTGG	sequence:102, 9234	-	55	LTR U3 NEF
GTCAGATATCCACTGACCTTTGG	sequence:108, 9240	+	45	LTR U3 NEF
GATATCCACTGACCTTTGGATGG	sequence:112, 9244	+	45	LTR U3 NEF
TAGCTTGTAGCACCATCCAAAGG	sequence:124, 9256	-	45	LTR U3 NEF
AGTACCAGTTGAGCCAGATAAGG	sequence:146, 9278	+	45	LTR U3 NEF
TCTACCTTATCTGGCTCAACTGG	sequence:150, 9282	-	45	LTR U3 NEF
TGAGCCAGATAAGGTAGAAGAGG	sequence:155, 9287	+	45	LTR U3 NEF
TTGGCCTCTTCTACCTTATCTGG	sequence:159, 9291	-	45	LTR U3 NEF
AAGGTAGAAGAGGCCAATAAAGG	sequence:165, 9297	+	40	LTR U3 NEF
GCTGGTGTCTCTCCTTTATTGG	sequence:178, 9310	-	45	LTR U3 NEF
GCTCACAGGGTGTAAACAAGCTGG	sequence:196, 9328	-	55	LTR U3 NEF
TTACACCTGTGAGCCTGCATGG	sequence:204, 9336	+	55	LTR U3 NEF
CCCTGTGAGCCTGCATGGAATGG	sequence:209, 9341	+	60	LTR U3 NEF
TCCATTCCATGCAGGCTCACAGG	sequence:210, 9342	-	55	LTR U3 NEF
CAGGGTCATCCATTCCATGCAGG	sequence:218, 9350	-	55	LTR U3 NEF
ACTCTAACACTTCTCTCAGGG	sequence:236, 9368	-	40	LTR U3 NEF
CTGAGAGAGAAGTGTAGAGTGG	sequence:238, 9370	+	45	LTR U3 NEF
AGAGAGAAGTGTAGAGTGGAGG	sequence:241, 9373	+	45	LTR U3 NEF
CCGCCTAGCATTTCATCACGTGG	sequence:272, 9404	+	55	LTR U3 NEF
GGGCCACGTGATGAAATGCTAGG	sequence:275, 9407	-	55	LTR U3 NEF
CGTGGCCCGAGAGCTGCATCCGG	sequence:290, 9422	+	70	LTR U3 NEF
GTACTIONGGATGCAGCTCTCGGG	sequence:295, 9427	-	60	LTR U3 NEF
AGTACTCCGGATGCAGCTCTCGG	sequence:296, 9428	-	55	LTR U3 NEF
CAGCAGTTCTTGAAGTACTCCGG	sequence:309, 9441	-	45	LTR U3 NEF
CTGACATCGAGCTTGCTACAAGG	sequence:329, 9461	+	50	LTR U3
TGACATCGAGCTTGCTACAAGGG	sequence:330, 9462	+	45	LTR U3
TGCTACAAGGGACTTTCCGCTGG	sequence:342, 9474	+	55	LTR U3
GCTACAAGGGACTTTCCGCTGGG	sequence:343, 9475	+	55	LTR U3
CTACAAGGGACTTTCCGCTGGGG	sequence:344, 9476	+	55	LTR U3
CCTTCCGCTGGGGACTTTCCAGG	sequence:354, 9486	+	60	LTR U3
TTTCCGCTGGGGACTTTCCAGGG	sequence:355, 9487	+	55	LTR U3
CCGCTGGGGACTTTCCAGGGAGG	sequence:358, 9490	+	70	LTR U3
GGGGACTTTCCAGGGAGGGCTGG	sequence:363, 9495	+	70	LTR U3
CCTTCCAGGGAGGGCTGGCCTGG	sequence:368, 9500	+	70	LTR U3
TTTCCAGGGAGGGCTGGCCTGGG	sequence:369, 9501	+	65	LTR U3
TGCTTATATGCAGCATCTGAGGG	sequence:412, 9544	-	40	LTR U3
CTGCTTATATGCAGCATCTGAGG	sequence:413, 9545	-	45	LTR U3
CAGCTGCTTTTTGCCTGTACTGG	sequence:433, 9565	+	50	LTR U3
AGCTGCTTTTTGCCTGTACTGGG	sequence:434, 9566	+	45	LTR U3
TTGCCTGTACTGGGTCTCTCTGG	sequence:443, 9575	+	55	LTR U3
TAACCAGAGAGACCCAGTACAGG	sequence:446, 9578	-	50	LTR U3
GGTTAGACCAGATTTGAGCCTGG	sequence:464, 9596	+	50	LTR R
GTTAGACCAGATTTGAGCCTGGG	sequence:465, 9597	+	45	LTR R
AGAGCTCCCAGGCTCAAATCTGG	sequence:471, 9603	-	55	LTR R
ATTTGAGCCTGGGAGCTCTCTGG	sequence:475, 9607	+	55	LTR R
TAGTTAGCCAGAGAGCTCCCAGG	sequence:482, 9614	-	55	LTR R
TGGGAGCTCTCTGGCTAACTAGG	sequence:484, 9616	+	55	LTR R
GGGAGCTCTCTGGCTAACTAGGG	sequence:485, 9617	+	55	LTR R
CCTTATTGAGGCTTAAGCAGTGG	sequence:511, 9643	-	40	LTR R
ACTCAAGGCAAGCTTTATTGAGG	sequence:523, 9655	-	40	LTR R
CACACTACTTGAAGCACTCAAGG	sequence:538, 9670	-	45	LTR U5
GCCCGTCTGTTGTGTGACTCTGG	sequence:562, 9694	+	60	LTR U5
ACCAGAGTCACACAACAGACGGG	sequence:563, 9695	-	50	LTR U5
TACCAGAGTCACACAACAGACGG	sequence:564, 9696	-	45	LTR U5
CACTGACTAAAAGGGTCTGAGGG	sequence:596, 9728	-	45	LTR U5
ACACTGACTAAAAGGGTCTGAGG	sequence:597, 9729	-	45	LTR U5

TCAGACCCTTTTAGTCAGTGTGG	sequence:599, 9731	+	45	LTR U5
GTGTGGAAAATCTCTAGCAGTGG	sequence:616	+	45	LTR U5
TCTAGCAGTGGCGCCCGAACAGG	sequence:628	+	65	LTR U5
CTAGCAGTGGCGCCCGAACAGGG	sequence:629	+	65	LTR U5
TCGCTTTCAAGTCCCTGTTCGGG	sequence:641	-	50	PBS
TTCGCTTTCAAGTCCCTGTTCGG	sequence:642	-	45	PBS
AACAGGGACTTGAAAGCGAAAGG	sequence:645	+	45	PBS
ACAGGGACTTGAAAGCGAAAGGG	sequence:646	+	45	PBS
GAAAGCGAAAGGGAAACCAGAGG	sequence:656	+	50	
CTGCGTCGAGAGAGCTCCTCTGG	sequence:672	-	65	
CAGAGGAGCTCTCTCGACGCAGG	sequence:673	+	65	
AGCTCTCTCGACGCAGGACTCGG	sequence:679	+	60	
TCGGCTTGCTGAAGCGCGCACGG	sequence:698	+	65	DIS
GCTGAAGCGCGCACGGCAAGAGG	sequence:705	+	70	DIS
AAGAGGCGAGGGGAGGCGACTGG	sequence:722	+	70	DIS
CCAAAAATTTGACTAGCGGAGG	sequence:753	+	40	PSI
TTTGACTAGCGGAGGCTAGAAGG	sequence:761	+	50	PSI
GAGGCTAGAAGGAGAGAGATGGG	sequence:772	+	50	PSI
GCGAGAGCGTCAGTATTAAGCGG	sequence:796	+	50	GAG
CGAGAGCGTCAGTATTAAGCGGG	sequence:797	+	50	GAG
GAGAGCGTCAGTATTAAGCGGGG	sequence:798	+	50	GAG
AGAGCGTCAGTATTAAGCGGGGG	sequence:799	+	50	GAG
GCGGGGGAGAATTAGATCGATGG	sequence:815	+	55	GAG
CGGGGGAGAATTAGATCGATGGG	sequence:816	+	50	GAG
GAAAAAATTCGGTTAAGGCCAGG	sequence:838	+	40	GAG
AAAATTCGGTTAAGGCCAGGGGG	sequence:841	+	45	GAG
AACATATAGTATGGGCAAGCAGG	sequence:884	+	40	GAG
ACATATAGTATGGGCAAGCAGGG	sequence:885	+	40	GAG
GAACGATTCGCAGTTAATCCTGG	sequence:913	+	45	GAG
TCTGATGTTTCTAACAGGCCAGG	sequence:931	-	45	GAG
GGCCTGTTAGAAACATCAGAAGG	sequence:934	+	45	GAG
AGCCTTCTGATGTTTCTAACAGG	sequence:936	-	40	GAG
AGAAGGCTGTAGACAAATACTGG	sequence:951	+	40	GAG
GAAGGCTGTAGACAAATACTGGG	sequence:952	+	40	GAG
CTACAACCATCCCTTCAGACAGG	sequence:979	+	50	GAG
TCTGATCCTGTCTGAAGGGATGG	sequence:985	-	50	GAG
TTCTTCTGATCCTGTCTGAAGGG	sequence:989	-	40	GAG
GTTCTTCTGATCCTGTCTGAAGG	sequence:990	-	45	GAG
CCCTCTATTGTGTGCATCAAAGG	sequence:1040	+	45	GAG
TCCTTTGATGCACACAATAGAGG	sequence:1041	-	40	GAG
GGAAGCTTTAGACAAGATAGAGG	sequence:1083	+	40	GAG
TGACACAGGACACAGCAGCCAGG	sequence:1149	+	60	GAG
TAGGGTAATTTGGCTGACCTGG	sequence:1167	-	45	GAG
CCCTATAGTGCAGAACATCCAGG	sequence:1185	+	50	GAG
CCTATAGTGCAGAACATCCAGGG	sequence:1186	+	45	GAG
CTATAGTGCAGAACATCCAGGGG	sequence:1187	+	45	GAG
GCAGAACATCCAGGGGCAAATGG	sequence:1194	+	55	GAG
CCAGGGGCAAATGGTACATCAGG	sequence:1203	+	55	GAG
GAAAACATGGGTATCACTTCTGG	sequence:1285	-	40	GAG
GGTGTAAATCTTGTGGGGTGG	sequence:1325	-	45	GAG
AAACACCATGCTAAACACAGTGG	sequence:1341	+	40	GAG
AACACCATGCTAAACACAGTGGG	sequence:1342	+	40	GAG
ACACCATGCTAAACACAGTGGGG	sequence:1343	+	45	GAG
CACCATGCTAAACACAGTGGGGG	sequence:1344	+	50	GAG
ACCATGCTAAACACAGTGGGGGG	sequence:1345	+	50	GAG
TCCCCCACTGTGTTTAGCATGG	sequence:1346	-	55	GAG
TTCTGCAGCTTCTCATTGATGG	sequence:1400	-	45	GAG
TCAATGAGGAAGCTGCAGAATGG	sequence:1403	+	45	GAG
CAATGAGGAAGCTGCAGAATGGG	sequence:1404	+	45	GAG
AGAGTGCATCCAGTGCATGCAGG	sequence:1429	+	55	GAG
GAGTGCATCCAGTGCATGCAGGG	sequence:1430	+	55	GAG
GCAATAGGCCCTGCATGCACTGG	sequence:1438	-	60	GAG
CATGCAGGGCCTATTGCACCAGG	sequence:1444	+	60	GAG
CTCATCTGGCCTGGTGAATAGG	sequence:1453	-	55	GAG
CTTGTTCTCTCATCTGGCCTGG	sequence:1462	-	55	GAG
CAGGCCAGATGAGAGAACCAAGG	sequence:1463	+	55	GAG
AGGCCAGATGAGAGAACCAAGGG	sequence:1464	+	50	GAG
GGCCAGATGAGAGAACCAAGGGG	sequence:1465	+	55	GAG
TTCCCTTGGTTCTCTCATCTGG	sequence:1467	-	50	GAG
CCAAGGGGAAGTGACATAGCAGG	sequence:1480	+	55	GAG
AGGAACTACTAGTACCCTTCAGG	sequence:1500	+	45	GAG

AGTACCCTTCAGGAACAAATAGG	sequence:1510	+	40	GAG
CCCTTCAGGAACAAATAGGATGG	sequence:1514	+	45	GAG
TCCATCCTATTTGTTCTGAAGG	sequence:1515	-	40	GAG
AATAATCCACCTATCCCAGTAGG	sequence:1543	+	40	GAG
ATTTCTCCTACTGGGATAGGTGG	sequence:1549	-	45	GAG
GTATAGCCCTACCAGCATTCTGG	sequence:1617	+	50	GAG
TTATGTCCAGAATGCTGGTAGGG	sequence:1623	-	40	GAG
CTTATGTCCAGAATGCTGGTAGG	sequence:1624	-	45	GAG
AGCATTCTGGACATAAGACAAGG	sequence:1630	+	40	GAG
CGGTCTACATAGTCTCTAAAGGG	sequence:1663	-	40	GAG
CCTTTAGAGACTATGTAGACCGG	sequence:1664	+	40	GAG
AAGAGCCGAGCAAGCTTCACAGG	sequence:1701	+	55	GAG
AGCCGAGCAAGCTTCACAGGAGG	sequence:1704	+	60	GAG
TACCTCCTGTGAAGCTTGCTCGG	sequence:1706	-	50	GAG
TTGGATGACAGAAACCTTGTTGG	sequence:1734	+	40	GAG
GTTTCGATTTTGGACCAACAAGG	sequence:1748	-	45	GAG
TACAATCTGGGTTTCGATTTTGG	sequence:1758	-	40	GAG
ATTTCTTCTAGTGTAGCTGCTGG	sequence:1804	-	40	GAG
AGAAATGATGACAGCATGTCAGG	sequence:1821	+	40	GAG
GAAATGATGACAGCATGTCAGGG	sequence:1822	+	40	GAG
GATGACAGCATGTCAGGGAGTGG	sequence:1827	+	55	GAG
ATGACAGCATGTCAGGGAGTGGG	sequence:1828	+	50	GAG
ACAGCATGTCAGGGAGTGGGAGG	sequence:1831	+	60	GAG
TGTCAGGGAGTGGGAGGACCCGG	sequence:1837	+	65	GAG
AGTGGGAGGACCCGGCCATAAGG	sequence:1845	+	65	GAG
AAAACCTTGCCTTATGGCCGGG	sequence:1855	-	45	GAG
CAAACTCTTGCCTTATGGCCGG	sequence:1856	-	45	GAG
CGGCCATAAGGCAAGAGTTTTGG	sequence:1857	+	50	GAG
CAGCCAAAACCTTGCCTTATGG	sequence:1860	-	45	GAG
GCTACCATAATGATGCAAAGAGG	sequence:1909	+	40	GAG
GGCACATAGCCAGAAATTGCAGG	sequence:1985	+	50	GAG
GCACATAGCCAGAAATTGCAGGG	sequence:1986	+	45	GAG
CCAGAAATTGCAGGGCCCCTAGG	sequence:1994	+	60	GAG
TTGCAGGGCCCCTAGGAAAAAGG	sequence:2001	+	55	GAG
TGCAGGGCCCCTAGGAAAAAGGG	sequence:2002	+	55	GAG
CCCCTAGGAAAAAGGGCTGTTGG	sequence:2009	+	55	GAG
TCCAACAGCCCTTTTTCTAGGG	sequence:2010	-	45	GAG
TTCCAACAGCCCTTTTTCTAGG	sequence:2011	-	45	GAG
AAAAAGGGCTGTTGGAAATGTGG	sequence:2017	+	40	GAG
GGGCTGTTGGAAATGTGGAAAGG	sequence:2022	+	50	GAG
TGTTGGAAATGTGGAAAGGAAGG	sequence:2026	+	40	GAG
GAAAGATTGTAAGTACTGAGAGACAGG	sequence:2058	+	40	GAG
GAAGATCTGGCCTTCTACAAGG	sequence:2094	+	50	GAG POL
AAGATCTGGCCTTCTACAAGGG	sequence:2095	+	45	GAG POL
TCTGGCCTTCTACAAGGGAAGG	sequence:2099	+	55	GAG POL
CCTTCTACAAGGGAAGGCCAGG	sequence:2104	+	60	GAG POL
CTTCTACAAGGGAAGGCCAGGG	sequence:2105	+	55	GAG POL
ATTCCCTGGCCTTCCCTGTAGG	sequence:2108	-	55	GAG POL
CTGCTCTGAAGAAATTCCTGG	sequence:2122	-	45	GAG POL
AATGGTGGGGCTGTTGGCTCTGG	sequence:2146	-	60	GAG POL
TGAAGAAATGGTGGGGCTGTTGG	sequence:2152	-	50	GAG POL
TCTGCTCTGAAGAAATGGTGGGG	sequence:2159	-	45	GAG POL
GTCTGCTCTGAAGAAATGGTGGG	sequence:2160	-	45	GAG POL
GGTCTGCTCTGAAGAAATGGTGG	sequence:2161	-	50	GAG POL
TCTGGTCTGCTCTGAAGAAATGG	sequence:2164	-	45	GAG POL
TCTGGTGGGGCTGTTGGCTCTGG	sequence:2182	-	65	GAG POL
CTCTCTTCTGGTGGGGCTGTTGG	sequence:2188	-	60	GAG POL
CCCCACCAGAAGAGAGCTTCAGG	sequence:2195	+	60	GAG POL
ACCTGAAGCTCTCTTCTGGTGGG	sequence:2196	-	50	GAG POL
GACCTGAAGCTCTCTTCTGGTGG	sequence:2197	-	55	GAG POL
CCAGAAGAGAGCTTCAGGTCTGG	sequence:2200	+	55	GAG POL
CAGAAGAGAGCTTCAGGTCTGGG	sequence:2201	+	50	GAG POL
AGAAGAGAGCTTCAGGTCTGGGG	sequence:2202	+	50	GAG POL
AACAACCTCCTCTCAGAAGCAGG	sequence:2232	+	50	GAG POL
ATCGGCTCCTGCTTCTGAGAGGG	sequence:2239	-	55	GAG POL
TATCGGCTCCTGCTTCTGAGAGG	sequence:2240	-	55	GAG POL
GAAGCAGGAGCCGATAGACAAGG	sequence:2247	+	55	GAG POL
GGATACAGTTCCTGTCTATCGG	sequence:2257	-	40	GAG POL
AGTGATCTGAGGGAAGTTAAAGG	sequence:2278	-	40	GAG POL
ACTTCCCTCAGATCACTCTTTGG	sequence:2284	+	45	GAG POL
GTTGCCAAAGAGTGATCTGAGGG	sequence:2288	-	45	GAG POL

CGTTGCCAAAGAGTGATCTGAGG	sequence:2289	-	50	GAG POL
CCCCTCGTCACAATAAAGATAGG	sequence:2313	+	45	GAG POL
CCCTCGTCACAATAAAGATAGGG	sequence:2314	+	40	GAG POL
CCTCGTCACAATAAAGATAGGGG	sequence:2315	+	40	GAG POL
CTCGTCACAATAAAGATAGGGGG	sequence:2316	+	40	GAG POL
TCGTCACAATAAAGATAGGGGGG	sequence:2317	+	40	GAG POL
AAAGATAGGGGGGCAACTAAAGG	sequence:2327	+	45	GAG POL
AAATGAGTTTGCCAGGAAGATGG	sequence:2392	+	40	POL
CCAAAAATGATAGGGGGAATTGG	sequence:2418	+	40	POL
AAAATGATAGGGGGAATTGGAGG	sequence:2421	+	40	POL
CCTACACCTGTCAACATAATTGG	sequence:2523	+	40	POL
ACTTTTGGGCCATCCATTCCTGG	sequence:2625	-	50	POL
GACTTCTGGGAAGTTCAATTAGG	sequence:2841	+	40	POL
TTAGGAATACCACATCCCGCAGG	sequence:2859	+	50	POL
TAGGAATACCACATCCCGCAGGG	sequence:2860	+	50	POL
TTTTTTAACCTGCGGGATGTGG	sequence:2868	-	45	POL
ATCAGTAACAGTACTGGATGTGG	sequence:2897	+	40	POL
TCAGTAACAGTACTGGATGTGGG	sequence:2898	+	40	POL
TTCCCTTAGATGAAGACTTCAGG	sequence:2938	+	40	POL
CTTCCTGAAGTCTTCATCTAAGG	sequence:2941	-	40	POL
TCAGTACAATGTGCTTCCACAGG	sequence:3017	+	45	POL
CAGTACAATGTGCTTCCACAGGG	sequence:3018	+	45	POL
ACAATGTGCTTCCACAGGGATGG	sequence:3022	+	50	POL
GTGCTTCCACAGGGATGGAAAGG	sequence:3027	+	55	POL
GGTGATCCTTTCCATCCCTGTGG	sequence:3033	-	55	POL
AGCTGAGACAACATCTGTTGAGG	sequence:3196	+	45	POL
TGAGACAACATCTGTTGAGGTGG	sequence:3199	+	45	POL
GAGACAACATCTGTTGAGGTGGG	sequence:3200	+	45	POL
AGACAACATCTGTTGAGGTGGGG	sequence:3201	+	45	POL
AGAAAGAACCTCCATTCTTTGG	sequence:3250	+	40	POL
AGAACCTCCATTCTTTGGATGG	sequence:3254	+	45	POL
GAACCTCCATTCTTTGGATGGG	sequence:3255	+	45	POL
TCATAACCCATCCAAAGGAATGG	sequence:3261	-	40	POL
GGAGTTCATAACCCATCCAAAGG	sequence:3266	-	45	POL
GTAAGTCCATTTATCAGGATGG	sequence:3287	-	40	POL
GGCTGTACTGTCCATTTATCAGG	sequence:3291	-	45	POL
TCTTTTTCTGGCAGCACTATAGG	sequence:3312	-	40	POL
TGCTGCCAGAAAAGACAGCTGG	sequence:3319	+	50	POL
ACAGTCCAGCTGTCTTTTCTGG	sequence:3324	-	45	POL
TGGGCAAGTCAGATTTACCCAGG	sequence:3381	+	50	POL
GGGCAAGTCAGATTTACCCAGGG	sequence:3382	+	50	POL
TTAGTGCTTTGGTTCCTCTAAGG	sequence:3431	-	40	POL
AGCTCTGCTTCTCTGTTAGTGG	sequence:3465	-	45	POL
AGAAGAAGCAGAGCTAGAACTGG	sequence:3473	+	45	POL
TAGCAGAAATACAGAAGCAGGGG	sequence:3562	+	40	POL
GAAATACAGAAGCAGGGGCAAGG	sequence:3567	+	50	POL
AGAAGCAGGGGCAAGGCCAATGG	sequence:3574	+	60	POL
CAGGAAAATATGCAAGAACGAGG	sequence:3637	+	40	POL
AGGAAACATGGGAAACATGGTGG	sequence:3769	+	45	POL
AAACATGGTGGACAGAGTATTGG	sequence:3781	+	40	POL
CAGAGTATTGGCAAGCCACCTGG	sequence:3793	+	55	POL
AAGCCACCTGGATTCTGAGTGG	sequence:3805	+	55	POL
AGCCACCTGGATTCTGAGTGGG	sequence:3806	+	55	POL
AACTCCCACTCAGGAATCCAGG	sequence:3811	-	50	POL
GTATTGACAACTCCCACTCAGG	sequence:3819	-	45	POL
CCCCTCCTTAGTGAAATTATGG	sequence:3841	+	40	POL
CTATGGGTTCTTCTCTAACTGG	sequence:3866	-	40	POL
AACGTTTCTGCTCCTACTATGGG	sequence:3882	-	40	POL
GAACGTTTCTGCTCCTACTATGG	sequence:3883	-	45	POL
GCAGAAACGTTCTATGTAGATGG	sequence:3894	+	40	POL
ATGTAGATGGGGCAGCTAGCAGG	sequence:3907	+	55	POL
TGTAGATGGGGCAGCTAGCAGGG	sequence:3908	+	55	POL
GCTAGCAGGGAGACTAAATTAGG	sequence:3921	+	45	POL
AGCAATTCATCTAGCTTTGCAGG	sequence:4025	+	40	POL
TCATCTAGCTTTGCAGGATTCGG	sequence:4031	+	40	POL
CATCTAGCTTTGCAGGATTCGGG	sequence:4032	+	45	POL
AGGAAAAGGTCTATCTGGCATGG	sequence:4168	+	45	POL
GGAAAAGGTCTATCTGGCATGGG	sequence:4169	+	45	POL
CCAGCACACAAAGGAATTGGAGG	sequence:4194	+	50	POL
AATTAGTCAGTGCTGGAATCAGG	sequence:4234	+	40	POL
TCACAGTAATTGGAGAGCAATGG	sequence:4310	+	40	POL

CTTTTGCTACTACAGGTGGCAGG	sequence:4346	-	50	POL
TAGCTGACATTTATCACAGCTGG	sequence:4378	-	40	POL
CTAAAAGGAGAAGCCATGCATGG	sequence:4398	+	45	POL
GGACAAGTAGACTGTAGTCCAGG	sequence:4419	+	50	POL
TAGACTGTAGTCCAGGAATATGG	sequence:4426	+	40	POL
CAATCTAGTTGCCATATTCCTGG	sequence:4437	-	40	POL
TGGCTACATGAACTGCTACCAGG	sequence:4484	-	50	POL
GTAGCAGTTCATGTAGCCAGTGG	sequence:4488	+	50	POL
GAAGTTATTCCAGCAGAAACAGG	sequence:4524	+	40	POL
TATTCCAGCAGAAACAGGGCAGG	sequence:4529	+	50	POL
CAGCAATTCACCACTACTACGG	sequence:4619	+	40	POL
TTTCACCAGTACTACGGTTAAGG	sequence:4625	+	40	POL
GGCGGCCTTAACCGTAGTACTGG	sequence:4630	-	60	POL
CTACGGTTAAGGCCGCTGTTGG	sequence:4636	+	60	POL
CGGTTAAGGCCGCTGTTGGTGG	sequence:4639	+	65	POL
GGTTAAGGCCGCTGTTGGTGGG	sequence:4640	+	60	POL
TAAGGCCGCTGTTGGTGGGCGG	sequence:4643	+	65	POL
AAGGCCGCTGTTGGTGGGCGGG	sequence:4644	+	70	POL
CTTGATTCGCGCCACCAACAGG	sequence:4651	-	60	POL
TTGGTGGGCGGGAATCAAGCAGG	sequence:4655	+	60	POL
CCCTACAATCCCAAAGTCAAGG	sequence:4689	+	50	POL
TCCTTGACTTTGGGGATTGTAGG	sequence:4690	-	45	POL
TCTACTACTCCTTGACTTTGGGG	sequence:4698	-	40	POL
TATAGGCCAGGTAAGAGATCAGG	sequence:4748	+	45	POL
GTTCAGCCTGATCTTACCTGG	sequence:4754	-	50	POL
TAAAAGAAAAGGGGGGATTGGGG	sequence:4820	+	40	POL
AAAAGAAAAGGGGGGATTGGGGG	sequence:4821	+	45	POL
AAAGAAAAGGGGGGATTGGGGGG	sequence:4822	+	50	POL
GGGATTGGGGGGTACAGTGCAGG	sequence:4833	+	65	POL
GGATTGGGGGGTACAGTGCAGGG	sequence:4834	+	60	POL
GATTGGGGGGTACAGTGCAGGGG	sequence:4835	+	60	POL
GATTCGCGCCACCAACAGGCGG	sequence:4648	-	65	POL
GGGACAGCAGAGATCCACTTTGG	sequence:4948	+	55	POL
AGCAGAGATCCACTTTGGAAAGG	sequence:4953	+	45	POL
TTTGCTGGTCCTTTCCAAAGTGG	sequence:4962	-	45	POL
AAGGACCAGCAAAGCTCCTCTGG	sequence:4972	+	55	POL
CCAGCAAAGCTCCTCTGGAAAGG	sequence:4977	+	55	POL
AAGCTCCTCTGGAAAGGTGAAGG	sequence:4983	+	50	POL
AGCTCCTCTGGAAAGGTGAAGGG	sequence:4984	+	50	POL
GCTCCTCTGGAAAGGTGAAGGGG	sequence:4985	+	55	POL
CTGCCCCCTCACCTTTCCAGAGG	sequence:4988	-	60	POL
GATTATGGAAAACAGATGGCAGG	sequence:5073	+	40	POL
TGATTGTGTGGCAAGTAGACAGG	sequence:5099	+	45	POL
TGTGGCAAGTAGACAGGATGAGG	sequence:5105	+	50	POL
ACAGGATGAGGATTAGAACATGG	sequence:5117	+	40	POL
TGTATGTTTCAGGGAAAGCTAGG	sequence:5162	+	40	VIF
GTATGTTTCAGGGAAAGCTAGGG	sequence:5163	+	40	VIF
TATGTTTCAGGGAAAGCTAGGGG	sequence:5164	+	40	VIF
TTTCAGGGAAAGCTAGGGGATGG	sequence:5168	+	50	VIF
TCAGAAGTACACATCCCCTAGG	sequence:5233	+	45	VIF
CAGAAGTACACATCCCCTAGGG	sequence:5234	+	45	VIF
AGAAGTACACATCCCCTAGGGG	sequence:5235	+	45	VIF
CCCCTAGGGGATGCTAGATTGG	sequence:5247	+	55	VIF
ACCAATCTAGCATCCCCTAGTGG	sequence:5248	-	50	VIF
ACATATTGGGGTCTGCATACAGG	sequence:5278	+	45	VIF
TGCATACAGGAGAAAGAGACTGG	sequence:5291	+	45	VIF
AGGAGAAAGAGACTGGCATCTGG	sequence:5298	+	50	VIF
GGAGAAAGAGACTGGCATCTGGG	sequence:5299	+	50	VIF
AAGAGACTGGCATCTGGGTCAGG	sequence:5304	+	55	VIF
AGAGACTGGCATCTGGGTCAGGG	sequence:5305	+	55	VIF
GTCAGGGAGTCTCCATAGAATGG	sequence:5321	+	50	VIF
AGGGAGTCTCCATAGAATGGAGG	sequence:5324	+	50	VIF
TTAGTTGGTCTGCTAGTTCAGGG	sequence:5373	-	40	VIF
ATTAGTTGGTCTGCTAGTTCAGG	sequence:5374	-	40	VIF
TAGGACATATAGTTAGCCCTAGG	sequence:5450	+	40	VIF
CTGCTTGATATTCACACCTAGGG	sequence:5466	-	40	VIF
CCTAGGTGTGAATATCAAGCAGG	sequence:5467	+	45	VIF
CAAGCAGGACATAACAAGGTAGG	sequence:5482	+	45	VIF
GGTAGGATCTCTACAATACTTGG	sequence:5499	+	40	VIF
TTCGTAACACTAGGCAAAGGTGG	sequence:5557	-	45	VIF
AGTTTCGTAACACTAGGCAAAGG	sequence:5560	-	40	VIF

TCTGTCAGTTTCGTAACACTAGG	sequence:5566	-	40	VIF
TAGTGTTACGAAACTGACAGAGG	sequence:5568	+	40	VIF
CGAAACTGACAGAGGATAGATGG	sequence:5576	+	45	VIF
GAACAAGCCCCAGAAGACCAAGG	sequence:5598	+	55	VPR
AACAAGCCCCAGAAGACCAAGGG	sequence:5599	+	50	VPR
CTGTGGCCCTTGGTCTTCTGGGG	sequence:5605	-	60	VPR
TCTGTGGCCCTTGGTCTTCTGGG	sequence:5606	-	55	VPR
CTCTGTGGCCCTTGGTCTTCTGG	sequence:5607	-	60	VPR
AGAAGACCAAGGGCCACAGAGGG	sequence:5609	+	55	VPR
GTGGCTCCCTCTGTGGCCCTTGG	sequence:5615	-	70	VPR
AGAGGGAGCCACACAATGAATGG	sequence:5626	+	50	VPR
TCTAGTGTCCATTCATTGTGTGG	sequence:5634	-	40	VPR
ATGGACTAGAGCTTTTAGAGG	sequence:5645	+	40	VPR
TTTCCTAGGATTGGCTCCATGG	sequence:5694	+	45	VPR
AAGCCATGGAGCCAAATCCTAGG	sequence:5697	-	50	VPR
AGGATTTGGCTCCATGGCTTAGG	sequence:5700	+	50	VPR
GGATTTGGCTCCATGGCTTAGGG	sequence:5701	+	50	VPR
AGATATGTTGCCCTAAGCCATGG	sequence:5711	-	45	VPR
ACTTATGGGGATACTTGGGCAGG	sequence:5739	+	50	VPR
TGGGGATACTTGGGCAGGAGTGG	sequence:5744	+	60	VPR
GTCGACACCCAATTCTGAAATGG	sequence:5804	-	45	VPR
GGGTGTCGACATAGCAGAATAGG	sequence:5817	+	50	VPR
GAATAGGCGTACTCAACAGAGG	sequence:5833	+	45	VPR
CAACAGAGGAGAGCAAGAAATGG	sequence:5847	+	45	VPR
GGCTCTAGTCTAGGATCTACTGG	sequence:5872	-	50	5' TAT
TAGATCCTAGACTAGAGCCCTGG	sequence:5876	+	50	5' TAT
TGCTTCCAGGGCTCTAGTCTAGG	sequence:5881	-	55	5' TAT
CTAGAGCCCTGGAAGCATCCAGG	sequence:5887	+	60	5' TAT
TGACTTCTGGATGCTTCCAGGG	sequence:5893	-	50	5' TAT
CTGACTTCTGGATGCTTCCAGG	sequence:5894	-	55	5' TAT
GCAGTTTTAGGCTGACTTCTGG	sequence:5905	-	50	5' TAT
CAAGTGGTACAAGCAGTTTTAGG	sequence:5917	-	40	5' TAT
AAAGCCTTAGGCATCTCCTATGG	sequence:5986	+	45	5' TAT
CCTTAGGCATCTCCTATGGCAGG	sequence:5990	+	55	5' TAT
TCTCCTATGGCAGGAAGAAGCGG	sequence:5999	+	50	5' TAT
TCTCCGCTTCTTCTGCCATAGG	sequence:6002	-	55	5' TAT
TGAGTCTGACTGCCTTGAGGAGG	sequence:6037	-	55	5' TAT, 5' REV
TGATGAGTCTGACTGCCTTGAGG	sequence:6040	-	50	5' TAT, 5' REV
ATAGAAAGAGCAGAAGACAGTGG	sequence:6232	+	40	VPU
GACAGTGGCAATGAGAGTGAAGG	sequence:6247	+	50	VPU
AGGAGAAATATCAGCACTTGTGG	sequence:6267	+	40	VPU
AATATCAGCACTTGTGGAGATGG	sequence:6273	+	40	VPU
ATATCAGCACTTGTGGAGATGGG	sequence:6274	+	40	VPU
TATCAGCACTTGTGGAGATGGGG	sequence:6275	+	45	VPU
ATCAGCACTTGTGGAGATGGGGG	sequence:6276	+	50	VPU
AGCACTTGTGGAGATGGGGGTGG	sequence:6279	+	60	VPU
AAATGGGGCACCATGCTCCTTGG	sequence:6302	+	55	VPU
AATGGGGCACCATGCTCCTTGGG	sequence:6303	+	55	VPU
CATCAATATCCCAAGGAGCATGG	sequence:6312	-	45	VPU
GGGAGCAGCAGGAAGCACTATGG	sequence:7840	+	60	RRE
GGAGCAGCAGGAAGCACTATGGG	sequence:7841	+	55	RRE
CAGGAAGCACTATGGGCGCACGG	sequence:7848	+	60	RRE
CGCACGGTCAATGACGCTGACGG	sequence:7864	+	60	RRE
GTCAATGACGCTGACGGTACAGG	sequence:7870	+	55	RRE
CAGGCCAGACAATTATTGTCTGG	sequence:7889	+	45	RRE
GCAGCAGAACAATTTGCTGAGGG	sequence:7921	+	45	RRE
CAATTTGCTGAGGGCTATTGAGG	sequence:7930	+	45	RRE
ATCTGTTGCAACTCACAGTCTGG	sequence:7962	+	45	RRE
TCTGTTGCAACTCACAGTCTGGG	sequence:7963	+	45	RRE
CTGTTGCAACTCACAGTCTGGGG	sequence:7964	+	50	RRE
CTGGGGCATCAAGCAGCTCCAGG	sequence:7981	+	65	RRE
GCAGCTCCAGGCAAGAATCCTGG	sequence:7993	+	60	RRE
CCAGGCAAGAATCCTGGCTGTGG	sequence:7999	+	60	RRE
TTAGGTATCTTTCCACAGCCAGG	sequence:8011	-	45	RRE
GGCTGTGGAAAGATACCTAAAGG	sequence:8014	+	45	RRE
AGACCCACCTCCCAACCCCGAGG	sequence:8424	+	70	3' TAT, 3' REV
GACCCACCTCCCAACCCCGAGGG	sequence:8425	+	70	3' TAT, 3' REV
ACCCACCTCCCAACCCCGAGGGG	sequence:8426	+	70	3' TAT, 3' REV
TCCCCTCGGGGTTGGGAGGTGGG	sequence:8427	-	70	3' TAT, 3' REV
TGTCGGGTCCCCTCGGGGTTGGG	sequence:8434	-	70	3' TAT, 3' REV
TCTATTCTTCGGGCCTGTCGGG	sequence:8450	-	55	3' TAT, 3' REV

TTCTATTCCTTCGGGCCTGTTCGG	sequence:8451	-	50	3' TAT, 3' REV
CCCGAAGGAATAGAAGAAGAAGG	sequence:8459	+	45	3' TAT, 3' REV
GGATCCGTTCACTAATCGAATGG	sequence:8505	-	45	3' REV
ACGGATCCTTAGCACTTATCTGG	sequence:8520	+	45	3' REV
CGGATCCTTAGCACTTATCTGGG	sequence:8521	+	45	3' REV
ATCGTCCCAGATAAGTGCTAAGG	sequence:8526	-	45	3' REV
CACTTATCTGGGACGATCTGCGG	sequence:8532	+	50	3' REV
GGTGGTAGCTGAAGAGGCACAGG	sequence:8557	-	60	3' REV
TCAAGCGGTGGTAGCTGAAGAGG	sequence:8563	-	55	3' REV
AGAGTAAGTCTCTCAAGCGGTGG	sequence:8575	-	50	3' REV
TCAAGAGTAAGTCTCTCAAGCGG	sequence:8578	-	40	3' REV
CTTGATTGTAACGAGGATTGTGG	sequence:8596	+	40	3' REV
AACGAGGATTGTGGAACCTCTGG	sequence:8605	+	45	3' REV
ACGAGGATTGTGGAACCTCTGGG	sequence:8606	+	45	3' REV
TTGTGGAACCTCTGGGACGCAGG	sequence:8613	+	55	3' REV
TGTGGAACCTCTGGGACGCAGGG	sequence:8614	+	55	3' REV
GTGGAACCTCTGGGACGCAGGGG	sequence:8615	+	60	3' REV
TGGAACCTCTGGGACGCAGGGGG	sequence:8616	+	60	3' REV
ACTTCTGGGACGCAGGGGGTGGG	sequence:8620	+	65	3' REV
GGTGGGAAGCCCTCAAATATTGG	sequence:8637	+	50	3' REV
GGGAAGCCCTCAAATATTGGTGG	sequence:8640	+	50	3' REV
GGAGATTCCACCAATATTTGAGG	sequence:8647	-	40	3' REV
TCTCCTACAGTATTGGAGTCAGG	sequence:8665	+	45	3' REV
GTTCTGACTCCAATACTGTAGG	sequence:8668	-	45	3' REV
CAAGTGGTCAAAAAGTAGTGTGG	sequence:8852	+	40	NEF
TGGTCAAAAAGTAGTGTGGTTGG	sequence:8856	+	40	NEF
CAAAAAGTAGTGTGGTTGGATGG	sequence:8860	+	40	NEF
TGGTTGGATGGCCTACTGTAAGG	sequence:8872	+	50	NEF
GGTTGGATGGCCTACTGTAAGGG	sequence:8873	+	50	NEF
CTCATTCTTTCCCTTACAGTAGG	sequence:8883	-	40	NEF
CGAGCTGAGCCAGCAGCAGATGG	sequence:8907	+	65	NEF
GAGCTGAGCCAGCAGCAGATGGG	sequence:8908	+	60	NEF
AGCTGAGCCAGCAGCAGATGGGG	sequence:8909	+	60	NEF
TGAGCCAGCAGCAGATGGGGTGG	sequence:8912	+	65	NEF
GCTCCCACCCATCTGCTGCTGG	sequence:8916	-	70	NEF
GGGAGCAGCATCTCGAGACCTGG	sequence:8933	+	65	NEF
TCTCGAGACCTGGAAAAACATGG	sequence:8943	+	45	NEF
TGATTGCTCCATGTTTTCCAGG	sequence:8951	-	40	NEF
CTACCAATGCTGCTTGTGCCTGG	sequence:8992	+	55	NEF
TAGCCAGGCACAAGCAGCATTGG	sequence:8995	-	55	NEF
TGCCTGGCTAGAAGCACAAGAGG	sequence:9008	+	55	NEF
CTCCTCTTGTGCTTCTAGCCAGG	sequence:9010	-	55	NEF
CTGGCTAGAAGCACAAGAGGAGG	sequence:9011	+	55	NEF
GCTAGAAGCACAAGAGGAGGAGG	sequence:9014	+	55	NEF
GGGTTTTCCAGTCACACCTCAGG	sequence:9041	+	55	NEF
AAAGGTACCTGAGGTGTGACTGG	sequence:9048	-	50	NEF
ATTGGTCTTAAAGGTACCTGAGG	sequence:9057	-	40	NEF
AAAGAAAAGGGGGGACTGGAAGG	sequence:9117	+	50	NEF
AAGAAAAGGGGGGACTGGAAGGG	sequence:9118	+	50	NEF

LAI Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
TACCACACACAAGGCTACTTCCCTGATT	sequence:54, 9186	+	45	LTR U3 NEF
CCTTATCTGGCTCAACTGGTACTAGCTT	sequence:141, 9273	-	50	LTR U3 NEF
AGGCCAATAAAGGAGAGAACACCAGCTT	sequence:175, 9307	+	45	LTR U3 NEF
TACTTCAAGAACTGCTGACATCGAGCTT	sequence:315, 9447	+	40	LTR U3 NEF
CTCAGATGCTGCATATAAGCAGCTGCTT	sequence:414, 9546	+	45	LTR U3
ACCCAGTACAGGCAAAAAGCAGCTGCTT	sequence:430, 9562	-	50	LTR U3
TACTGGTCTCTCTGGTTAGACCAGATT	sequence:450, 9582	+	50	LTR U3
CTCTGGCTAACTAGGGAACCCACTGCTT	sequence:492, 9624	+	55	LTR R
ACCCACTGCTTAAGCCTCAATAAAGCTT	sequence:509, 9641	+	50	LTR R
GCACTCAAGGCAAGCTTTATTGAGGCTT	sequence:520, 9652	-	45	LTR R
GCCTCAATAAAGCTTGCTTGAGTGCTT	sequence:522, 9654	+	45	LTR R
CACTACTTGAAGCACTCAAGGCAAGCTT	sequence:531, 9663	-	45	LTR R
GAGAGCTCCTCTGGTTTCCCTTTCGCTT	sequence:658	-	60	
GGAGCTCTCTCGACGCAGGACTCGGCTT	sequence:677	+	65	
CTCCCCTCGCCTCTTGCCGTGCGCGCTT	sequence:709	-	70	DIS
TTCCCATCGATCTAATTCTCCCCGCTT	sequence:813	-	40	GAG
TATGGGCAAGCAGGGAGCTAGAACGATT	sequence:893	+	55	GAG
AACTGCGAATCGTTCTAGCTCCCTGCTT	sequence:900	-	45	GAG
CTTCTGATGTTTCTAACAGGCCAGGATT	sequence:928	-	40	GAG
TGCTCTTCTCTATCTTGTCTAAAGCTT	sequence:1085	-	45	GAG
CCTGTGTCAGCTGCTGCTTGCTGTGCTT	sequence:1130	-	60	GAG
CTGCTGTGTCCTGTGTCAGCTGCTGCTT	sequence:1139	-	60	GAG
TTATCAGAAGGAGCCACCCACAAGATT	sequence:1312	+	55	GAG
GGATGCACTCTATCCCATTCTGCAGCTT	sequence:1412	-	50	GAG
TTGTTGGTCCAAAATGCGAACCCAGATT	sequence:1750	+	40	GAG
TCTAGTGTAGCTGCTGGTCCAATGCTT	sequence:1793	-	55	GAG
GCTGAATTTGTTACTTGGCTCATTGCTT	sequence:1883	-	40	GAG
GGAAAGGAAGGACACCAAATGAAAGATT	sequence:2038	+	45	GAG
AGCCAACAGCCCACCAGAAGAGAGCTT	sequence:2186	+	60	GAG POL
ACAGTTCCTTGTCTATCGGCTCCTGCTT	sequence:2248	-	50	GAG POL
TCATCTGCTCCTGTATCTAATAGAGCTT	sequence:2350	-	40	POL
GGTCCTACTAATACTGTACCTATAGCTT	sequence:2497	-	45	POL
AAGTGCAACCAATCTGAGTCAACAGATT	sequence:2550	-	45	POL
TAACTTTTGGGCCATCCATTCCTGGCTT	sequence:2622	-	40	POL
TGGCAAATACTGGAGTATTGTATGGATT	sequence:2745	-	40	POL
TCACCCACATCCAGTACTGTTACTGATT	sequence:2896	-	50	POL
GGGAAAATTGAATTGGGCAAGTCAGATT	sequence:3368	+	40	POL
TTGGGCAAGTCAGATTTACCCAGGGATT	sequence:3380	+	45	POL
CTGTTTTCTGCCAGTTCTAGCTCTGCTT	sequence:3478	-	45	POL
GCTAGAATGGCAGAAAACAGAGAGATT	sequence:3485	+	45	POL
GATATGTCCATTGGCCTTGCCCCTGCTT	sequence:3576	-	50	POL
AGGAGTCTTCCCATATTACTATGCTT	sequence:3719	-	40	POL
GACAGAGTATTGGCAAGCCACCTGGATT	sequence:3791	+	50	POL
AACTCCCACTCAGGAATCCAGGTGGCTT	sequence:3805	-	50	POL
GAATTGCTTGAACTCAGTCTTCTGATT	sequence:4005	-	40	POL
GAATCCTGCAAAGCTAGATGAATTGCTT	sequence:4024	-	45	POL
TTCATCTAGCTTTCAGGATTCGGGATT	sequence:4030	+	40	POL
AGTAATTGGAGAGCAATGGCTAGTGATT	sequence:4314	+	45	POL
CTACAGTCTACTTGTCCATGCATGGCTT	sequence:4408	-	45	POL
TGTAGTCCAGGAATATGGCAACTAGATT	sequence:4431	+	45	POL
CCTGTTTCTGCTGGAATAACTTCTGCTT	sequence:4519	-	45	POL
GTAGGGAATCCAAATTCCTGCTTGATT	sequence:4667	-	40	POL
CTGCCCTTACCTTCCAGAGGAGCTT	sequence:4983	-	60	POL
TATGGAAAACAGATGGCAGGTGATGATT	sequence:5076	+	45	POL
TGTGTGGCAAGTAGACAGGATGAGGATT	sequence:5103	+	50	POL
TACACATCCCCTAGGGGATGCTAGATT	sequence:5240	+	50	VIF
AGAGATCCTACCTTGTTATGTCCTGCTT	sequence:5483	-	40	VIF
GTTTCGTAACACTAGGCAAAGGTGGCTT	sequence:5554	-	40	VIF
CTCTGTGGCCCTTGGTCTTCTGGGGCTT	sequence:5602	-	60	VPR
GCCACACAATGAATGGACTAGAGCTT	sequence:5633	+	50	VPR
ATGGACTAGAGCTTTAGAGGAGCTT	sequence:5645	+	40	VPR
ATTTTCTAGGATTTGGCTCCATGGCTT	sequence:5692	+	40	VPR
CAGTTTTAGGCTGACTTCTGGATGCTT	sequence:5899	-	45	5' TAT
CATCCAGGAAGTCAGCCTAAAAGCTT	sequence:5902	+	50	5' TAT
TTCCTGCCATAGGAGATGCCTAAGGCTT	sequence:5987	-	55	5' TAT
GAGGAGGTCTTCGTCGCTGTCTCCGCTT	sequence:6016	-	60	5' TAT 5' REV
CGTCATTGACCGTGCGCCATAGTGCTT	sequence:7852	-	65	RRE
CAGCCAGGATTCCTGCCTGGAGCTGCTT	sequence:7991	-	60	RRE
CTTTAGGTATCTTCCACAGCCAGGATT	sequence:8008	-	40	RRE
AGAGAGACAGAGACAGATCCATTCGATT	sequence:8487	+	50	3' REV
GCCTGTGCCTCTCAGCTACCACCGCTT	sequence:8556	+	60	3' REV
CCACCGCTTGAGAGACTTACTCTTGATT	sequence:8575	+	55	3' REV
GTTCTGACTCCAATACTGTAGGAGATT	sequence:8663	-	45	3' REV

NM

GGTAGCTGCTGTATTGCTACTTGTGATT	sequence:8969	-	50	NEF
AGCAATACAGCAGCTACCAATGCTGCTT	sequence:8979	+	45	NEF
AAACCCACCTCCTCCTCCTTGTGCTT	sequence:9019	-	60	NEF

LAI Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
GAGCGTCAGTATTAAGCGGGGAGAAT	sequence:800	+	55	GAG
ATTCGGTTAAGGCCAGGGGAAAGAAA	sequence:844	+	60	GAG
TCGCAGTTAATCCTGGCCTGTTAGAAA	sequence:920	+	55	GAG
GAGGAAGAGCAAAACAAAAGTAAGAAA	sequence:1102	+	40	GAG
AGACCATCAATGAGGAAGCTGCAGAAT	sequence:1397	+	45	GAG
ATAATCCACCTATCCCAGTAGGAGAAA	sequence:1544	+	40	GAG
TGGGACCAGCAGCTACACTAGAAGAAA	sequence:1799	+	55	GAG
AAGAGGCAATTTTAGGAACCAAAGAAA	sequence:1926	+	40	GAG
TGGCAAAGAAGGGCACATAGCCAGAAA	sequence:1974	+	50	GAG
TGTTGGCTCTGGTCTGCTCTGAAGAAA	sequence:2131	-	55	GAG POL
TGTTGGCTCTGGTCTGCTCTGAAGAAA	sequence:2167	-	55	GAG POL
GACAGTATGATCAGATACTCATAGAAA	sequence:2458	+	40	POL
CCACATCCCGCAGGGTTAAAAAAGAAA	sequence:2868	+	55	POL
ACCACACCAGACAAAAAACATCAGAAA	sequence:3228	+	40	POL
CAGTACAGCCTATAGTGCTGCCAGAAA	sequence:3304	+	50	POL
AAGAAGCAGAGCTAGAAGTGGCAGAAA	sequence:3475	+	45	POL
CACTCCATGTACTGGTTCTTTTAGAAT	sequence:3510	-	45	POL
AGGCAGTGCAAAAAATAACCACAGAAA	sequence:3694	+	40	POL
AGAAAGAACCCATAGTAGGAGCAGAAA	sequence:3874	+	40	POL
GATAAGGCCCAAGATGAACATGAGAAA	sequence:4281	+	45	POL
ACCTGCCACCTGTAGTAGCAAAAAGAAA	sequence:4345	+	55	POL
ATCCCCAAAGTCAAGGAGTAGTAGAAT	sequence:4696	+	45	POL
TGGGGGGTACAGTGCAGGGGAAAGAAT	sequence:4838	+	70	POL cPPT
TCACTATGAAAGCCCTCATCCAAGAAT	sequence:5202	+	45	VIF
CATATTGGGGTCTGCATACAGGAGAAA	sequence:5279	+	45	VIF
ATCTGGGTGAGGAGTCTCCATAGAAT	sequence:5315	+	60	VIF
CTGTTTTTCAGACTCTGCTATAAGAAA	sequence:5415	+	40	VIF
CTAGAGCTTTTAGAGGAGCTTAAGAAT	sequence:5652	+	45	VPR
GGCAGGAGTGAAGCCATAATAAGAAT	sequence:5756	+	55	VPR
CAGAATTGGGTGTCGACATAGCAGAAT	sequence:5810	+	45	VPR
CGTTACTCAACAGAGGAGCAAGAAA	sequence:5840	+	50	VPR
TTGTGTGGTCCATAGTAATCATAGAAT	sequence:6155	+	40	VPU
ACAGTGGCAATGAGAGTGAAGGAGAAA	sequence:6248	+	45	VPU
GGGCATCAAGCAGCTCCAGGCAAGAAT	sequence:7984	+	65	RRE
CAGTATTGGAGTCAGGAATAAGAAT	sequence:8672	+	45	3' REV
TGGATGGCCTACTGTAAGGGAAAGAAT	sequence:8876	+	55	NEF

LAI Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
ATATCTTGTCTTCGTTGGGAGTGAAT	sequence:12, 9144	-	40	LTR U3 NEF
GTGTGGTAGATCCACAGATCAAGGAT	sequence:36, 9168	-	50	LTR U3 NEF
CAGAACTACACACCAGGGCCAGGGGT	sequence:84, 9216	+	60	LTR U3 NEF
GGTCAGATATCCACTGACCTTTGGAT	sequence:107, 9239	+	50	LTR U3 NEF
GTAGCACCATCAAAGGTCAGTGGAT	sequence:115, 9247	-	50	LTR U3 NEF
TTACACCCTGTGAGCCTGCATGGAAT	sequence:204, 9336	+	55	LTR U3 NEF
ACCCTGTGAGCCTGCATGGAATGGAT	sequence:208, 9340	+	60	LTR U3 NEF
TGACCCTGAGAGAGAAGTGTAGAGT	sequence:233, 9365	+	50	LTR U3 NEF
CCACTTAACACTTCTCTCAGGGT	sequence:235, 9367	-	45	LTR U3 NEF
CGTGGCCCGAGAGCTGCATCCGGAGT	sequence:290, 9422	+	70	LTR U3 NEF
TCAGCAGTTCTTGAAGTACTCCGGAT	sequence:307, 9439	-	40	LTR U3 NEF
GCAGCTGCTTTTTGCCTGTACTGGGT	sequence:432, 9564	+	50	LTR U3
GCTTTATTGAGGCTTAAGCAGTGGGT	sequence:509, 9641	-	40	LTR R
AAGCCTCAATAAAGCTTGCTTGAGT	sequence:520, 9652	+	45	LTR R
TCTGAGGGATCTCTAGTTACCAGAGT	sequence:578, 9710	-	45	LTR U5
ACACTGACTAAAAGGGTCTGAGGGAT	sequence:594, 9726	-	45	LTR U5
GCCGTGCGCGCTTCAGCAAGCCGAGT	sequence:696	-	70	DIS
CGGAGGCTAGAAGGAGAGAGATGGGT	sequence:770	+	60	Psi
AGCGTCAGTATTAAGCGGGGAGAAT	sequence:801	+	55	GAG
TTCTTTCCCTGCGCTTAACCGAAT	sequence:844	-	50	GAG
CTAACAGGCCAGGATTAAGTGCAGAT	sequence:918	-	45	GAG
TTCTGATGTTTCTAACAGGCCAGGAT	sequence:929	-	40	GAG
GCTACAACCATCCCTTCAGACAGGAT	sequence:978	+	50	GAG
GTTCTTCTGATCCTGTCTGAAGGGAT	sequence:987	-	45	GAG
ACCCTCTATTGTGTGCATCAAAGGAT	sequence:1039	+	45	GAG
CCCCTGGATGTTCTGCACTATAGGGT	sequence:1184	-	55	GAG
GCCTGATGTACCATTTGCCCTGGAT	sequence:1201	-	55	GAG
GACCATCAATGAGGAAGCTGCAGAAT	sequence:1398	+	50	GAG
TCAATGAGGAAGCTGCAGAATGGGAT	sequence:1403	+	45	GAG
GAGGAAGCTGCAGAATGGGATAGAGT	sequence:1408	+	55	GAG
TGCAATAGGCCCTGCATGCACTGGAT	sequence:1436	-	55	GAG
TAGTACCCTTCAGGAACAAATAGGAT	sequence:1509	+	40	GAG
ACCCTTCAGGAACAAATAGGATGGAT	sequence:1513	+	45	GAG
AATTTCTCCTACTGGGATAGGTGGAT	sequence:1547	-	40	GAG
TTTGGTCTTGTCTTATGTCCAGAAT	sequence:1633	-	40	GAG
ACCGGTCTACATAGTCTCTAAAGGGT	sequence:1662	-	45	GAG
TGTGAAGCTTGCTCGGCTCTTAGAGT	sequence:1696	-	55	GAG
GCTTCACAGGAGGTAATAAATTGGAT	sequence:1714	+	40	GAG
GAAATGATGACAGCATGTCAGGGAGT	sequence:1822	+	40	GAG
GGAGGACCCGGCCATAAGGCAAGAGT	sequence:1849	+	70	GAG
CCAAAACCTCTTGCCTTATGGCCGGGT	sequence:1854	-	45	GAG
CTTCCTACAAGGGAAGGCCAGGGAAT	sequence:2105	+	55	GAG POL
CCAGAAGAGAGCTTCAGGTCTGGGGT	sequence:2200	+	55	GAG POL
ATCGGCTCCTGCTTCTGAGAGGGAGT	sequence:2236	-	55	GAG POL
GAGTGATCTGAGGGAAGTTAAAGGAT	sequence:2276	-	45	GAG POL
GTGACGAGGGGTCGTTGCCAAAGAGT	sequence:2298	-	65	GAG POL
CCCCTATCTTTATTGTGACGAGGGGT	sequence:2312	-	45	GAG POL
GGCAAATACTGGAGTATTGTATGGAT	sequence:2746	-	40	POL
AATTGAACTTCCAGAAGTCTTGAGT	sequence:2835	-	40	POL
GACTTCTGGGAAGTTCAATTAGGAAT	sequence:2841	+	40	POL
ATTAGGAATACCACATCCCGCAGGGT	sequence:2858	+	45	POL
TCAGTACAATGTGCTTCCACAGGGAT	sequence:3017	+	45	POL
TGTGCTTCCACAGGGATGGAAAGGAT	sequence:3026	+	55	POL
CAGAAAGAACCTCCATTCCTTTGGAT	sequence:3249	+	45	POL
AAGAACCTCCATTCCTTTGGATGGGT	sequence:3253	+	45	POL
GGAGTTCATAACCCATCCAAAGGAAT	sequence:3263	-	45	POL
GTACTGTCCATTTATCAGGATGGAGT	sequence:3284	-	40	POL
AGGCTGTACTGTCCATTTATCAGGAT	sequence:3289	-	40	POL
TGGGCAAGTCAGATTTACCCAGGGAT	sequence:3381	+	50	POL
TTAGTGCTTTGGTTCCTCTAAGGAGT	sequence:3428	-	40	POL
CAGGAAAATATGCAAGAACGAGGGGT	sequence:3637	+	40	POL
AACATGGGAAACATGGTGGACAGAGT	sequence:3773	+	45	POL
ACAGAGTATTGGCAAGCCACCTGGAT	sequence:3792	+	50	POL
TTGGCAAGCCACCTGGATTCTGAGT	sequence:3800	+	55	POL
AGCCACCTGGATTCTGAGTGGGAGT	sequence:3806	+	55	POL
GTATTGACAACTCCCACTCAGGAAT	sequence:3816	-	45	POL
AGAACGTTTCTGCTCCTACTATGGGT	sequence:3881	-	45	POL
CCCGAATCCTGCAAAGCTAGATGAAT	sequence:4029	-	55	POL
TCATCTAGCTTTGCAGGATTCGGGAT	sequence:4031	+	40	POL
TCACTTTTATCTGGTTGTGCTTGAAT	sequence:4101	-	40	POL
AAGGAAAAGGTCTATCTGGCATGGGT	sequence:4167	+	45	POL
GCATGGGTACCAGCACACAAAGGAAT	sequence:4185	+	55	POL
GGACAAGTAGACTGTAGTCCAGGAAT	sequence:4419	+	50	POL

CTGGCTACATGAACTGCTACCAGGAT	sequence:4482	-	50	POL
GGTAGCAGTTCATGTAGCCAGTGGAT	sequence:4487	+	50	POL
GTTTCCTGCCCTGTTTCTGCTGGAAT	sequence:4530	-	55	POL
AAGGCCGCCTGTTGGTGGGCGGGAAT	sequence:4644	+	70	POL
TTGGTGGGCGGGAATCAAGCAGGAAT	sequence:4655	+	60	POL
GCGGGAATCAAGCAGGAATTTGGAAT	sequence:4662	+	50	POL
CCTTGACTTTGGGGATTGTAGGGAAT	sequence:4686	-	45	POL
CCCTACAATCCCCAAAGTCAAGGAGT	sequence:4689	+	50	POL
TCCCCAAAGTCAAGGAGTAGTAGAAT	sequence:4697	+	50	POL
GTCAAGGAGTAGTAGAATCTATGAAT	sequence:4705	+	40	POL
TAAAAGAAAAGGGGGATTGGGGGGT	sequence:4820	+	40	POL cPPT
GGGGGGTACAGTGCAGGGGAAAGAAT	sequence:4839	+	70	POL cPPT
CTTTGCTGGTCTTTCCAAAGTGGAT	sequence:4960	-	45	POL
ATGATTGTGTGGCAAGTAGACAGGAT	sequence:5098	+	40	POL
GTGTGGCAAGTAGACAGGATGAGGAT	sequence:5104	+	50	POL
GTATGTTTCAGGGAAAGCTAGGGGAT	sequence:5163	+	40	VIF
CACTATGAAAGCCCTCATCCAAGAAT	sequence:5203	+	50	VIF
CAGAAGTACACATCCCACTAGGGGAT	sequence:5234	+	45	VIF
ACCAATCTAGCATCCCCTAGTGGGAT	sequence:5245	-	50	VIF
CAGGAGAAAGAGACTGGCATCTGGGT	sequence:5297	+	50	VIF
AGAGACTGGCATCTGGGTCAGGGAGT	sequence:5305	+	55	VIF
TCTGGGTCAGGGAGTCTCCATAGAAT	sequence:5316	+	60	VIF
ACATATAGTTAGCCCTAGGTGTGAAT	sequence:5454	+	40	VIF
TCAAGCAGGACATAACAAGGTAGGAT	sequence:5481	+	45	VIF
CTAGTGTTACGAAACTGACAGAGGAT	sequence:5567	+	40	VIF
GCCACAGAGGGAGCCACACAATGAAT	sequence:5621	+	65	VPR
TAGAGCTTTTAGAGGAGCTTAAGAAT	sequence:5653	+	40	VPR
GAAGCTGTTAGACATTTTCCTAGGAT	sequence:5679	+	40	VPR
ACTTATGGGGATACTTGGGCAGGAGT	sequence:5739	+	50	VPR
GCAGGAGTGGAAAGCCATAATAAGAAT	sequence:5757	+	50	VPR
TGGATAAACAGCAGTTGTTGCAGAAT	sequence:5781	-	40	VPR
TGTCGACACCCAATTCTGAAATGGAT	sequence:5802	-	45	VPR
AGAATTGGGTGTCGACATAGCAGAAT	sequence:5811	+	45	VPR
CCATTTCTGCTCCTCTGTTGAGT	sequence:5844	-	50	VPR
ATGCTTCCAGGGCTCTAGTCTAGGAT	sequence:5879	-	55	5' TAT
AGCAGTTTTAGGCTGACTTCCTGGAT	sequence:5903	-	45	5' TAT
TGTGTGGTCCATAGTAATCATAGAAT	sequence:6156	+	40	VPU
GAGCAGAAGACAGTGGCAATGAGAGT	sequence:6239	+	50	VPU
ATATCAGCACTTGTGGAGATGGGGGT	sequence:6274	+	40	VPU
AAATGGGGCACCATGCTCCTTGGGAT	sequence:6302	+	55	VPU
GCTGCTTATGCCCCAGACTGTGAGT	sequence:7972	-	60	RRE
GGCATCAAGCAGCTCCAGGCAAGAAT	sequence:7985	+	65	RRE
TTTAGGTATCTTTCCACAGCCAGGAT	sequence:8009	-	40	RRE
TGGCTGTGAAAAGATACCTAAAGGAT	sequence:8013	+	45	RRE
CTTCTATTCCTTCGGGCCTGTCGGGT	sequence:8449	-	55	3' TAT 3' REV
CAGATCCATTCGATTAGTGAACGGAT	sequence:8500	+	40	3' REV
AGGATCCGTTCACTAATCGAATGGAT	sequence:8503	-	45	3' REV
GCTAAGGATCCGTTCACTAATCGAAT	sequence:8507	-	45	3' REV
GATCGTCCCAGATAAGTGCTAAGGAT	sequence:8524	-	50	3' REV
CCACAATCCTCGTTACAATCAAGAGT	sequence:8593	-	45	3' REV
TGTGGAATTCTGGGACGCAGGGGGT	sequence:8614	+	55	3' REV
GGGAAGCCCTCAAATATTGGTGGAAAT	sequence:8640	+	50	3' REV
GGTGGAATCTCCTACAGTATTGGAGT	sequence:8658	+	45	3' REV
AGTATTGGAGTCAGGAACTAAAGAAT	sequence:8673	+	40	3' REV
GTGGTCAAAAAGTAGTGTGGTTGGAT	sequence:8855	+	45	NEF
GGATGGCCTACTGTAAGGGAAAGAAT	sequence:8877	+	55	NEF
CGAGCTGAGCCAGCAGCAGATGGGGT	sequence:8907	+	65	NEF

LAI Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
TTTGGATGGTGCTACAAGCTAGTACCAG	sequence:126, 9258	+	50	LTR U3 NEF
TTTATTGGCCTCTTCTACCTTATCTGGC	sequence:158, 9290	-	50	LTR U3 NEF
TTTGACAGCCGCTAGCATTTCATCACG	sequence:264, 9396	+	54	LTR U3 NEF
TTTCATCACGTGGCCCGAGAGCTGCATC	sequence:282, 9414	+	62	LTR U3 NEF
TTTTTGCTGTACTGGGTCTCTCTGGTT	sequence:440, 9572	+	54	LTR U3
TTTTGCCTGTACTGGGTCTCTCTGGTTA	sequence:441, 9573	+	54	LTR U3
TTTGCCTGTACTGGGTCTCTCTGGTTAG	sequence:442, 9574	+	54	LTR U3
TTTGAGCCTGGGAGCTCTCTGGCTAACT	sequence:476, 9608	+	58	LTR R
TTTATTGAGGCTTAAGCAGTGGGTTCCC	sequence:505, 9637	-	54	LTR R
TTTCCACACTGACTAAAAGGGTCTGAGG	sequence:597, 9729	-	50	LTR U5
TTTTCCACACTGACTAAAAGGGTCTGAG	sequence:598, 9730	-	50	LTR U5
TTTTAGTCAGTGTGGAAAATCTCTAGCA	sequence:607, 9739	+	42	LTR U5
TTTAGTCAGTGTGGAAAATCTCTAGCAG	sequence:608	+	46	LTR U5
TTTCAAGTCCCTGTTCCGGGCGCCACTGC	sequence:632	-	67	LTR U5
TTTCGCTTTCAAGTCCCTGTTCCGGGCGC	sequence:638	-	62	PBS
TTTCCCTTTTCGCTTTCAAGTCCCTGTTCC	sequence:644	-	50	PBS
TTTGGCGTACTCACCAGTCGCCTCCCCT	sequence:730	-	67	DIS
TTTTTGGCGTACTCACCAGTCGCCTCCC	sequence:732	-	67	DIS
TTTTGACTAGCGGAGGCTAGAAGGAGAG	sequence:760	+	58	Psi
TTTGACTAGCGGAGGCTAGAAGGAGAGA	sequence:761	+	54	Psi
TTTCCCATCGATCTAATTCTCCCCGCT	sequence:814	-	54	GAG
TTTTCCCATCGATCTAATTCTCCCCGCG	sequence:815	-	58	GAG
TTTTTCCCATCGATCTAATTCTCCCCG	sequence:816	-	54	GAG
TTTTTTCCCATCGATCTAATTCTCCCC	sequence:817	-	50	GAG
TTTCCCCTGGCCTTAACCGAATTTTTT	sequence:839	-	46	GAG
TTTCTTTCCCCTGGCCTTAACCGAATT	sequence:843	-	50	GAG
TTTTCTTTCCCCTGGCCTTAACCGAAT	sequence:844	-	54	GAG
TTTTTCTTTCCCCTGGCCTTAACCGAA	sequence:845	-	54	GAG
TTTTTTCTTTCCCCTGGCCTTAACCGA	sequence:846	-	54	GAG
TTTATATTTTTCTTTCCCCTGGCCTT	sequence:852	-	42	GAG
TTTCTAACAGGCCAGGATTAACCTGCGAA	sequence:919	-	46	GAG
TTTGATGCACACAATAGAGGGTTGCTAC	sequence:1033	-	46	GAG
TTTGCTCTTCTCTATCTTGTCTAAAGC	sequence:1087	-	42	GAG
TTTTGCTCTTCTCTATCTTGTCTAAAG	sequence:1088	-	42	GAG
TTTAGACAAGATAGAGGAAGAGCAAAAC	sequence:1089	+	42	GAG
TTTTGTTTTGCTCTTCTCTATCTTGTCT	sequence:1093	-	42	GAG
TTTGGCTGACCTGGCTGCTGTGCTCCTGT	sequence:1153	-	62	GAG
TTTTGGCTGACCTGGCTGCTGTGCTCCTG	sequence:1154	-	67	GAG
TTTGCCCTGGATGTTCTGCACTATAGG	sequence:1186	-	54	GAG
TTTAAAGTCTAGGTGATATGGCCTGAT	sequence:1220	-	42	GAG
TTTACCCATGCATTTAAAGTTCTAGGTG	sequence:1232	-	42	GAG
TTTCAGCCAGAAGTGATACCCATGTTT	sequence:1278	+	46	GAG
TTTTCAGCATTATCAGAAGGAGCCACCC	sequence:1303	+	54	GAG
TTTCAGCATTATCAGAAGGAGCCACCCC	sequence:1304	+	54	GAG
TTTAAATCTTGTGGGGTGGCTCCTTCTG	sequence:1316	-	54	GAG
TTTAGCATGGTGTTTAAATCTTGTGGGG	sequence:1328	-	46	GAG
TTTAAACACCATGCTAAACACAGTGGGG	sequence:1338	+	50	GAG
TTTGCATGGCTGCTTGATGTCCCCCAC	sequence:1360	-	62	GAG
TTTAAACATTTGCATGGCTGCTTGATGTC	sequence:1367	-	46	GAG
TTTTAACATTTGCATGGCTGCTTGATGT	sequence:1368	-	42	GAG
TTTGTTCTGAAGGGTACTAGTAGTTCC	sequence:1501	-	46	GAG
TTTCTCCTACTGGGATAGGTGGATTATT	sequence:1543	-	42	GAG
TTTGGTCTTGTCTTATGTCCAGAATGC	sequence:1631	-	46	GAG
TTTTGGTCTTGTCTTATGTCCAGAATG	sequence:1632	-	46	GAG
TTTAGAGACTATGTAGACCGGTTCTATA	sequence:1666	+	42	GAG
TTTATAGAACCGGTCTACATAGTCTCTA	sequence:1668	-	42	GAG
TTTTATAGAACCGGTCTACATAGTCTCT	sequence:1669	-	42	GAG
TTTACCTCCTGTGAAGCTTGCTCGGCTC	sequence:1703	-	62	GAG
TTTTACCTCCTGTGAAGCTTGCTCGGCT	sequence:1704	-	58	GAG
TTTTTACCTCCTGTGAAGCTTGCTCGGC	sequence:1705	-	58	GAG
TTTTTTACCTCCTGTGAAGCTTGCTCGG	sequence:1706	-	54	GAG
TTTGGACCAACAAGGTTTCTGTCATCCA	sequence:1735	-	46	GAG
TTTTGGACCAACAAGGTTTCTGTCATCC	sequence:1736	-	50	GAG
TTTTAAAAGCATTGGGACCAGCAGCTAC	sequence:1787	+	50	GAG
TTTAAAAGCATTGGGACCAGCAGCTACA	sequence:1788	+	50	GAG
TTTCTTCTAGTGTAGCTGCTGGTCCCAA	sequence:1798	-	50	GAG
TTTTGGCTGAAGCAATGAGCCAAGTAAC	sequence:1874	+	50	GAG
TTTGGCTGAAGCAATGAGCCAAGTAACA	sequence:1875	+	46	GAG
TTTGTTACTTGGCTCATTGCTTCAGCCA	sequence:1877	-	46	GAG
TTTCAATTGTGGCAAAGAAGGGCACATA	sequence:1965	+	42	GAG

Cpf1

TTTCTGGCTATGTGCCCTTCTTTGCCAC	sequence:1973	-	54	GAG
TTTCCTAGGGGCCCTGCAATTTCTGGCT	sequence:1992	-	58	GAG
TTTTCTAGGGGCCCTGCAATTTCTGGC	sequence:1993	-	62	GAG
TTTTTCTAGGGGCCCTGCAATTTCTGG	sequence:1994	-	58	GAG
TTTCCAACAGCCCTTTTTCTAGGGGCC	sequence:2007	-	58	GAG
TTTCCACATTTCCAACAGCCCTTTTTCC	sequence:2015	-	46	GAG
TTTGGTGTCTTCTTTCCACATTTCCA	sequence:2029	-	46	GAG
TTTTTTAGGGAAGATCTGGCCTTCTAC	sequence:2085	+	50	GAG POL
TTTTTAGGGAAGATCTGGCCTTCTACA	sequence:2086	+	50	GAG POL
TTTTAGGGAAGATCTGGCCTTCTACAA	sequence:2087	+	50	GAG POL
TTTAGGGAAGATCTGGCCTTCTACAAG	sequence:2088	+	54	GAG POL
TTTTCTTCAGAGCAGACCAGAGCCAACA	sequence:2130	+	54	GAG POL
TTTCTTCAGAGCAGACCAGAGCCAACAG	sequence:2131	+	54	GAG POL
TTTCTTCAGAGCAGACCAGAGCCAACAG	sequence:2167	+	54	GAG POL
TTAACTTCCCTCAGATCACTCTTTGGC	sequence:2280	+	50	GAG POL
TTTGCCAACGACCCCTCGTCACAATAAA	sequence:2302	+	50	GAG POL
TTTAGTTGCCCCCTATCTTTATTGTGA	sequence:2320	-	46	GAG POL
TTTCCATCTTCTGGCAAACCTATTCT	sequence:2390	-	42	POL
TTTGGTTTCCATCTTCTGGCAAACCTCA	sequence:2395	-	46	POL
TTTGGTTTCCATCTTCTGGCAAACCTC	sequence:2396	-	50	POL
TTTTTGGTTTCCATCTTCTGGCAAACCT	sequence:2397	-	46	POL
TTTGCCAGGAAGATGGAAACCAAAAATG	sequence:2399	+	42	POL
TTTGATAAAACCTCCAATTTCCCTATC	sequence:2426	-	42	POL
TTTAAAGTGCAACCAATCTGAGTCAACA	sequence:2554	-	42	POL
TTTGGGCCATCCATTCCTGGCTTTAATT	sequence:2617	-	46	POL
TTTTGGGCCATCCATTCCTGGCTTTAAT	sequence:2618	-	50	POL
TTTAACTTTTGGGCCATCCATTCCTGGC	sequence:2624	-	54	POL
TTTTCCCTTCTTTTCCATTTCTGTACA	sequence:2697	-	42	POL
TTTAAACCCTGCGGGATGTGGTATTCTTA	sequence:2860	-	54	POL
TTTTAACCCTGCGGGATGTGGTATTCTT	sequence:2861	-	54	POL
TTTTTAACCCTGCGGGATGTGGTATTCC	sequence:2862	-	54	POL
TTTTTTAACCCTGCGGGATGTGGTATTC	sequence:2863	-	50	POL
TTTCTTTTTTAACCCTGCGGGATGTGGT	sequence:2867	-	46	POL
TTTTCTTTTTTAACCCTGCGGGATGTGG	sequence:2868	-	50	POL
TTTTTCTTTTTTAACCCTGCGGGATGTG	sequence:2869	-	46	POL
TTTTTCAGTTCCCTTAGATGAAGACTTC	sequence:2930	+	42	POL
TTTTCAGTTCCCTTAGATGAAGACTTCA	sequence:2931	+	42	POL
TTTCAGTTCCCTTAGATGAAGACTTCAG	sequence:2932	+	42	POL
TTTCCATCCCTGTGGAAGCACATTGTAC	sequence:3020	-	50	POL
TTTGGAATATTGCTGGTGATCCTTTCCA	sequence:3042	-	42	POL
TTTTGTTCTATGCTGCCCTATTTCTAAG	sequence:3161	-	42	POL
TTTGTCTGGTGTGGTAAGTCCCCACCTC	sequence:3215	-	58	POL
TTTTGTCTGGTGTGGTAAGTCCCCACCT	sequence:3216	-	58	POL
TTTTTGTCTGGTGTGGTAAGTCCCCACC	sequence:3217	-	58	POL
TTTTTTGTCTGGTGTGGTAAGTCCCCAC	sequence:3218	-	54	POL
TTTGGATGGGTTATGAACTCCATCCTGA	sequence:3268	+	46	POL
TTTATCAGGATGGAGTTCATAACCCATC	sequence:3272	-	46	POL
TTTCTGGCAGCACTATAGGCTGTACTGT	sequence:3303	-	50	POL
TTTTCTGGCAGCACTATAGGCTGTACTG	sequence:3304	-	54	POL
TTTTTCTGGCAGCACTATAGGCTGTACT	sequence:3305	-	50	POL
TTTTCCCACTAACTTCTGTATGTCATTG	sequence:3347	-	42	POL
TTTAAATCCCTGGGTAAATCTGACTTGCC	sequence:3383	-	50	POL
TTTACCCAGGGATTAAGTAAGGCAATT	sequence:3394	+	42	POL
TTTCTGCCAGTTCTAGCTCTGCTTCTTC	sequence:3474	-	50	POL
TTTTCTGCCAGTTCTAGCTCTGCTTCTT	sequence:3475	-	50	POL
TTTAGAATCTCTCTGTTTTCTGCCAGTT	sequence:3490	-	42	POL
TTTTAGAATCTCTCTGTTTTCTGCCAGT	sequence:3491	-	42	POL
TTTTGATGGGTGATAATACTCCATGT	sequence:3527	-	42	POL
TTTGATATGTCCATTGGCCTTGCCCCTG	sequence:3579	-	54	POL
TTTACATCATTAGTGTGGGCACCCCTCG	sequence:3655	-	58	POL
TTTTACATCATTAGTGTGGGCACCCCTC	sequence:3656	-	54	POL
TTTCTGTGGTTATTTTTTGCCTGCTC	sequence:3693	-	42	POL
TTTCCCATATTACTATGCTTTCTGTGG	sequence:3712	-	42	POL
TTTAGGAGTCTTTCCCATATTACTATG	sequence:3722	-	42	POL
TTTCCATGTTTCTTTTGTATGGGTAG	sequence:3756	-	42	POL
TTTCACTAAAGGAGGGGTATTGACAAAC	sequence:3830	-	42	POL
TTTCTGCTCTACTATGGGTTCTTTCTC	sequence:3873	-	46	POL
TTTAGTCTCCCTGCTAGCTGCCCCATCT	sequence:3911	-	62	POL
TTTCCTAATTTAGTCTCCCTGCTAGCTG	sequence:3919	-	46	POL
TTTTCTAATTTAGTCTCCCTGCTAGCT	sequence:3920	-	46	POL
TTTGTTGTGTCAGTTAGGGTGACAACCT	sequence:3979	-	42	POL

Cpf1

TTTACTTCTAATCCCGAATCCTGCAAAG	sequence:4039	-	46	POL
TTTATCTGGTTGTGCTTGAATGATTCT	sequence:4094	-	42	POL
TTTTATCTGGTTGTGCTTGAATGATTCC	sequence:4095	-	42	POL
TTTGTGTGCTGGTACCCATGCCAGATAG	sequence:4178	-	54	POL
TTTCCTCCAATTCCTTTGTGTGCTGGTA	sequence:4192	-	46	POL
TTTTTAGATGGAATAGATAAGGCCCAAG	sequence:4266	+	42	POL
TTTTAGATGGAATAGATAAGGCCCAAGA	sequence:4267	+	42	POL
TTTAGATGGAATAGATAAGGCCCAAGAT	sequence:4268	+	42	POL
TTTCTCATGTTTCATCTTGGGCCTTATCT	sequence:4280	-	42	POL
TTTGCTACTACAGGTGGCAGGTTAAAAT	sequence:4339	-	42	POL
TTTTAACCTGCCACCTGTAGTAGCAAAA	sequence:4340	+	46	POL
TTTTAACCTGCCACCTGTAGTAGCAAAAAG	sequence:4341	+	50	POL
TTTCTTTTGTACTACAGGTGGCAGGTT	sequence:4344	-	46	POL
TTTAGCTGACATTTATCACAGCTGGCTA	sequence:4375	-	46	POL
TTTTAGCTGACATTTATCACAGCTGGCT	sequence:4376	-	46	POL
TTTAGAAGGAAAAGTTATCCTGGTAGCA	sequence:4466	+	42	POL
TTTCCTGCCCTGTTTCTGCTGGAATAAC	sequence:4527	-	50	POL
TTTTAAAAGAAAGTATGCTGTTTCTGCC	sequence:4546	-	42	POL
TTTTAAAATTAGCAGGAAGATGGCCAGT	sequence:4567	+	42	POL
TTTTAAAATTAGCAGGAAGATGGCCAGTA	sequence:4568	+	42	POL
TTTACTGGCCATCTTCTGCTAATTTTA	sequence:4570	-	42	POL
TTTTACTGGCCATCTTCTGCTAATTTT	sequence:4571	-	42	POL
TTTTTACTGGCCATCTTCTGCTAATTT	sequence:4572	-	42	POL
TTTCACCAGTACTACGGTTAAGGCCGCC	sequence:4625	+	58	POL
TTTGGGGATTGTAGGGAATCCAAATTC	sequence:4677	-	42	POL
TTTGGAATCCCTACAATCCCCAAAGTC	sequence:4680	+	46	POL
TTTGTACTGCTGTCTAAGATGTTCAGC	sequence:4770	-	42	POL
TTTTAAAAGAAAAGGGGGGATTGGGGGG	sequence:4817	+	54	POL cPPT
TTTTAAAAGAAAAGGGGGGATTGGGGGGT	sequence:4818	+	54	POL cPPT
TTTCCCCTGCACTGTACCCCCAATCCC	sequence:4833	-	67	POL cPPT
TTTTCGGGTTTATTACAGGGACAGCAGA	sequence:4931	+	50	POL
TTTCGGGTTTATTACAGGGACAGCAGAG	sequence:4932	+	50	POL
TTTATTACAGGGACAGCAGAGATCCACT	sequence:4939	+	50	POL
TTTCAAAGTGGATCTCTGCTGTCCCTG	sequence:4946	-	54	POL
TTTGCTGGTCTTTCCAAAGTGGATCTC	sequence:4957	-	50	POL
TTTGAAAGGACCAGCAAAGCTCCTCTG	sequence:4966	+	54	POL
TTTCCAGAGGAGCTTTGCTGGTCCTTTC	sequence:4970	-	54	POL
TTTCCATGTTCTAATCCTCATCCTGTCT	sequence:5115	-	42	POL
TTTTCCATGTTCTAATCCTCATCCTGTC	sequence:5116	-	46	POL
TTTCAGGGAAAGCTAGGGGATGGTTTTA	sequence:5168	+	46	VIF
TTTTATAGACATCACTATGAAAGCCCTC	sequence:5191	+	42	VIF
TTTATAGACATCACTATGAAAGCCCTCA	sequence:5192	+	42	VIF
TTTCTCCTGTATGCAGACCCCAATATGT	sequence:5278	-	46	VIF
TTTCCTCCATTCTATGGAGACTCCCTGA	sequence:5322	-	50	VIF
TTTTCTCCATTCTATGGAGACTCCCTG	sequence:5323	-	54	VIF
TTTTTCTCCATTCTATGGAGACTCCCT	sequence:5324	-	50	VIF
TTTTTCAGACTCTGCTATAAGAAAGGCC	sequence:5418	+	46	VIF
TTTTCAGACTCTGCTATAAGAAAGGCCT	sequence:5419	+	46	VIF
TTTCAGACTCTGCTATAAGAAAGGCCTT	sequence:5420	+	42	VIF
TTTGGTGTTATTAATGCTGCTAGTGCCA	sequence:5519	-	42	VIF
TTTTGGTGTTATTAATGCTGCTAGTGCC	sequence:5520	-	46	VIF
TTTTTGGTGTTATTAATGCTGCTAGTGC	sequence:5521	-	42	VIF
TTTCGTAACACTAGGCAAAGGTGGCTTT	sequence:5553	-	46	VIF
TTTGCCTAGTGTTACGAAACTGACAGAG	sequence:5562	+	46	VIF
TTTTAGAGGAGCTTAAGAATGAAGCTGT	sequence:5659	+	42	VPR
TTTAGAGGAGCTTAAGAATGAAGCTGTT	sequence:5660	+	42	VPR
TTTTCTAGGATTTGGCTCCATGGCTTA	sequence:5693	+	50	VPR
TTTCCTAGGATTTGGCTCCATGGCTTAG	sequence:5694	+	50	VPR
TTTGGCTCCATGGCTTAGGGCAACATAT	sequence:5704	+	50	VPR
TTTCATAGATATGTTGCCCTAAGCCATG	sequence:5712	-	42	VPR
TTTATCCATTTCAGAAATGGGTGTGCGAC	sequence:5799	+	46	VPR
TTTCAGAAATGGGTGTGCGACATAGCAGA	sequence:5807	+	46	VPR
TTTCTTGCTCTCCTCTGTTGAGTAACGC	sequence:5839	-	50	VPR
TTTAGGCTGACTTCCTGGATGCTTCCAG	sequence:5895	-	58	5' TAT
TTTTAGGCTGACTTCCTGGATGCTTCCA	sequence:5896	-	54	5' TAT
TTTACAATAGCAAGTGGTACAAGCAGTT	sequence:5922	-	42	5' TAT
TTTTACAATAGCAAGTGGTACAAGCAGT	sequence:5923	-	42	5' TAT
TTTTTACAATAGCAAGTGGTACAAGCAG	sequence:5924	-	42	5' TAT
TTTCACAACAAAAGCCTTAGGCATCTCC	sequence:5976	+	46	5' TAT
TTTCTCCTTCACTCTCATTGCCACTGTC	sequence:6247	-	50	VPU
TTTCCACCCCATCTCCACAAGTGCTGA	sequence:6277	-	58	VPU

Cpf1

TTTGCTGAGGGCTATTGAGGCGCAACAG	sequence:7933	+	58	RRE
TTTCCACAGCCAGGATTCTTGCCTGGAG	sequence:7997	-	58	RRE
TTTAGGTATCTTTCCACAGCCAGGATTC	sequence:8007	-	50	RRE
TTTAGTTCCTGACTCCAATACTGTAGGA	sequence:8667	-	46	3' REV
TTTCCCTTACAGTAGGCCATCCAACCAC	sequence:8871	-	54	NEF
TTTCCAGGTCTCGAGATGCTGCTCCCAC	sequence:8931	-	62	NEF
TTTTCCAGGTCTCGAGATGCTGCTCCA	sequence:8932	-	62	NEF
TTTTTCCAGGTCTCGAGATGCTGCTCCC	sequence:8933	-	62	NEF
TTTTCCAGTCACACCTCAGGTACCTTTA	sequence:9044	+	50	NEF
TTTCCAGTCACACCTCAGGTACCTTTAA	sequence:9045	+	46	NEF
TTTAAGACCAATGACTTACAAGGCAGCT	sequence:9068	+	46	NEF
TTTAAAAAGTGGCTAAGATCTACAGCTG	sequence:9091	-	42	NEF
TTTTAAAAGAAAAGGGGGGACTGGAAGG	sequence:9112	+	50	NEF
TTTTAAAAGAAAAGGGGGGACTGGAAGGG	sequence:9113	+	54	NEF

NL4-3 Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
GACAAGAGATCCTTGATCTGTGG	sequence:28	+	45	LTR U3
TGTGGTAGATCCACAGATCAAGG	sequence:38, 9113	-	45	LTR U3 Nef
CTGTGGATCTACCACACACAAGG	sequence:45, 9120	+	50	LTR U3 Nef
CAGGGAAGTAGCCTTGTGTGTGG	sequence:56, 9131	-	55	LTR U3 Nef
CACAAGGCTACTTCCCTGATTGG	sequence:61, 9136	+	50	LTR U3 Nef
GTGTGTAGTTCTGCCAATCAGGG	sequence:74, 9149	-	45	LTR U3 Nef
GGTGTGTAGTTCTGCCAATCAGG	sequence:75, 9150	-	50	LTR U3 Nef
GATTGGCAGAACTACACACCAGG	sequence:78, 9153	+	50	LTR U3 Nef
ATTGGCAGAACTACACACCAGGG	sequence:79, 9154	+	45	LTR U3 Nef
CAGAACTACACACCAGGGCCAGG	sequence:84, 9159	+	60	LTR U3 Nef
AGAACTACACACCAGGGCCAGGG	sequence:85, 9160	+	55	LTR U3 Nef
GGATATCTGATCCCTGGCCCTGG	sequence:96	-	60	LTR U3
GTCAGTGGATATCTGATCCCTGG	sequence:102	-	50	LTR U3
ATCAGATATCCACTGACCTTTGG	sequence:108	+	40	LTR U3
GATATCCACTGACCTTTGGATGG	sequence:112, 9187	+	45	LTR U3 Nef
AAGCACCATCCAAAGGTCAGTGG	sequence:117	-	50	LTR U3
TAACCTGAAGCACCATCCAAAGG	sequence:124	-	40	LTR U3
TCTACTTGCTCTGGTTCAACTGG	sequence:150	-	45	LTR U3
TGAACCAGAGCAAGTAGAAGAGG	sequence:155	+	45	LTR U3
TTGGCCTCTTCTACTTGCTCTGG	sequence:159	-	50	LTR U3
CAAGTAGAAGAGGCCAATGAAGG	sequence:165	+	45	LTR U3
GCTGTTGTTCTCTCCTTCATTGG	sequence:178	-	45	LTR U3
TTACACCCTATGAGCCAGCATGG	sequence:204	+	50	LTR U3
TACACCCTATGAGCCAGCATGGG	sequence:205	+	50	LTR U3
CCCTATGAGCCAGCATGGGATGG	sequence:209	+	60	LTR U3
TCCATCCCATGCTGGCTCATAGG	sequence:210	-	55	LTR U3
TATGAGCCAGCATGGGATGGAGG	sequence:212	+	55	LTR U3
CCAGCATGGGATGGAGGACCCGG	sequence:218	+	65	LTR U3
GCATGGGATGGAGGACCCGGAGG	sequence:221	+	70	LTR U3
CATGGGATGGAGGACCCGGAGGG	sequence:222	+	65	LTR U3
ACACTAATACTTCTCCCTCCGGG	sequence:236	-	45	LTR U3
CACACTAATACTTCTCCCTCCGG	sequence:237	-	45	LTR U3
CGGAGGGAGAAGTATTAGTGTGG	sequence:238	+	50	LTR U3
CCTCCTAGCATTTCGTCACATGG	sequence:272	+	50	LTR U3
GGGCCATGTGACGAAATGCTAGG	sequence:275	-	55	LTR U3
CATGGCCGAGAGCTGCATCCGG	sequence:290	+	65	LTR U3
GTACTIONCGGATGCAGCTCTCGGG	sequence:295, 9370	-	60	LTR U3 Nef
AGTACTCCGGATGCAGCTCTCGG	sequence:296, 9371	-	55	LTR U3 Nef
CAGCAGTCTTTGTAGTACTCCGG	sequence:309	-	45	LTR U3
CTGACATCGAGCTTTCTACAAGG	sequence:329	+	45	LTR U3
TGACATCGAGCTTTCTACAAGGG	sequence:330	+	40	LTR U3
TTCTACAAGGGACTTTCCGCTGG	sequence:342	+	50	LTR U3
TCTACAAGGGACTTTCCGCTGGG	sequence:343	+	50	LTR U3
CTACAAGGGACTTTCCGCTGGGG	sequence:344, 9419	+	55	LTR U3
CTTTCCGCTGGGGACTTTCCAGG	sequence:354, 9429	+	60	LTR U3
TTTCCGCTGGGGACTTTCCAGGG	sequence:355, 9430	+	55	LTR U3
CCGCTGGGGACTTTCCAGGGAGG	sequence:358, 9433	+	70	LTR U3
GGGGACTTTCCAGGGAGGTGTGG	sequence:363	+	65	LTR U3
TTTCCAGGGAGGTGTGGCCTGGG	sequence:369	+	60	LTR U3
CAGCTGCTTTTTGCCTGTACTGG	sequence:433, 9508	+	50	LTR U3
AGCTGCTTTTTGCCTGTACTGGG	sequence:434, 9509	+	45	LTR U3
TTGCCTGTACTGGTCTCTCTGG	sequence:443, 9518	+	55	LTR U3
TAACCAGAGAGACCCAGTACAGG	sequence:446, 9521	-	50	LTR U3
GGTTAGACCAGATCTGAGCCTGG	sequence:464,9539	+	55	LTR R
GTTAGACCAGATCTGAGCCTGGG	sequence:465, 9540	+	50	LTR R
AGAGCTCCCAGGCTCAGATCTGG	sequence:471, 9546	-	60	LTR R
ATCTGAGCCTGGGAGCTCTCTGG	sequence:475,9550	+	60	LTR R
TAGTTAGCCAGAGAGCTCCCAGG	sequence:482, 9557	-	55	LTR R
TGGGAGCTCTCTGGCTAACTAGG	sequence:484, 9559	+	55	LTR R
GGGAGCTCTCTGGCTAACTAGGG	sequence:485, 9560	+	55	LTR R
CTTTATTGAGGCTTAAGCAGTGG	sequence:511, 9586	-	40	LTR R
ACTCAAGGCAAGCTTTATTGAGG	sequence:523,9598	-	40	LTR R
CACACTACTTTGAGCACTCAAGG	sequence:538	-	45	LTR R
GCCCGTCTGTTGTGTGACTCTGG	sequence:562, 9637	+	60	LTR U5
ACCAGAGTCACACAACAGACGGG	sequence:563, 9638	-	50	LTR U5
TACCAGAGTCACACAACAGACGG	sequence:564, 9639	-	45	LTR U5
CACTGACTAAAAGGGTCTGAGGG	sequence:596, 9671	-	45	LTR U5
ACACTGACTAAAAGGGTCTGAGG	sequence:597, 9672	-	45	LTR U5
TCAGACCCTTTTAGTCAGTGTGG	sequence:599, 9674	+	45	LTR U5
GTGTGGAAAATCTCTAGCAGTGG	sequence:616	+	45	LTR U5

TCTAGCAGTGGCGCCCGAACAGG	sequence:628	+	65	LTR U5
CTAGCAGTGGCGCCCGAACAGGG	sequence:629	+	65	LTR U5
TCGCTTTCAAGTCCCTGTTCCGGG	sequence:641	-	50	
TTCGCTTTCAAGTCCCTGTTCCGG	sequence:642	-	45	
GAAAGCGAAAGTAAAGCCAGAGG	sequence:656	+	45	
CTGCGTCGAGAGATCTCCTCTGG	sequence:672	-	60	
CAGAGGAGATCTCTCGACGCAGG	sequence:673	+	60	
AGATCTCTCGACGCAGGACTCGG	sequence:679	+	55	
TCGGCTTGCTGAAGCGCGCACGG	sequence:698	+	65	
GCTGAAGCGCGCACGGCAAGAGG	sequence:705	+	70	
CCAAAAATTTGACTAGCGGAGG	sequence:753	+	40	Psi
TTTGACTAGCGGAGGCTAGAAGG	sequence:761	+	50	Psi
GAGGCTAGAAGGAGAGATGGG	sequence:772	+	50	Psi
AGAGATGGGTGCGAGAGCGTCCGG	sequence:786	+	60	Psi
GCGAGAGCGTCGGTATTAAGCGG	sequence:796	+	55	Gag
CGAGAGCGTCGGTATTAAGCGGG	sequence:797	+	55	Gag
GAGAGCGTCGGTATTAAGCGGGG	sequence:798	+	55	Gag
AGAGCGTCGGTATTAAGCGGGGG	sequence:799	+	55	Gag
GCGGGGGAGAATTAGATAAATGG	sequence:815	+	45	Gag
CGGGGGAGAATTAGATAAATGGG	sequence:816	+	40	Gag
GAAAAAATTCGGTTAAGGCCAGG	sequence:838	+	40	Gag
AAAATTCGGTTAAGGCCAGGGGG	sequence:841	+	45	Gag
AACATATAGTATGGGCAAGCAGG	sequence:884	+	40	Gag
ACATATAGTATGGGCAAGCAGGG	sequence:885	+	40	Gag
GAACGATTCGCAGTTAATCCTGG	sequence:913	+	45	Gag
GGCCTTTAGAGACATCAGAAGG	sequence:934	+	45	Gag
AGCCTTCTGATGTCTCTAAAAGG	sequence:936	-	40	Gag
AGAAGGCTGTAGACAAATACTGG	sequence:951	+	40	Gag
GAAGGCTGTAGACAAATACTGGG	sequence:952	+	40	Gag
CTACAACCATCCCTTCAGACAGG	sequence:979	+	50	Gag
TCTGATCCTGTCTGAAGGGATGG	sequence:985	-	50	Gag
TTCTTCTGATCCTGTCTGAAGGG	sequence:989	-	40	Gag
GTTCTTCTGATCCTGTCTGAAGG	sequence:990	-	45	Gag
TCCTCTATTGTGTGCATCAAAGG	sequence:1040	+	40	Gag
TCCTTTGATGCACACAATAGAGG	sequence:1041	-	40	Gag
GGAAGCCTTAGATAAAGATAGAGG	sequence:1083	+	40	Gag
TGACACAGGAAACAACAGCCAGG	sequence:1149	+	50	Gag
TAGGGTAATTTGGCTGACCTGG	sequence:1167	-	45	Gag
CCCTATAGTGCAGAACCTCCAGG	sequence:1185	+	55	Gag
CCTATAGTGCAGAACCTCCAGGG	sequence:1186	+	50	Gag
CTATAGTGCAGAACCTCCAGGGG	sequence:1187	+	50	Gag
GCAGAACCTCCAGGGGCAAATGG	sequence:1194	+	60	Gag
GATGTACCATTTGCCCTGGAGG	sequence:1200	-	55	Gag
CCAGGGGCAAATGGTACATCAGG	sequence:1203	+	55	Gag
GGTATTTAAATCTGTGGGGTGG	sequence:1325	-	40	Gag
ATACCATGCTAAACACAGTGGGG	sequence:1343	+	40	Gag
TACCATGCTAAACACAGTGGGGG	sequence:1344	+	45	Gag
ACCATGCTAAACACAGTGGGGGG	sequence:1345	+	50	Gag
TCCCCCACTGTGTTTAGCATGG	sequence:1346	-	55	Gag
TTCTGCAGCTTCTCATTGATGG	sequence:1400	-	45	Gag
TCAATGAGGAAGCTGCAGAATGG	sequence:1403	+	45	Gag
CAATGAGGAAGCTGCAGAATGGG	sequence:1404	+	45	Gag
AGATTGCATCCAGTGCATGCAGG	sequence:1429	+	50	Gag
GATTGCATCCAGTGCATGCAGGG	sequence:1430	+	50	Gag
GCAATAGGCCCTGCATGCACTGG	sequence:1438	-	60	Gag
CATGCAGGGCCTATTGCACCAGG	sequence:1444	+	60	Gag
CTCATCTGGCCTGGTGAATAGG	sequence:1453	-	55	Gag
CTTGGTTCTCTCATCTGGCCTGG	sequence:1462	-	55	Gag
CAGGCCAGATGAGAGAACCAAGG	sequence:1463	+	55	Gag
AGGCCAGATGAGAGAACCAAGGG	sequence:1464	+	50	Gag
GGCCAGATGAGAGAACCAAGGGG	sequence:1465	+	55	Gag
TTCCCTTGGTTCTCTCATCTGG	sequence:1467	-	50	Gag
CCAAGGGGAAGTGACATAGCAGG	sequence:1480	+	55	Gag
AGGAACTACTAGTACCCTTCAGG	sequence:1500	+	45	Gag
AGTACCCTTCAGGAACAAATAGG	sequence:1510	+	40	Gag
CCCTTCAGGAACAAATAGGATGG	sequence:1514	+	45	Gag
TCCATCCTATTTGTTCTGAAGG	sequence:1515	-	40	Gag
CATAATCCACCTATCCCAGTAGG	sequence:1543	+	45	Gag
ATTTCTCCTACTGGGATAGGTGG	sequence:1549	-	45	Gag
TAGATTTCTCCTACTGGGATAGG	sequence:1552	-	40	Gag
GTATAGCCCTACCAGCATTCTGG	sequence:1617	+	50	Gag

TTATGTCCAGAATGCTGGTAGGG	sequence:1623	-	40	Gag
CTTATGTCCAGAATGCTGGTAGG	sequence:1624	-	45	Gag
AGCATTCTGGACATAAGACAAGG	sequence:1630	+	40	Gag
GGACATAAGACAAGGACCAAAGG	sequence:1638	+	45	Gag
TAGTCTCTAAAGGGTTCCTTTGG	sequence:1654	-	40	Gag
CGGTCTACATAGTCTCTAAAGGG	sequence:1663	-	40	Gag
TCGGTCTACATAGTCTCTAAAGG	sequence:1664	-	40	Gag
AGCCGAGCAAGCTTCACAAGAGG	sequence:1704	+	55	Gag
TACCTCTTGTGAAGCTTGCTCGG	sequence:1706	-	45	Gag
TTGGATGACAGAAACCTTGTTGG	sequence:1734	+	40	Gag
GTTCGCATTTTGGACCAACAAGG	sequence:1748	-	45	Gag
TACAATCTGGGTTTCGATTTTGG	sequence:1758	-	40	Gag
ATTTCTTCTAGTGTGCTCCTGG	sequence:1804	-	45	Gag
AGAAATGATGACAGCATGTCAGG	sequence:1821	+	40	Gag
GAAATGATGACAGCATGTCAGGG	sequence:1822	+	40	Gag
GATGACAGCATGTCAGGGAGTGG	sequence:1827	+	55	Gag
ATGACAGCATGTCAGGGAGTGGG	sequence:1828	+	50	Gag
TGACAGCATGTCAGGGAGTGGGG	sequence:1829	+	55	Gag
GACAGCATGTCAGGGAGTGGGGG	sequence:1830	+	60	Gag
ACAGCATGTCAGGGAGTGGGGGG	sequence:1831	+	60	Gag
TGTCAGGGAGTGGGGGGACCCGG	sequence:1837	+	70	Gag
AAAACCTTGTCTTATGGCCGGG	sequence:1855	-	40	Gag
CAAACTCTTGCTTTATGGCCGG	sequence:1856	-	40	Gag
CGGCCATAAAGCAAGAGTTTTGG	sequence:1857	+	45	Gag
CAGCCAAAACCTTGCTTTATGG	sequence:1860	-	40	Gag
TGGTAGCTGGATTTGTTACTTGG	sequence:1893	-	40	Gag
GGCACATAGCCAAAATTGCAGG	sequence:1985	+	45	Gag
GCACATAGCCAAAATTGCAGGG	sequence:1986	+	40	Gag
CCAAAATTGCAGGGCCCTAGG	sequence:1994	+	55	Gag
TTGCAGGGCCCTAGGAAAAGG	sequence:2001	+	55	Gag
TGCAGGGCCCTAGGAAAAGGG	sequence:2002	+	55	Gag
CCCTAGGAAAAGGGCTGTTGG	sequence:2009	+	55	Gag
TCCAACAGCCCTTTTCTAGGG	sequence:2010	-	45	Gag
TTCCAACAGCCCTTTTCTAGG	sequence:2011	-	45	Gag
AAAAAGGGCTGTTGGAAATGTGG	sequence:2017	+	40	Gag
GGGCTGTTGGAAATGTGGAAAGG	sequence:2022	+	50	Gag
TGTTGGAAATGTGGAAAGGAAGG	sequence:2026	+	40	Gag
GAAAGATTGACTGAGAGACAGG	sequence:2058	+	40	Gag
GAAGATCTGGCCTTCCCACAAGG	sequence:2094	+	55	Gag Pol
AAGATCTGGCCTTCCCACAAGGG	sequence:2095	+	50	Gag Pol
TCTGGCCTTCCCACAAGGGAAGG	sequence:2099	+	60	Gag Pol
CCTTCCCACAAGGGAAGGCCAGG	sequence:2104	+	65	Gag Pol
CTTCCCACAAGGGAAGGCCAGGG	sequence:2105	+	60	Gag Pol
AATTCCCTGGCCTTCCCTTGTGG	sequence:2109	-	55	Gag Pol
CTGCTCTGAAGAAAATTCCCTGG	sequence:2122	-	45	Gag Pol
TCTGGTGGGGCTGTTGGCTCTGG	sequence:2146	-	65	Gag Pol
CTCTCTTCTGGTGGGGCTGTTGG	sequence:2152	-	60	Gag Pol
CCCACCAGAAGAGAGCTTCAGG	sequence:2159	+	60	Gag Pol
ACCTGAAGCTCTCTTCTGGTGGG	sequence:2160	-	50	Gag Pol
AACCTGAAGCTCTCTTCTGGTGG	sequence:2161	-	50	Gag Pol
CCAGAAGAGAGCTTCAGGTTTGG	sequence:2164	+	50	Gag Pol
CAGAAGAGAGCTTCAGGTTTGGG	sequence:2165	+	45	Gag Pol
AGAAGAGAGCTTCAGGTTTGGGG	sequence:2166	+	45	Gag Pol
AACAACCTCCTCAGAAGCAGG	sequence:2196	+	50	Gag Pol
ATCGGCTCCTGCTTCTGAGAGGG	sequence:2203	-	55	Gag Pol
TATCGGCTCCTGCTTCTGAGAGG	sequence:2204	-	55	Gag Pol
GAAGCAGGAGCCGATAGACAAGG	sequence:2211	+	55	Gag Pol
GGATACAGTTCCTTGTCTATCGG	sequence:2221	-	40	Gag Pol
AGTGATCTGAGGGAAGCTAAAGG	sequence:2242	-	45	Gag Pol
GCTTCCCTCAGATCACTCTTTGG	sequence:2248	+	50	Gag Pol
GCTGCCAAAGAGTGATCTGAGGG	sequence:2252	-	50	Gag Pol
CGTGCCAAAGAGTGATCTGAGG	sequence:2253	-	55	Gag Pol
CCCTCGTCACAATAAAGATAGG	sequence:2277	+	45	Gag Pol
CCCTCGTCACAATAAAGATAGGG	sequence:2278	+	40	Gag Pol
CCTCGTCACAATAAAGATAGGGG	sequence:2279	+	40	Gag Pol
CTCGTCACAATAAAGATAGGGGG	sequence:2280	+	40	Gag Pol
TCGTACAATAAAGATAGGGGGG	sequence:2281	+	40	Gag Pol
AAAGATAGGGGGGCAATTAAGG	sequence:2291	+	40	Pol
CCAAAATGATAGGGGGAATTGG	sequence:2382	+	40	Pol
AAAATGATAGGGGGAATTGGAGG	sequence:2385	+	40	Pol
ATCTGCGGACATAAAGCTATAGG	sequence:2448	+	40	Pol

CCTACACCTGTCAACATAATTGG	sequence:2487	+	40	Pol
ACTTTTGGGCCATCCATTCCTGG	sequence:2589	-	50	Pol
TTAGGAATACCACATCCTGCAGG	sequence:2823	+	45	Pol
TAGGAATACCACATCCTGCAGGG	sequence:2824	+	45	Pol
TGTTTTAACCTGCAGGATGTGG	sequence:2832	-	45	Pol
ATCAGTAACAGTACTGGATGTGG	sequence:2861	+	40	Pol
TCAGTAACAGTACTGGATGTGGG	sequence:2862	+	40	Pol
TCAGTACAATGTGCTTCCACAGG	sequence:2981	+	45	Pol
CAGTACAATGTGCTTCCACAGGG	sequence:2982	+	45	Pol
ACAATGTGCTTCCACAGGGATGG	sequence:2986	+	50	Pol
GTGCTTCCACAGGGATGGAAAGG	sequence:2991	+	55	Pol
GGTGATCCTTCCATCCCTGTGG	sequence:2997	-	55	Pol
ATGCTACACTGGAATATTGCTGG	sequence:3018	-	40	Pol
AACTGAGACAACATCTGTTGAGG	sequence:3160	+	40	Pol
TGAGACAACATCTGTTGAGGTGG	sequence:3163	+	45	Pol
GAGACAACATCTGTTGAGGTGGG	sequence:3164	+	45	Pol
AGACAACATCTGTTGAGGTGGGG	sequence:3165	+	45	Pol
AGAAAGAACCTCCATTCCTTTGG	sequence:3214	+	40	Pol
AGAACCTCCATTCCTTTGGATGG	sequence:3218	+	45	Pol
GAACCTCCATTCCTTTGGATGGG	sequence:3219	+	45	Pol
TCATAACCCATCCAAAGGAATGG	sequence:3225	-	40	Pol
GGAGTTCATAACCCATCCAAAGG	sequence:3230	-	45	Pol
GTAAGTCCATTTATCAGGATGG	sequence:3251	-	40	Pol
GGCTGTACTGTCCATTTATCAGG	sequence:3255	-	45	Pol
GCCTATAGTGCTGCCAGAAAAGG	sequence:3275	+	50	Pol
TCCTTTTCTGGCAGCACTATAGG	sequence:3276	-	45	Pol
ACAGTCCAGCTGTCTTTTCTGG	sequence:3288	-	50	Pol
TGGCAAGTCAGATTTATGCAGG	sequence:3345	+	45	Pol
GGGCAAGTCAGATTTATGCAGGG	sequence:3346	+	45	Pol
AGCTCTGCTTCTTCTGTTAGTGG	sequence:3429	-	45	Pol
AGAAGAAGCAGAGCTAGAACTGG	sequence:3437	+	45	Pol
AGCTAGAAGTGGCAGAAAACAGG	sequence:3448	+	45	Pol
GCTAGAAGTGGCAGAAAACAGGG	sequence:3449	+	45	Pol
CAGGGAGATTCTAAAAGAACCGG	sequence:3467	+	40	Pol
TAGCAGAAATACAGAAGCAGGGG	sequence:3526	+	40	Pol
GAAATACAGAAGCAGGGGCAAGG	sequence:3531	+	50	Pol
AGAAGCAGGGGCAAGGCCAATGG	sequence:3538	+	60	Pol
AAAAGGAAACATGGGAAGCATGG	sequence:3730	+	40	Pol
AGGAAACATGGGAAGCATGGTGG	sequence:3733	+	50	Pol
AAGCATGGTGGACAGAGTATTGG	sequence:3745	+	45	Pol
CAGAGTATTGGCAAGCCACCTGG	sequence:3757	+	55	Pol
AAGCCACCTGGATTCCTGAGTGG	sequence:3769	+	55	Pol
AGCCACCTGGATTCCTGAGTGGG	sequence:3770	+	55	Pol
AAACTCCCACTCAGGAATCCAGG	sequence:3775	-	50	Pol
GTATTGACAAACTCCCACTCAGG	sequence:3783	-	45	Pol
CCCCTCCCTTAGTGAAGTTATGG	sequence:3805	+	50	Pol
ACCATAACTTCACTAAGGGAGGG	sequence:3806	-	40	Pol
TACCATAACTTCACTAAGGGAGG	sequence:3807	-	40	Pol
CTGGTACCATAACTTCACTAAGG	sequence:3811	-	40	Pol
ATGTAGATGGGGCAGCCAATAGG	sequence:3871	+	50	Pol
TGTAGATGGGGCAGCCAATAGGG	sequence:3872	+	50	Pol
ACAAAAGTTGTCCCCCTAACGG	sequence:3938	+	40	Pol
GATTTGTTGTGTCCGTTAGGGGG	sequence:3950	-	45	Pol
TGATTTGTTGTGTCCGTTAGGGG	sequence:3951	-	40	Pol
CTGATTTGTTGTGTCCGTTAGGG	sequence:3952	-	40	Pol
TCTGATTTGTTGTGTCCGTTAGG	sequence:3953	-	40	Pol
AGCAATTCATCTAGCTTTGCAGG	sequence:3989	+	40	Pol
TCATCTAGCTTTGCAGGATTCGG	sequence:3995	+	40	Pol
CATCTAGCTTTGCAGGATTCGGG	sequence:3996	+	45	Pol
AGGAAAAAGTCTACCTGGCATGG	sequence:4132	+	45	Pol
GGAAAAAGTCTACCTGGCATGGG	sequence:4133	+	45	Pol
TGTGTGCTGGTACCCATGCCAGG	sequence:4145	-	60	Pol
CCAGCACACAAAGGAATTGGAGG	sequence:4158	+	50	Pol
GTAGATAAATTGGTCAGTGCTGG	sequence:4191	+	40	Pol
TCACAGTAATTGGAGAGCAATGG	sequence:4274	+	40	Pol
CTTTTGCTACTACAGGTGGTAGG	sequence:4310	-	45	Pol
TAGCTGACATTTATCACAGCTGG	sequence:4342	-	40	Pol
CTAAAAGGGGAAGCCATGCATGG	sequence:4362	+	50	Pol
GGACAAGTAGACTGTAGCCCAGG	sequence:4383	+	55	Pol
TAGACTGTAGCCCAGGAATATGG	sequence:4390	+	45	Pol
AATCTAGCTGCCATATTCCTGGG	sequence:4400	-	40	Pol

CAATCTAGCTGCCATATTCCTGG	sequence:4401	-	45	Pol
GTAGCAGTTCATGTAGCCAGTGG	sequence:4452	+	50	Pol
GAAGTAATTCCAGCAGAGACAGG	sequence:4488	+	45	Pol
GTTTCTTGCCCTGTCTCTGCTGG	sequence:4497	-	55	Pol
GGCGGCCTTAACTGTAGTACTGG	sequence:4594	-	55	Pol
CTACAGTTAAGGCCCGCTGTTGG	sequence:4600	+	55	Pol
CAGTTAAGGCCCGCTGTTGGTGG	sequence:4603	+	60	Pol
AGTTAAGGCCCGCTGTTGGTGGG	sequence:4604	+	55	Pol
TAAGGCCCGCTGTTGGTGGGCGG	sequence:4607	+	65	Pol
AAGGCCCGCTGTTGGTGGGCGGG	sequence:4608	+	70	Pol
GATCCCCGCCACCAACAGGCGG	sequence:4612	-	70	Pol
CTTGATCCCCGCCACCAACAGG	sequence:4615	-	65	Pol
TTGGTGGGCGGGGATCAAGCAGG	sequence:4619	+	65	Pol
CCCTACAATCCCCAAAGTCAAGG	sequence:4653	+	50	Pol
TCCTTGACTTTGGGGATTGTAGG	sequence:4654	-	45	Pol
TATAGGACAGGTAAGAGATCAGG	sequence:4712	+	40	Pol
TAAAAGAAAAGGGGGGATTGGGG	sequence:4784	+	40	Pol
AAAAGAAAAGGGGGGATTGGGGG	sequence:4785	+	45	Pol cPPT
AAAGAAAAGGGGGGATTGGGGG	sequence:4786	+	50	Pol cPPT
GGGATTGGGGGTACAGTGCAGG	sequence:4797	+	65	Pol
GGATTGGGGGTACAGTGCAGGG	sequence:4798	+	60	Pol
GATTGGGGGTACAGTGCAGGGG	sequence:4799	+	60	Pol
GGGACAGCAGAGATCCAGTTTGG	sequence:4912	+	55	Pol
AGCAGAGATCCAGTTTGAAAGG	sequence:4917	+	45	Pol
TTTGCTGGTCCTTTCCAACTGG	sequence:4926	-	45	Pol
AAGGACCAGCAAAGCTCCTCTGG	sequence:4936	+	55	Pol
CCAGCAAAGCTCCTCTGGAAAGG	sequence:4941	+	55	Pol
AAGCTCCTCTGGAAAGGTGAAGG	sequence:4947	+	50	Pol
AGCTCCTCTGGAAAGGTGAAGGG	sequence:4948	+	50	Pol
GCTCCTCTGGAAAGGTGAAGGGG	sequence:4949	+	55	Pol
CTGCCCTTACCTTTCCAGAGG	sequence:4952	-	60	Pol
GCAAAGATCATCAGGGATTATGG	sequence:5022	+	40	Pol
GATTATGGAAAACAGATGGCAGG	sequence:5037	+	40	Pol
TGATTGTGTGGCAAGTAGACAGG	sequence:5063	+	45	Pol Vif
TGTGGCAAGTAGACAGGATGAGG	sequence:5069	+	50	Pol Vif
ACAGGATGAGGATTAACACATGG	sequence:5081	+	40	Pol Vif
TTTCAAGGAAAGCTAAGGACTGG	sequence:5132	+	40	Vif
TCAGAAGTACACATCCCCTAGG	sequence:5197	+	45	Vif
CAGAAGTACACATCCCCTAGGG	sequence:5198	+	45	Vif
AGAAGTACACATCCCCTAGGGG	sequence:5199	+	45	Vif
CTAATTTAGCATCCCCTAGTGGG	sequence:5211	-	40	Vif
ACTAATTTAGCATCCCCTAGTGG	sequence:5212	-	40	Vif
ACATATTGGGGTCTGCATACAGG	sequence:5242	+	45	Vif
TGCATACAGGAGAAAGAGACTGG	sequence:5255	+	45	Vif
GTCAGGGAGTCTCCATAGAATGG	sequence:5285	+	50	Vif
AGGGAGTCTCCATAGAATGGAGG	sequence:5288	+	50	Vif
TTAGTTGGTCTGCTAGGTCAGGG	sequence:5337	-	45	Vif
ATTAGTTGGTCTGCTAGGTCAGG	sequence:5338	-	45	Vif
GATGAATTAGTTGGTCTGCTAGG	sequence:5343	-	40	Vif
TAGGACGTATAGTTAGTCTAGG	sequence:5414	+	40	Vif
CCTAGGTGTGAATATCAAGCAGG	sequence:5431	+	45	Vif
CAAGCAGGACATAACAAGGTAGG	sequence:5446	+	45	Vif
GGTAGGATCTCTACAGTACTTGG	sequence:5463	+	45	Vif
AGCCACCTTTGCCTAGTGTAGG	sequence:5519	+	50	Vif
TTCCTAACACTAGGCAAAGGTGG	sequence:5521	-	45	Vif
AGTTTCTAACACTAGGCAAAGG	sequence:5524	-	40	Vif
TCTGTGAGTTTCTAACACTAGG	sequence:5530	-	40	Vif
GGAAACTGACAGAGGACAGATGG	sequence:5540	+	50	Vif
GAACAAGCCCCAGAAGACCAAGG	sequence:5562	+	55	Vif Vpr
AACAAGCCCCAGAAGACCAAGGG	sequence:5563	+	50	Vif Vpr
CTGTGGCCCTTGGTCTTCTGGGG	sequence:5569	-	60	Vif Vpr
TCTGTGGCCCTTGGTCTTCTGGG	sequence:5570	-	55	Vif Vpr
CTCTGTGGCCCTTGGTCTTCTGG	sequence:5571	-	60	Vif Vpr
AGAAGACCAAGGGCCACAGAGGG	sequence:5573	+	55	Vif Vpr
ATGGCTCCCTCTGTGGCCCTTGG	sequence:5579	-	65	Vif Vpr
TCATTGTATGGCTCCCTCTGTGG	sequence:5586	-	50	Vif Vpr
AGAGGGAGCCATAACAATGAATGG	sequence:5590	+	45	Vif Vpr
ATGGACACTAGAGCTTTTAGAGG	sequence:5609	+	40	Vif Vpr
AAGTTATGGAGCCATATCCTAGG	sequence:5661	-	40	Vpr
AGGATATGGCTCCATAACTTAGG	sequence:5664	+	40	Vpr
ATGAACTTACGGGGATACTTGG	sequence:5698	+	40	Vpr

TGAAACTTACGGGGATACTTGGG	sequence:5699	+	40	Vpr
ACTTACGGGGATACTTGGGCAGG	sequence:5703	+	55	Vpr
CGGGGATACTTGGGCAGGAGTGG	sequence:5708	+	65	Vpr
GTCGACACCCAATTCTGAAATGG	sequence:5768	-	45	Vpr
GGGTGTGCACATAGCAGAATAGG	sequence:5781	+	50	Vpr
GAATAGGCGTTACTCGACAGAGG	sequence:5797	+	50	Vpr
CGACAGAGGAGAGCAAGAAATGG	sequence:5811	+	50	Vpr
GGCTCTAGTCTAGGATCTACTGG	sequence:5836	-	50	Vpr
TAGATCCTAGACTAGAGCCCTGG	sequence:5840	+	50	Vpr 5' Tat
TGCTTCCAGGGCTCTAGTCTAGG	sequence:5845	-	55	Vpr 5' Tat
CTAGAGCCCTGGAAGCATCCAGG	sequence:5851	+	60	5' Tat
TGACTTCCTGGATGCTTCCAGGG	sequence:5857	-	50	5' Tat
CTGACTTCCTGGATGCTTCCAGG	sequence:5858	-	55	5' Tat
GCAGTTTTAGGCTGACTTCCTGG	sequence:5869	-	50	5' Tat
AAAGCCTTAGGCATCTCCTATGG	sequence:5950	+	45	5' Tat
CCTTAGGCATCTCCTATGGCAGG	sequence:5954	+	55	5' Tat
TCTCCTATGGCAGGAAGAAGCGG	sequence:5963	+	50	5' Tat
TCTCCGCTTCTTCTGCCATAGG	sequence:5966	-	55	5' Tat
ATAGAAAGAGCAGAAGACAGTGG	sequence:6196	+	40	Vpu
GACAGTGGAATGAGAGTGAAGG	sequence:6211	+	50	Vpu
AGGAGAAGTATCAGCACTTGTGG	sequence:6231	+	45	Vpu
AGTATCAGCACTTGTGGAGATGG	sequence:6237	+	45	Vpu
GTATCAGCACTTGTGGAGATGGG	sequence:6238	+	45	Vpu
TATCAGCACTTGTGGAGATGGGG	sequence:6239	+	45	Vpu
ATCAGCACTTGTGGAGATGGGGG	sequence:6240	+	50	Vpu
AGCACTTGTGGAGATGGGGGTGG	sequence:6243	+	60	Vpu
AAATGGGGCACCATGCTCCTTGG	sequence:6266	+	55	Vpu
AATGGGGCACCATGCTCCTTGGG	sequence:6267	+	55	Vpu
CATCAATATCCCAAGGAGCATGG	sequence:6276	-	45	Vpu
GGGAGCAGCAGGAAGCACTATGG	sequence:7783	+	60	RRE
GGAGCAGCAGGAAGCACTATGGG	sequence:7784	+	55	RRE
CGCAGCGTCAATGACGCTGACGG	sequence:7807	+	60	RRE
GTCAATGACGCTGACGGTACAGG	sequence:7813	+	55	RRE
GCAGCAGAACAATTTGCTGAGGG	sequence:7864	+	45	RRE
CAATTTGCTGAGGGCTATTGAGG	sequence:7873	+	45	RRE
ATCTGTTGCAACTCACAGTCTGG	sequence:7905	+	45	RRE
TCTGTTGCAACTCACAGTCTGGG	sequence:7906	+	45	RRE
CTGTTGCAACTCACAGTCTGGGG	sequence:7907	+	50	RRE
CTGGGGCATCAAACAGCTCCAGG	sequence:7924	+	60	RRE
ACAGCTCCAGGCAAGAATCCTGG	sequence:7936	+	55	RRE
CCAGGCAAGAATCCTGGCTGTGG	sequence:7942	+	60	RRE
TTAGGTATCTTCCACAGCCAGG	sequence:7954	-	45	RRE
GGCTGTGGAAAGATACCTAAAGG	sequence:7957	+	45	RRE
ACCCACCTCCCAATCCCGAGGGG	sequence:8369	+	65	3' Tat 3' Rev
TCCCCTCGGGATTGGGAGGTGGG	sequence:8370	-	65	3' Tat 3' Rev
GTCCCCTCGGGATTGGGAGGTGG	sequence:8371	-	70	3' Tat 3' Rev
TGTCGGGTCCCCTCGGGATTGGG	sequence:8377	-	65	3' Tat 3' Rev
CTGTCGGGTCCCCTCGGGATTGG	sequence:8378	-	70	3' Tat 3' Rev
CAATCCCGAGGGGACCCGACAGG	sequence:8379	+	70	3' Tat 3' Rev
TCTATTCTTCGGGCCTGTCGGG	sequence:8393	-	55	3' Tat 3' Rev
TTCTATTCTTCGGGCCTGTCGG	sequence:8394	-	50	3' Tat 3' Rev
CCCGAAGGAATAGAAGAAGAAGG	sequence:8402	+	45	3' Tat 3' Rev
GGATCCGTTCACTAATCGAATGG	sequence:8448	-	45	3' Rev
ACGGATCCTTAGCACTTATCTGG	sequence:8463	+	45	3' Rev
CGGATCCTTAGCACTTATCTGGG	sequence:8464	+	45	3' Rev
ATCGTCCCAGATAAGTGCTAAGG	sequence:8469	-	45	3' Rev
CATTATCTGGGACGATCTGCGG	sequence:8475	+	50	3' Rev
GGTGGTAGCTGAAGAGGCACAGG	sequence:8500	-	60	3' Rev
TCAAGCGGTGGTAGCTGAAGAGG	sequence:8506	-	55	3' Rev
AGAGTAAGTCTCTCAAGCGGTGG	sequence:8518	-	50	3' Rev
TCAAGAGTAAGTCTCTCAAGCGG	sequence:8521	-	40	3' Rev
CTTGATTGTAACGAGGATTGTGG	sequence:8539	+	40	3' Rev
AACGAGGATTGTGGAACCTTCTGG	sequence:8548	+	45	3' Rev
ACGAGGATTGTGGAACCTTCTGGG	sequence:8549	+	45	3' Rev
TTGTGGAACCTTCTGGGACGCAGG	sequence:8556	+	55	3' Rev
TGTGGAACCTTCTGGGACGCAGGG	sequence:8557	+	55	3' Rev
GTGGAACCTTCTGGGACGCAGGGG	sequence:8558	+	60	3' Rev
TGGAACCTTCTGGGACGCAGGGGG	sequence:8559	+	60	3' Rev
ACTTCTGGGACGCAGGGGGTGGG	sequence:8563	+	65	3' Rev
GGTGGGAAGCCCTCAAATATTGG	sequence:8580	+	50	3' Rev
GGAAGCCCTCAAATATTGGTGG	sequence:8583	+	50	3' Rev

GGAGATTCCACCAATATTTGAGG	sequence:8590	-	40	3' Rev
TCTCCTACAGTATTGGAGTCAGG	sequence:8608	+	45	3' Rev
GTTCCCTGACTCCAATACTGTAGG	sequence:8611	-	45	3' Rev
TGATTGGATGGCCTGCTGTAAGG	sequence:8815	+	50	Nef
GATTGGATGGCCTGCTGTAAGGG	sequence:8816	+	50	Nef
CTCATTCTTTCCCTTACAGCAGG	sequence:8826	-	45	Nef
CGAGCTGAGCCAGCAGCAGATGG	sequence:8850	+	65	Nef
GAGCTGAGCCAGCAGCAGATGGG	sequence:8851	+	60	Nef
AGCTGAGCCAGCAGCAGATGGGG	sequence:8852	+	60	Nef
TGAGCCAGCAGCAGATGGGGTGG	sequence:8855	+	65	Nef
GCTCCCACCCCATCTGCTGCTGG	sequence:8859	-	70	Nef
TCTCGAGACCTAGAAAAACATGG	sequence:8886	+	40	Nef
CTAACAATGCTGCTTGTGCCTGG	sequence:8935	+	50	Nef
TGCCTGGCTAGAAGCACAAAGAGG	sequence:8951	+	55	Nef
CTCCTTTGTGCTTCTAGCCAGG	sequence:8953	-	55	Nef
CTGGCTAGAAGCACAAAGAGGAGG	sequence:8954	+	55	Nef
GCACAAGAGGAGGAAGAGGTGGG	sequence:8964	+	55	Nef
GGGTTTTCCAGTCACACCTCAGG	sequence:8984	+	55	Nef
AAAGGTACCTGAGGTGTGACTGG	sequence:8991	-	50	Nef
ATTGGTCTTAAAGGTACCTGAGG	sequence:9000	-	40	Nef
AAAGAAAAGGGGGGACTGGAAGG	sequence:9060	+	50	Nef
AAGAAAAGGGGGGACTGGAAGGG	sequence:9061	+	50	Nef
GACAAGATATCCTTGATCTGTGG	sequence:9103	+	40	LTR U3 Nef
GAACTACACACCAGGGCCAGGGG	sequence:9161	+	60	LTR U3 Nef
GGATATCTGACCCCTGGCCCTGG	sequence:9171	-	65	LTR U3 Nef
GTCAGTGGATATCTGACCCCTGG	sequence:9177	-	55	LTR U3 Nef
GTCAGATATCCACTGACCTTTGG	sequence:9183	+	45	LTR U3 Nef
TAGCTTGTAGCACCATCCAAGG	sequence:9199	-	45	LTR U3 Nef
AGTACCAGTTGAGCCAGATAAGG	sequence:9221	+	45	LTR U3 Nef
TCTACCTTATCTGGCTCAACTGG	sequence:9225	-	45	LTR U3 Nef
TGAGCCAGATAAGGTAGAAGAGG	sequence:9230	+	45	LTR U3 Nef
TTGGCCTCTTCTACCTTATCTGG	sequence:9234	-	45	LTR U3 Nef
AAGGTAGAAGAGGCCAATAAAGG	sequence:9240	+	40	LTR U3 Nef
GCTGGTGTCTCTCCTTTATTGG	sequence:9253	-	45	LTR U3 Nef
GCTCACAGGGTGTAAACAAGCTGG	sequence:9271	-	55	LTR U3 Nef
TTACACCTGTGAGCCTGCATGG	sequence:9279	+	55	LTR U3 Nef
CCCTGTGAGCCTGCATGGAATGG	sequence:9284	+	60	LTR U3 Nef
TCCATTCCATGCAGGCTCACAGG	sequence:9285	-	55	LTR U3 Nef
CAGGGTCATCCATTCCATGCAGG	sequence:9293	-	55	LTR U3 Nef
ACTCTAACACTTCTCTCAGGG	sequence:9311	-	40	LTR U3 Nef
CTGAGAGAGAAGTGTTAGAGTGG	sequence:9313	+	45	LTR U3 Nef
AGAGAGAAGTGTTAGAGTGGAGG	sequence:9316	+	45	LTR U3 Nef
CCGCCTAGCATTTCATCACGTGG	sequence:9347	+	55	LTR U3 Nef
GGGCCACGTGATGAAATGCTAGG	sequence:9350	-	55	LTR U3 Nef
CGTGGCCCGAGAGCTGCATCCGG	sequence:9365	+	70	LTR U3 Nef
CAGCAGTTCTGAAGTACTCCGG	sequence:9384	-	45	LTR U3 Nef
CTGACATCGAGCTTGCTACAAGG	sequence:9404	+	50	LTR U3 Nef
TGACATCGAGCTTGCTACAAGGG	sequence:9405	+	45	LTR U3 Nef
TGCTACAAGGGACTTTCCGCTGG	sequence:9417	+	55	LTR U3
GCTACAAGGGACTTTCCGCTGGG	sequence:9418	+	55	LTR U3
GGGGACTTTCCAGGGAGGCGTGG	sequence:9438	+	70	LTR U3
CTTTCCAGGGAGGCGTGGCCTGG	sequence:9443	+	70	LTR U3
TTTCCAGGGAGGCGTGGCCTGGG	sequence:9444	+	65	LTR U3
TGCTTATATGCAGCATCTGAGGG	sequence:9487	-	40	LTR U3
CTGCTTATATGCAGCATCTGAGG	sequence:9488	-	45	LTR U3
CACACTACTTGAAGCACTCAAGG	sequence:9613	-	45	LTR U5

BK132 Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
CTGGTTTCTCTTTCGCTTTCAGG	sequence:4	-	45	
GAAAGCGAAAGAGAAACCAGAGG	sequence:7	+	45	
CTGCGTCGAGAGAGCTCCTCTGG	sequence:23	-	65	
CAGAGGAGCTCTCTCGACGCAGG	sequence:24	+	65	
AGCTCTCTCGACGCAGGACTCGG	sequence:30	+	60	
TCGGCTTGCTGAAGCGCGCACGG	sequence:49	+	65	
GCTGAAGCGCGCACGGCAAGAGG	sequence:56	+	70	
CCAAAAATTTTACTAGCGGAGG	sequence:104	+	40	
TTTACTAGCGGAGGCTAGAAGG	sequence:112	+	50	
GAGGCTAGAAGGAGAGATGGG	sequence:123	+	50	Gag
GCGAGAGCGTCAGTATTAAGCGG	sequence:147	+	50	Gag
CGAGAGCGTCAGTATTAAGCGGG	sequence:148	+	50	Gag
GAGAGCGTCAGTATTAAGCGGGG	sequence:149	+	50	Gag
AGAGCGTCAGTATTAAGCGGGGG	sequence:150	+	50	Gag
GCGGGGACAATTAGATAGATGG	sequence:166	+	50	Gag
CGGGGGACAATTAGATAGATGGG	sequence:167	+	45	Gag
GATGGGAGAAAATTCGGTTACGG	sequence:184	+	40	Gag
GAGAAAATTCGGTTACGGCCAGG	sequence:189	+	50	Gag
AGAAAATTCGGTTACGGCCAGGG	sequence:190	+	45	Gag
GAAAATTCGGTTACGGCCAGGGG	sequence:191	+	50	Gag
AAAATTCGGTTACGGCCAGGGGG	sequence:192	+	50	Gag
AACATATAGTATGGGCAAGCAGG	sequence:235	+	40	Gag
ACATATAGTATGGGCAAGCAGGG	sequence:236	+	40	Gag
GAACGATTCGAGTTAATCCTGG	sequence:264	+	45	Gag
CGCAGTTAATCCTGGCCTATTGG	sequence:272	+	50	Gag
TCTGATGTTTCCAATAGGCCAGG	sequence:282	-	45	Gag
GGCCTATTGGAAACATCAGAAGG	sequence:285	+	45	Gag
AGCCTTCTGATGTTTCCAATAGG	sequence:287	-	40	Gag
AGAAGGCTGTAGACAAATACTGG	sequence:302	+	40	Gag
GAAGGCTGTAGACAAATACTGGG	sequence:303	+	40	Gag
CTACAACCAAGCCTTCAGACAGG	sequence:330	+	50	Gag
TCTGATCCTGTCTGAAGGCTTGG	sequence:336	-	50	Gag
GTTCTTCTGATCCTGTCTGAAGG	sequence:341	-	45	Gag
GATAGAGGTAAGACACCAAGG	sequence:413	+	40	Gag
GGAAGCTTTAGAGAAGATAGAGG	sequence:434	+	40	Gag
TAGGGTAATTTGGCTAACCTGG	sequence:518	-	40	Gag
CCCTATAGTGCAAATATGCAGG	sequence:536	+	40	Gag
GCAAATATGCAGGGGCAAATGG	sequence:545	+	45	Gag
GCAGGGGCAAATGGTACATCAGG	sequence:554	+	55	Gag
GGTGTAAATCTTGTGGGGTGG	sequence:676	-	45	Gag
AAACACCATGCTAAACACAGTGG	sequence:692	+	40	Gag
AACACCATGCTAAACACAGTGGG	sequence:693	+	40	Gag
ACACCATGCTAAACACAGTGGGG	sequence:694	+	45	Gag
CACCATGCTAAACACAGTGGGGG	sequence:695	+	50	Gag
ACCATGCTAAACACAGTGGGGGG	sequence:696	+	50	Gag
TCCCCCACTGTGTTTAGCATGG	sequence:697	-	55	Gag
TTCTGCAGCTTCTCATTGATGG	sequence:751	-	45	Gag
TCAATGAGGAAGCTGCAGAATGG	sequence:754	+	45	Gag
CAATGAGGAAGCTGCAGAATGGG	sequence:755	+	45	Gag
AGATTGCATCCAGTGCATGCAGG	sequence:780	+	50	Gag
GATTGCATCCAGTGCATGCAGGG	sequence:781	+	50	Gag
GCAATAGGCCCTGCATGCACTGG	sequence:789	-	60	Gag
CATGCAGGGCCTATTGCACCAGG	sequence:795	+	60	Gag
CTCATCTGGCCTGGTGAATAGG	sequence:804	-	55	Gag
CTTGTTCTCTCATCTGGCCTGG	sequence:813	-	55	Gag
CAGGCCAGATGAGAGAACCAAGG	sequence:814	+	55	Gag
AGGCCAGATGAGAGAACCAAGGG	sequence:815	+	50	Gag
GGCCAGATGAGAGAACCAAGGGG	sequence:816	+	55	Gag
TTCCCTTGGTCTCTCATCTGG	sequence:818	-	50	Gag
CCAAGGGGAAGTGACATAGCAGG	sequence:831	+	55	Gag
AGGAACTACTAGTACCCTTCAGG	sequence:851	+	45	Gag
AGTACCCTCAGGAACAAATAGG	sequence:861	+	40	Gag
CCCTCAGGAACAAATAGGATGG	sequence:865	+	45	Gag
TCCATCCTATTTGTTCTGAAGG	sequence:866	-	40	Gag
ACATAATCCACCTATCCCAGTGG	sequence:893	+	45	Gag
CATAATCCACCTATCCCAGTGGG	sequence:894	+	45	Gag
ATTTCTCCCACTGGGATAGGTGG	sequence:900	-	50	Gag
TAAATTTCTCCCACTGGGATAGG	sequence:903	-	40	Gag
GTATAGCCCTACCAGCATTTTGG	sequence:968	+	45	Gag
GGACATAAGACAAGGACCAAAGG	sequence:989	+	45	Gag

TAGTCTCTAAAGGGTTCCTTTGG	sequence:1005	-	40	Gag
CGGTCTACATAGTCTCTAAAGGG	sequence:1014	-	40	Gag
CCTTTAGAGACTATGTAGACCGG	sequence:1015	+	40	Gag
AAGAGCCGAGCAAGCCTCACAGG	sequence:1052	+	60	Gag
AGCCGAGCAAGCCTCACAGGAGG	sequence:1055	+	65	Gag
TACCTCCTGTGAGGCTTGCTCGG	sequence:1057	-	55	Gag
CCTCACAGGAGGTAATAAATTGG	sequence:1066	+	40	Gag
TTGGATGACAGAAACCTTGTTGG	sequence:1085	+	40	Gag
GTTCGCATTTTGGACCAACAAGG	sequence:1099	-	45	Gag
TACAATCTGGGTTTCGCATTTTGG	sequence:1109	-	40	Gag
ATTTCTTAGTGTAGCTGCTGG	sequence:1155	-	40	Gag
AGAAATGATGACAGCATGTCAGG	sequence:1172	+	40	Gag
GAAATGATGACAGCATGTCAGGG	sequence:1173	+	40	Gag
GATGACAGCATGTCAGGGAGTGG	sequence:1178	+	55	Gag
ATGACAGCATGTCAGGGAGTGGG	sequence:1179	+	50	Gag
ACAGCATGTCAGGGAGTGGGAGG	sequence:1182	+	60	Gag
TGTCAGGGAGTGGGAGGTCCCGG	sequence:1188	+	65	Gag
AGTGGGAGGTCCCGGCCATAAGG	sequence:1196	+	65	Gag
AAAACCTTGCCTTATGGCCGGG	sequence:1206	-	45	Gag
CAAACTCTTGCCTTATGGCCGG	sequence:1207	-	45	Gag
CGGCCATAAGGCAAGAGTTTTGG	sequence:1208	+	50	Gag
CCATAAGGCAAGAGTTTTGGCGG	sequence:1211	+	45	Gag
TGGTAGCTGAATTTGTCACCTTGG	sequence:1244	-	40	Gag
GCTACCATAATGATGCAGAAAGG	sequence:1260	+	40	Gag
GGCACATAGCCCGAAATTGCAGG	sequence:1336	+	55	Gag
GCACATAGCCCGAAATTGCAGGG	sequence:1337	+	50	Gag
CCCGAAATTGCAGGGCCCCTAGG	sequence:1345	+	65	Gag
TCCTAGGGGCCCTGCAATTTCCG	sequence:1346	-	55	Gag
TTGCAGGGCCCCTAGGAAGAAGG	sequence:1352	+	60	Gag
TGCAGGGCCCCTAGGAAGAAGGG	sequence:1353	+	60	Gag
TCCAACAGCCCTTCTCCTAGGG	sequence:1361	-	50	Gag
CTCCAACAGCCCTTCTCCTAGG	sequence:1362	-	55	Gag
GGGCTGTTGGAGATGTGGAAAGG	sequence:1373	+	55	Gag
GAAAATCTGGCCTTCCCACAAGG	sequence:1445	+	50	Gag Pol
AAAATCTGGCCTTCCCACAAGGG	sequence:1446	+	45	Gag Pol
TCTGGCCTTCCCACAAGGGAAGG	sequence:1450	+	60	Gag Pol
CCTTCCCACAAGGGAAGGCCAGG	sequence:1455	+	65	Gag Pol
CTTCCCACAAGGGAAGGCCAGGG	sequence:1456	+	60	Gag Pol
AATCCCTGGCCTTCCCTTGTGG	sequence:1460	-	55	Gag Pol
CTGCTCTGAAGAAATTCCTGG	sequence:1473	-	45	Gag Pol
CTCTCTGCTGGTGGGGCTGTCGG	sequence:1503	-	65	Gag Pol
CCCACCAGCAGAGAGCTTCAGG	sequence:1510	+	65	Gag Pol
CCAGCAGAGAGCTTCAGGTTTGG	sequence:1515	+	55	Gag Pol
CAGCAGAGAGCTTCAGGTTTGGG	sequence:1516	+	50	Gag Pol
AGCAGAGAGCTTCAGGTTTGGGG	sequence:1517	+	50	Gag Pol
AACAACCTCCTCTCAGAAGCAGG	sequence:1547	+	50	Gag Pol
ATTGTCTCCTGCTTCTGAGAGGG	sequence:1554	-	45	Gag Pol
TATTGTCTCCTGCTTCTGAGAGG	sequence:1555	-	45	Gag Pol
GAAGCAGGAGACAATAGACAAGG	sequence:1562	+	45	Gag Pol
AGTGATTTGAGGGCAGTTAAAGG	sequence:1593	-	40	Gag Pol
ACTGCCCTCAAATCACTCTTTGG	sequence:1599	+	45	Gag Pol
GTTGCCAAAGAGTGATTTGAGGG	sequence:1603	-	40	Gag Pol
CCCCTCGTCACAATAAAGATAGG	sequence:1628	+	45	Gag Pol
CCCTCGTCACAATAAAGATAGGG	sequence:1629	+	40	Gag Pol
CCTCGTCACAATAAAGATAGGGG	sequence:1630	+	40	Gag Pol
CTCGTCACAATAAAGATAGGGGG	sequence:1631	+	40	Gag Pol
TCGTACAATAAAGATAGGGGGG	sequence:1632	+	40	Pol
AAAGATAGGGGGGCAACTAAAGG	sequence:1642	+	45	Pol
GGAAGCTCTATTAGATACAGGGG	sequence:1663	+	40	Pol
CCAAAAATGATAGGGGGAATTGG	sequence:1733	+	40	Pol
AAAATGATAGGGGGAATTGGAGG	sequence:1736	+	40	Pol
CCTACACCTGTCAACATAATTGG	sequence:1838	+	40	Pol
ACTTTTGGGCCATCCATTCCTGG	sequence:1940	-	50	Pol
GACTTCTGGGAAGTTCAATTAGG	sequence:2156	+	40	Pol
TTAGGAATACCACATCCCGCAGG	sequence:2174	+	50	Pol
TAGGAATACCACATCCCGCAGGG	sequence:2175	+	50	Pol
TTTTTCAACCCTGCGGGATGTGG	sequence:2183	-	50	Pol
ATCAGTAACGGTACTGGATGTGG	sequence:2212	+	45	Pol
TCAGTAACGGTACTGGATGTGGG	sequence:2213	+	45	Pol
CTCATTGTTTGTACTCGGTATGG	sequence:2292	-	40	Pol
GGTGTCTCATTGTTTGTACTCGG	sequence:2297	-	40	Pol

AGTACAAACAATGAGACACCAGG	sequence:2300	+	40	Pol
GTACAAACAATGAGACACCAGGG	sequence:2301	+	40	Pol
TCAGTACAATGTGCTTCCACAGG	sequence:2332	+	45	Pol
CAGTACAATGTGCTTCCACAGGG	sequence:2333	+	45	Pol
ACAATGTGCTTCCACAGGGATGG	sequence:2337	+	50	Pol
GTGCTTCCACAGGGATGGAAAGG	sequence:2342	+	55	Pol
GGTGATCCTTCCATCCCTGTGG	sequence:2348	-	55	Pol
AACTGAGACAACATCTGTTGAGG	sequence:2511	+	40	Pol
TGAGACAACATCTGTTGAGGTGG	sequence:2514	+	45	Pol
GAGACAACATCTGTTGAGGTGGG	sequence:2515	+	45	Pol
AGACAACATCTGTTGAGGTGGGG	sequence:2516	+	45	Pol
AGAAAGAACCTCCATTCTTTGG	sequence:2565	+	40	Pol
AGAACCTCCATTCTTTGGATGG	sequence:2569	+	45	Pol
GAACCTCCATTCTTTGGATGGG	sequence:2570	+	45	Pol
TCATAACCCATCCAAAGGAATGG	sequence:2576	-	40	Pol
GGAGTTCATAACCCATCCAAAGG	sequence:2581	-	45	Pol
GTAAGTCCATTTATCAGGATGG	sequence:2602	-	40	Pol
GGTTGTACTGTCCATTTATCAGG	sequence:2606	-	40	Pol
AGTACAACCTATAGTGCTGCCGG	sequence:2620	+	45	Pol
TCTTTTCCGGCAGCACTATAGG	sequence:2627	-	45	Pol
TGCTGCCGGAAAAAGACAGCTGG	sequence:2634	+	55	Pol
ACAGTCCAGCTGTCTTTTCCGG	sequence:2639	-	45	Pol
TGGGCAAGTCAGATTTACCCAGG	sequence:2696	+	50	Pol
GGGCAAGTCAGATTTACCCAGGG	sequence:2697	+	50	Pol
TTAGTGCTTTGGTTCCCTAAGG	sequence:2746	-	45	Pol
AGCTCTGCCTCTTTGTTAGTGG	sequence:2780	-	45	Pol
AAAAGAGGCAGAGCTAGAACTGG	sequence:2788	+	45	Pol
AGCTAGAACTGGCAGAAAACAGG	sequence:2799	+	45	Pol
GCTAGAACTGGCAGAAAACAGGG	sequence:2800	+	45	Pol
TAGCAGAAATACAGAAGCAGGGG	sequence:2877	+	40	Pol
GAAATACAGAAGCAGGGGCAAGG	sequence:2882	+	50	Pol
AGAAGCAGGGGCAAGGCCAATGG	sequence:2889	+	60	Pol
AAGAAACATGGGAAACATGGTGG	sequence:3084	+	40	Pol
AAACATGGTGGACAGAGTATTGG	sequence:3096	+	40	Pol
CAGAGTATTGGCAAGCCACCTGG	sequence:3108	+	55	Pol
AAGCCACCTGGATTCTGAGTGG	sequence:3120	+	55	Pol
AGCCACCTGGATTCTGAGTGGG	sequence:3121	+	55	Pol
AACTCCCACTCAGGAATCCAGG	sequence:3126	-	50	Pol
GTATTGACAAACTCCCACTCAGG	sequence:3134	-	45	Pol
CCCCTCCCTTAGTGAAATTATGG	sequence:3156	+	45	Pol
CTATGGGCTCTTTTCTAACTGG	sequence:3181	-	40	Pol
AAGGTTTCTGCTCCTACTATGGG	sequence:3197	-	40	Pol
GAAGGTTTCTGCTCCTACTATGG	sequence:3198	-	45	Pol
GCAGAAACCTTCTATGTAGATGG	sequence:3209	+	40	Pol
AACCTTCTATGTAGATGGGGCGG	sequence:3214	+	45	Pol
AGCCGCCCATCTACATAGAAGG	sequence:3216	-	55	Pol
ATGTAGATGGGGCGGCTAACAGG	sequence:3222	+	55	Pol
TGTAGATGGGGCGGCTAACAGGG	sequence:3223	+	55	Pol
GCTAACAGGGAGACTAAATTAGG	sequence:3236	+	40	Pol
GCAGGATATGTGACTAACAGAGG	sequence:3263	+	45	Pol
AGCAATTCACCTAGCTTTGCAGG	sequence:3340	+	45	Pol
TCACCTAGCTTTGCAGGATTCGG	sequence:3346	+	45	Pol
CACCTAGCTTTGCAGGATTCGGG	sequence:3347	+	50	Pol
ATCCCGAATCCTGCAAAGCTAGG	sequence:3349	-	50	Pol
ACAAAAGGAAAAGGTCTACCTGG	sequence:3478	+	40	Pol
AGGAAAAGGTCTACCTGGCATGG	sequence:3483	+	50	Pol
GGAAAAGGTCTACCTGGCATGGG	sequence:3484	+	50	Pol
TGTGTGCTGGTACCCATGCCAGG	sequence:3496	-	60	Pol
GGCATGGGTACCAGCACACAAGG	sequence:3499	+	60	Pol
GCATGGGTACCAGCACACAAGGG	sequence:3500	+	55	Pol
GTACCAGCACACAAGGGAATTGG	sequence:3506	+	50	Pol
CCAGCACACAAGGGAATTGGAGG	sequence:3509	+	55	Pol
AATTAGTCAGTGCTGGAATCAGG	sequence:3549	+	40	Pol
TCACAGTAATTGGAGAGCAATGG	sequence:3625	+	40	Pol
CTTTTGCTACTACAGGTGGCAGG	sequence:3661	-	50	Pol
TAGCTGACATTTATCACAGCTGG	sequence:3693	-	40	Pol
CTAAAAGGAGAAGCCATGCATGG	sequence:3713	+	45	Pol
AGGAGAAGCCATGCATGGACAGG	sequence:3718	+	55	Pol
ACAGTCTACCTGTCCATGCATGG	sequence:3726	-	50	Pol
GGACAGGTAGACTGTAGTCCAGG	sequence:3734	+	55	Pol
TAGACTGTAGTCCAGGAATATGG	sequence:3741	+	40	Pol

CAATCTAGTTGCCATATTCCTGG	sequence:3752	-	40	Pol
TGGCTACATGAACTGCTACCAGG	sequence:3799	-	50	Pol
GTAGCAGTTCATGTAGCCAGTGG	sequence:3803	+	50	Pol
GAAGTTATTCCAGCAGAGACAGG	sequence:3839	+	45	Pol
AAGTTATTCCAGCAGAGACAGGG	sequence:3840	+	40	Pol
TATTCCAGCAGAGACAGGGCAGG	sequence:3844	+	55	Pol
GTTTCCTGCCCTGTCTCTGCTGG	sequence:3848	-	60	Pol
CAGCAATTCACCACTACTACGG	sequence:3934	+	40	Pol
TTTACCAGTACTACGGTTAAGG	sequence:3940	+	40	Pol
GGCAGCCTTAACCGTAGTACTGG	sequence:3945	-	55	Pol
CTACGGTTAAGGCTGCCTGTTGG	sequence:3951	+	55	Pol
CGGTTAAGGCTGCCTGTTGGTGG	sequence:3954	+	60	Pol
GGTTAAGGCTGCCTGTTGGTGGG	sequence:3955	+	55	Pol
TAAGGCTGCCTGTTGGTGGGCGG	sequence:3958	+	60	Pol
AAGGCTGCCTGTTGGTGGGCGGG	sequence:3959	+	65	Pol
AGGCTGCCTGTTGGTGGGCGGGG	sequence:3960	+	70	Pol
CTTGATCCCCGCCACCAACAGG	sequence:3966	-	65	Pol
TTGGTGGGCGGGGATCAAGCAGG	sequence:3970	+	65	Pol
CTTGACTTTGGGGATTGTAGGGG	sequence:4003	-	45	Pol
CCCTACAATCCCCAAAGTCAAGG	sequence:4004	+	50	Pol
TCCTTGACTTTGGGGATTGTAGG	sequence:4005	-	45	Pol
TCTACTACTCCTTGACTTTGGGG	sequence:4013	-	40	Pol
TATAGGACAGGTAAGAGATCAGG	sequence:4063	+	40	Pol
TAAAAGAAAAGGGGGGATTGGGG	sequence:4135	+	40	Pol cPPT
AAAAGAAAAGGGGGGATTGGGGG	sequence:4136	+	45	Pol cPPT
AAAGAAAAGGGGGGATTGGGGGG	sequence:4137	+	50	Pol cPPT
GGGATTGGGGGGTACAGTGCAGG	sequence:4148	+	65	Pol cPPT
GGATTGGGGGGTACAGTGCAGGG	sequence:4149	+	60	Pol cPPT
GATTGGGGGGTACAGTGCAGGGG	sequence:4150	+	60	Pol cPPT
GGGACAGCAGAGATCCACTTTGG	sequence:4263	+	55	Pol
AGCAGAGATCCACTTTGGAAAGG	sequence:4268	+	45	Pol
TTTGCTGGTCCTTTCCAAAGTGG	sequence:4277	-	45	Pol
AAGGACCAGCAAAGCTTCTGTGG	sequence:4287	+	50	Pol
CCAGCAAAGCTTCTGTGGAAAGG	sequence:4292	+	50	Pol
AGCTTCTGTGGAAAGGTGAAGGG	sequence:4299	+	45	Pol
GCTTCTGTGGAAAGGTGAAGGGG	sequence:4300	+	50	Pol
GATTATGGAAAACAGATGGCAGG	sequence:4388	+	40	Pol Vif
TGATTGTGTGGCAAGTAGACAGG	sequence:4414	+	45	Pol Vif
TGTGGCAAGTAGACAGGATGAGG	sequence:4420	+	50	Pol Vif
TCAGAAGTACATATCCCCTAGG	sequence:4548	+	40	Vif
AGTACATATCCCCTAGGAGAGG	sequence:4553	+	45	Vif
CCCCTAGGAGAGGCTAGCTTGG	sequence:4562	+	60	Vif
ACCAAGCTAGCCTCTCCTAGTGG	sequence:4563	-	55	Vif
ACATATTGGGGTCTGCATACAGG	sequence:4593	+	45	Vif
TGCATACAGGAGAAAGAGAATGG	sequence:4606	+	40	Vif
AGGAGAAAGAGAATGGCATTGG	sequence:4613	+	40	Vif
AGAGAATGGCATTGGGTCAAGG	sequence:4620	+	45	Vif
GTCAAGGAGTCTCCATAGAATGG	sequence:4636	+	45	Vif
AAGGAGTCTCCATAGAATGGAGG	sequence:4639	+	45	Vif
GTCTCCATAGAATGGAGGAAAGG	sequence:4644	+	45	Vif
TCTTCCTTTCTCCATTCTATGG	sequence:4648	-	40	Vif
TATAACACACAAGTAGACCCAGG	sequence:4671	+	40	Vif
TTAGTTGGTCTGCTAGGCCTGGG	sequence:4688	-	50	Vif
ATTAGTTGGTCTGCTAGGCCTGG	sequence:4689	-	50	Vif
GATGAATTAGTTGGTCTGCTAGG	sequence:4694	-	40	Vif
CCTAGTTGTGAATATCAAGCAGG	sequence:4782	+	40	Vif
CAAGCAGGACATAACAAGGTAGG	sequence:4797	+	45	Vif
GGTAGGATCTCTACAGTACTTGG	sequence:4814	+	45	Vif
AGCCACCTTTGCCTAGTGTTAGG	sequence:4870	+	50	Vif
TTCCTAACACTAGGCAAAGGTGG	sequence:4872	-	45	Vif
AGTTTCCTAACACTAGGCAAAGG	sequence:4875	-	40	Vif
TCTGTCAAGTTTCCTAACACTAGG	sequence:4881	-	40	Vif
GGAAACTGACAGAGGATAGATGG	sequence:4891	+	45	Vif
GAACAAGCCCCAGAAGACCAAGG	sequence:4913	+	55	Vif Vpr
AACAAGCCCCAGAAGACCAAGGG	sequence:4914	+	50	Vif Vpr
CTGTGGCCCTTGGTCTTCTGGGG	sequence:4920	-	60	Vif Vpr
TCTGTGGCCCTTGGTCTTCTGGG	sequence:4921	-	55	Vif Vpr
CTCTGTGGCCCTTGGTCTTCTGG	sequence:4922	-	60	Vif Vpr
ATGGCTCTCTGTGGCCCTTGG	sequence:4930	-	60	Vif Vpr
TCATTGTATGGCTCTCTGTGG	sequence:4937	-	45	Vif Vpr
AGAGAGAGCCATACAATGAATGG	sequence:4941	+	40	Vif Vpr

ATGGACACTAGAGCTTTTAGAGG	sequence:4960	+	40	Vif Vpr
AAGCTGTTAGACACTTTCCTAGG	sequence:4995	+	40	Vpr
TTAGACACTTTCCTAGGACATGG	sequence:5001	+	40	Vpr
TTTCTAGGACATGGCTCCACGG	sequence:5009	+	50	Vpr
AATCCGTGGAGCCATGTCCTAGG	sequence:5012	-	55	Vpr
AGGACATGGCTCCACGGATTAGG	sequence:5015	+	55	Vpr
GGACATGGCTCCACGGATTAGGG	sequence:5016	+	55	Vpr
AGATATATTGCCCTAATCCGTGG	sequence:5026	-	40	Vpr
AATTATGGGGACACTTGGGCAGG	sequence:5054	+	50	Vpr
TGGGGACACTTGGGCAGGAGTGG	sequence:5059	+	65	Vpr
GGACACTTGGGCAGGAGTGGAGG	sequence:5062	+	65	Vpr
GGGTGTGCACATAGCAGAATAGG	sequence:5132	+	50	Vpr
CTACAAAGGAGAGCAAGAAATGG	sequence:5162	+	40	Vpr
GGCTCTAGTCTAGGATCTACTGG	sequence:5187	-	50	Vpr 5' Tat
TAGATCCTAGACTAGAGCCCTGG	sequence:5191	+	50	5' Tat
TGCTTCCAGGGCTCTAGTCTAGG	sequence:5196	-	55	5' Tat
CTAGAGCCCTGGAAGCATCCAGG	sequence:5202	+	60	5' Tat
TGACTTCTGGATGCTTCCAGGG	sequence:5208	-	50	5' Tat
CTGACTTCTGGATGCTTCCAGG	sequence:5209	-	55	5' Tat
TCCAGGAAGTCAGCTAAGACGG	sequence:5219	+	50	5' Tat
GCCGTCTTAGGCTGACTTCTGG	sequence:5220	-	60	5' Tat
CAAGTGTTACAAGCCGTCTTAGG	sequence:5232	-	45	5' Tat
AAAGCCTTAGGCATCTCCTATGG	sequence:5301	+	45	5' Tat
CCTTAGGCATCTCCTATGGCAGG	sequence:5305	+	55	5' Tat 5' Rev
TCTCCTATGGCAGGAAGAAGCGG	sequence:5314	+	50	5' Tat 5' Rev
TCTCCGCTTCTTCTGCCATAGG	sequence:5317	-	55	5' Tat 5' Rev
ACAGCGACGAAGAGCTCCTCAGG	sequence:5339	+	60	5' Tat 5' Rev
TGATGAGTCTGACGGTCTGAGG	sequence:5355	-	55	5' Tat 5' Rev
GAGAAGCTTGATGAGTCTGACGG	sequence:5363	-	45	5' Tat 5' Rev
ATAGAAAGAGCAGAAGACAGTGG	sequence:5550	+	40	Vpu
GACAGTGGAATGAGAGTGAAGG	sequence:5565	+	50	Vpu
CAATGAGAGTGAAGGAGATCAGG	sequence:5573	+	45	Vpu
GGAAGAATTGTCAGCACTTGTGG	sequence:5594	+	45	Vpu
ATTGTCAGCACTTGTGGAGATGG	sequence:5600	+	45	Vpu
TTGTCAGCACTTGTGGAGATGGG	sequence:5601	+	45	Vpu
TGTCAGCACTTGTGGAGATGGGG	sequence:5602	+	50	Vpu
AGATGGGGCATCTTGCTCCTTGG	sequence:5617	+	55	Vpu
GATGGGGCATCTTGCTCCTTGGG	sequence:5618	+	55	Vpu
GGGAGCAGCAGGAAGCACTATGG	sequence:7121	+	60	RRE
GGAGCAGCAGGAAGCACTATGGG	sequence:7122	+	55	RRE
CGCAGCGTCAGTGACGCTGACGG	sequence:7145	+	65	RRE
GTCAGTGACGCTGACGGTACAGG	sequence:7151	+	60	RRE
CAGGCCAGACTATTATTGTCTGG	sequence:7170	+	45	RRE
AACAGCAGAACAATCTGCTGAGG	sequence:7201	+	45	RRE
ACAGCAGAACAATCTGCTGAGGG	sequence:7202	+	45	RRE
CAATCTGCTGAGGGCTATTGAGG	sequence:7211	+	50	RRE
ATCTGTTGCAACTCACAGTCTGG	sequence:7243	+	45	RRE
TCTGTTGCAACTCACAGTCTGGG	sequence:7244	+	45	RRE
CTGTTGCAACTCACAGTCTGGGG	sequence:7245	+	50	RRE
CTGGGGCATCAAGCAGCTCCAGG	sequence:7262	+	65	RRE
CCAGGCAAGAGTCTTAGCTGTGG	sequence:7280	+	55	RRE
AGCTGTGGAAAGATACCTAAAGG	sequence:7295	+	40	RRE
TCGATTCTCCGGGCTGTCTGGG	sequence:7731	-	65	3' Tat 3' Rev
TTCGATTCTCCGGGCTGTCTGG	sequence:7732	-	60	3' Tat 3' Rev
CCCGGAGGAATCGAAGAAGAAGG	sequence:7740	+	55	3' Tat 3' Rev
ACCTTCTTCTCGATTCTCCGG	sequence:7741	-	45	3' Tat 3' Rev
GAGAGAGACAGAGACACATCCGG	sequence:7767	+	50	3' Tat 3' Rev
AGACACATCCGGACGCTTAGTGG	sequence:7778	+	55	3' Tat 3' Rev
ACATCCGGACGCTTAGTGGATGG	sequence:7782	+	55	3' Rev
GAATCCATCCACTAAGCGTCCGG	sequence:7786	-	50	3' Rev
CAATTTTCTGGGTCGATCTGCGG	sequence:7813	+	45	3' Rev
GGTGGTAGCTGAAGAGGCACAGG	sequence:7838	-	60	3' Rev
TCAAGCGGTGGTAGCTGAAGAGG	sequence:7844	-	55	3' Rev
AGAGTAAGTCTCTCAAGCGGTGG	sequence:7856	-	50	3' Rev
TCAAGAGTAAGTCTCTCAAGCGG	sequence:7859	-	40	3' Rev
CTTGATTGTAACGAGGATTGTGG	sequence:7877	+	40	3' Rev
AACGAGGATTGTGGAATTCTGG	sequence:7886	+	45	3' Rev
ACGAGGATTGTGGAATTCTGGG	sequence:7887	+	45	3' Rev
TTGTGGAATTCTGGGACGCAGG	sequence:7894	+	55	3' Rev
TGTGGAATTCTGGGACGCAGGG	sequence:7895	+	55	3' Rev
GTGGAATTCTGGGACGCAGGGG	sequence:7896	+	60	3' Rev

TGGAACTTCTGGGACGCAGGGGG	sequence:7897	+	60	3' Rev
ACTTCTGGGACGCAGGGGGTGGG	sequence:7901	+	65	3' Rev
GGTGGGAAGCCCTCAAATATTGG	sequence:7918	+	50	3' Rev
GGGAAGCCCTCAAATATTGGTGG	sequence:7921	+	50	3' Rev
GGAGATTCCACCAATATTTGAGG	sequence:7928	-	40	3' Rev
TCTCCTACAGTATTGGAGTCAGG	sequence:7946	+	45	3' Rev
GTTCTGACTCCAATACTGTAGG	sequence:7949	-	45	3' Rev
AGCAAGTGGTCAAAAATGAGTGG	sequence:8131	+	40	Nef
AGTGGTCAAAAATGAGTGGGTGG	sequence:8135	+	45	Nef
TGAGTGGGTGGCCTGCTGTAAGG	sequence:8147	+	60	Nef
GAGTGGGTGGCCTGCTGTAAGGG	sequence:8148	+	60	Nef
CTCATTCTTTCCCTTACAGCAGG	sequence:8158	-	45	Nef
TGAGCCAGCAGCAGATGGAGTGG	sequence:8196	+	60	Nef
GAGCCAGCAGCAGATGGAGTGGG	sequence:8197	+	60	Nef
GCTCCCACTCCATCTGCTGCTGG	sequence:8200	-	65	Nef
GGGAGCAGCATCTAGAGACCTGG	sequence:8217	+	60	Nef
TCTAGAGACCTGGAAAAACATGG	sequence:8227	+	40	Nef
TGATTGCTCCATGTTTTTCCAGG	sequence:8235	-	40	Nef
CTACCAATGCTGATTGTGCCTGG	sequence:8276	+	50	Nef
TAGCCAGGCACAATCAGCATTGG	sequence:8279	-	50	Nef
TGCCTGGCTAGAAGCACAAAGAGG	sequence:8292	+	55	Nef
CTCCTCTTGTGCTTCTAGCCAGG	sequence:8294	-	55	Nef
CTGGCTAGAAGCACAAAGAGGAGG	sequence:8295	+	55	Nef
GCTAGAAGCACAAAGAGGAGGAGG	sequence:8298	+	55	Nef
AAGGTACCTGAGGTTTACTGGG	sequence:8334	-	45	Nef
ATTGGTCTTAAAGGTACCTGAGG	sequence:8344	-	40	Nef
AGAGCTCCCTTGTAAGTCATTGG	sequence:8362	-	45	Nef
AAAGAAAAGGGGGGACTGGAAGG	sequence:8404	+	50	Nef
AAGAAAAGGGGGGACTGGAAGGG	sequence:8405	+	50	Nef
GACAAGATATCCTTGATCTGTGG	sequence:8447	+	40	Nef LTR U3
TATGGTAGACCCACAGATCAAGG	sequence:8457	-	45	Nef LTR U3
CTGTGGGTCTACCATACACAAGG	sequence:8464	+	50	Nef LTR U3
CAGGGAAGTAGCCTTGTGTATGG	sequence:8475	-	50	Nef LTR U3
CACAAGGCTACTTCCCTGATTGG	sequence:8480	+	50	Nef LTR U3
GTGTGTAATTCTGCCAATCAGGG	sequence:8493	-	40	Nef LTR U3
GGTGTGTAATTCTGCCAATCAGG	sequence:8494	-	45	Nef LTR U3
GATTGGCAGAATTACACACCAGG	sequence:8497	+	45	Nef LTR U3
ATTGGCAGAATTACACACCAGGG	sequence:8498	+	40	Nef LTR U3
CAGAATTACACACCAGGGCCAGG	sequence:8503	+	55	Nef LTR U3
AGAATTACACACCAGGGCCAGGG	sequence:8504	+	50	Nef LTR U3
GAATTACACACCAGGGCCAGGGG	sequence:8505	+	55	Nef LTR U3
GGGAATCTGACCCCTGGCCCTGG	sequence:8515	-	70	Nef LTR U3
GTCAGTGGGAATCTGACCCCTGG	sequence:8521	-	60	Nef LTR U3
GTCAGATTCCCACTGACCTTTGG	sequence:8527	+	50	Nef LTR U3
GATTCCCACTGACCTTTGGATGG	sequence:8531	+	50	Nef LTR U3
AGCACCATCCAAAGGTCAGTGGG	sequence:8535	-	50	Nef LTR U3
AAGCACCATCCAAAGGTCAGTGG	sequence:8536	-	50	Nef LTR U3
TAGCTTGAAGCACCATCCAAAGG	sequence:8543	-	45	Nef LTR U3
AGTACCAGTTGAGCCAGACAAGG	sequence:8565	+	50	Nef LTR U3
TCTACCTTGTCTGGCTCAACTGG	sequence:8569	-	50	Nef LTR U3
TGAGCCAGACAAGGTAGAAGAGG	sequence:8574	+	50	Nef LTR U3
TTGGCCTTCTACCTTGTCTGG	sequence:8578	-	50	Nef LTR U3
AAGGTAGAAGAGGCCAATGAAGG	sequence:8584	+	45	Nef LTR U3
AGGTAGAAGAGGCCAATGAAGGG	sequence:8585	+	45	Nef LTR U3
GGTAGAAGAGGCCAATGAAGGGG	sequence:8586	+	50	Nef LTR U3
GCAGTTGTTTTCCCTTCATTGG	sequence:8597	-	45	Nef LTR U3
TTACACCCTATGAGCCAGCATGG	sequence:8623	+	50	Nef LTR U3
TACACCCTATGAGCCAGCATGGG	sequence:8624	+	50	Nef LTR U3
CCCTATGAGCCAGCATGGGATGG	sequence:8628	+	60	Nef LTR U3
TCCATCCCATGCTGGCTCATAGG	sequence:8629	-	55	Nef LTR U3
CCAGCATGGGATGGAAGACCCGG	sequence:8637	+	60	Nef LTR U3
GTGGAAGTTTGACAGCCGCCTGG	sequence:8676	+	60	Nef LTR U3
CCGCCTGGCATTGCATCACATGG	sequence:8691	+	60	Nef LTR U3
GGGCCATGTGATGCAATGCCAGG	sequence:8694	-	60	Nef LTR U3
CATGGCCCCGAGAGAAGCATCCGG	sequence:8709	+	60	Nef LTR U3
GTA CTCCGGATGCTTCTCTCGGG	sequence:8714	-	55	Nef LTR U3
AGTACTCCGGATGCTTCTCTCGG	sequence:8715	-	50	Nef LTR U3
GAAGCATCCGGAGTACTACAAGG	sequence:8721	+	50	Nef LTR U3
CAGCAGTCCCTGTAGTACTCCGG	sequence:8728	-	50	Nef LTR U3
CTGACATCGAGCTTCTACAAGG	sequence:8748	+	45	LTR U3
TGACATCGAGCTTCTACAAGGG	sequence:8749	+	40	LTR U3

TTCTACAAGGGACTTTCCGCTGG	sequence:8761	+	50	LTR U3
TCTACAAGGGACTTTCCGCTGGG	sequence:8762	+	50	LTR U3
CTACAAGGGACTTTCCGCTGGGG	sequence:8763	+	55	LTR U3
CTTTCCGCTGGGGACTTTCCAGG	sequence:8773	+	60	LTR U3
TTTCCGCTGGGGACTTTCCAGGG	sequence:8774	+	55	LTR U3
CCGCTGGGGACTTTCCAGGGAGG	sequence:8777	+	70	LTR U3
GGGGACTTTCCAGGGAGGCGTGG	sequence:8782	+	70	LTR U3
CTTTCCAGGGAGGCGTGGCCTGG	sequence:8787	+	70	LTR U3
TTTCCAGGGAGGCGTGGCCTGGG	sequence:8788	+	65	LTR U3
TGCTTATATGCAGCATCTGAGGG	sequence:8831	-	40	LTR U3
CTGCTTATATGCAGCATCTGAGG	sequence:8832	-	45	LTR U3
CAGCTGCTTTTTGCCTGTACTGG	sequence:8852	+	50	LTR U3
AGCTGCTTTTTGCCTGTACTGGG	sequence:8853	+	45	LTR U3
TTGCCTGTACTGGGTCTCTCTGG	sequence:8862	+	55	LTR U3
TAACCAGAGAGACCCAGTACAGG	sequence:8865	-	50	LTR U3
GGTTAGACCAGATCAGAGCCTGG	sequence:8883	+	55	LTR R
GTTAGACCAGATCAGAGCCTGGG	sequence:8884	+	50	LTR R
AGAGCTCCCAGGCTCTGATCTGG	sequence:8890	-	60	LTR R
ATCAGAGCCTGGGAGCTCTCTGG	sequence:8894	+	60	LTR R
TAGTTAGCCAGAGAGCTCCCAGG	sequence:8901	-	55	LTR R
TGGGAGCTCTCTGGCTAACTAGG	sequence:8903	+	55	LTR R
GGGAGCTCTCTGGCTAACTAGGG	sequence:8904	+	55	LTR R
CTTTATTGAGGCTTAAGCAGTGG	sequence:8930	-	40	LTR R
ACTCAAGGCAAGCTTTATTGAGG	sequence:8942	-	40	LTR U5
CACACTACTTGAAGCACTCAAGG	sequence:8957	-	45	LTR U5

HIV-1 671 Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
CAGCTAGCCAGAGAGCTCCCAGG	sequence:4	-	65	LTR R
CTGGGAGCTCTCTGGCTAGCTGG	sequence:5	+	65	LTR R
TGGGAGCTCTCTGGCTAGCTGGG	sequence:6	+	60	LTR R
GGGAGCTCTCTGGCTAGCTGGGG	sequence:7	+	65	LTR R
CTTTATTGAGGCTTAAGCAGTGG	sequence:33	-	40	LTR R
ACTCAAGGCAAGCTTTATTGAGG	sequence:45	-	40	LTR R
CACACTACTTAAAGCACTCAAGG	sequence:60	-	40	LTR U5
GCCCGTCTGTTATGTGACTCTGG	sequence:84	+	55	LTR U5
ACCAGAGTCACATAACAGACGGG	sequence:85	-	45	LTR U5
TACCAGAGTCACATAACAGACGG	sequence:86	-	40	LTR U5
CACTGACTAAAAGGGTCTGAGGG	sequence:118	-	45	LTR U5
ACACTGACTAAAAGGGTCTGAGG	sequence:119	-	45	LTR U5
TCAGACCCTTTTAGTCAGTGTGG	sequence:121	+	45	LTR U5
GTGTGGAAAATCTCTAGCAGTGG	sequence:138	+	45	LTR U5
TCTAGCAGTGGCGCCCGAACAGG	sequence:150	+	65	LTR U5
CTAGCAGTGGCGCCCGAACAGGG	sequence:151	+	65	LTR U5
CTTTCTTCAAGTCCCTGTTCCGG	sequence:163	-	45	
ACTTTCTTCAAGTCCCTGTTCCGG	sequence:164	-	40	
CTGCGTCGAGAGATCTCCTCTGG	sequence:191	-	60	
CAGAGGAGATCTCTCGACGCAGG	sequence:192	+	60	
AGATCTCTCGACGCAGGACTCGG	sequence:198	+	55	
TGCTGAAGAAGAAGCGCGCGCGG	sequence:223	+	60	
GAAGAAGCGCGCGCGGCAAGAGG	sequence:230	+	70	
TTTGACTAGCGGAGGCTAGAAGG	sequence:287	+	50	
GAGGCTAGAAGGAGAGATGGG	sequence:298	+	50	
GCGAGAGCGTCAGTATTAAGCGG	sequence:322	+	50	Gag
CGAGAGCGTCAGTATTAAGCGGG	sequence:323	+	50	Gag
GAGAGCGTCAGTATTAAGCGGGG	sequence:324	+	50	Gag
AGAGCGTCAGTATTAAGCGGGGG	sequence:325	+	50	Gag
GCGGGGAGAATTAGATAGATGG	sequence:341	+	50	Gag
CGGGGGAGAATTAGATAGATGGG	sequence:342	+	45	Gag
GAAAAAATTCGGTTAAGGCCAGG	sequence:364	+	40	Gag
AAAATTCGGTTAAGGCCAGGGGG	sequence:367	+	45	Gag
GAACGATTCGCAGTTAATCCTGG	sequence:439	+	45	Gag
GGCCTTTTAGAGACATCAGAAGG	sequence:460	+	45	Gag
AGCCTTCTGATGTCTCTAAAAGG	sequence:462	-	40	Gag
AGAAGGCTGTAGACAAATACTGG	sequence:477	+	40	Gag
GAAGGCTGTAGACAAATACTGGG	sequence:478	+	40	Gag
CTACAACCAGCCCTTCAGACAGG	sequence:505	+	55	Gag
TCTGATCCTGTCTGAAGGGCTGG	sequence:511	-	55	Gag
TTCTTCTGATCCTGTCTGAAGGG	sequence:515	-	40	Gag
GTTCTTCTGATCCTGTCTGAAGG	sequence:516	-	45	Gag
CCCTCTATTGTGTGCATGCAAGG	sequence:566	+	50	Gag
TCCTTGCATGCACACAATAGAGG	sequence:567	-	45	Gag
TTGTGTGCATGCAAGGATAGAGG	sequence:573	+	45	Gag
GATAGAGGTAAGAACACCAAGG	sequence:588	+	40	Gag
CAGCAAGCAGAAGCTGACGCAGG	sequence:670	+	60	Gag
TGACGCAGGAAAAACAACCCGG	sequence:684	+	45	Gag
TAGGGTAATTCTGGCTGACCGGG	sequence:702	-	50	Gag
ATAGGGTAATTCTGGCTGACCGG	sequence:703	-	45	Gag
TCTGCACTATAGGGTAATTCTGG	sequence:711	-	40	Gag
CTTGAGATTCTGCACTATAGGG	sequence:720	-	40	Gag
CCTATAGTGCAGAATCTCCAAGG	sequence:721	+	45	Gag
CTATAGTGCAGAATCTCCAAGGG	sequence:722	+	40	Gag
GCAGAATCTCCAAGGGCAAATGG	sequence:729	+	50	Gag
CCAAGGGCAAATGGTACATCAGG	sequence:738	+	50	Gag
GGGTACTTCTGGGCTGAAGG	sequence:812	-	50	Gag
GGTGTAAATCTGTGGGGTGG	sequence:860	-	45	Gag
AAACACCATGCTAAACACAGTGG	sequence:876	+	40	Gag
AACACCATGCTAAACACAGTGGG	sequence:877	+	40	Gag
ACACCATGCTAAACACAGTGGGG	sequence:878	+	45	Gag
CACCATGCTAAACACAGTGGGGG	sequence:879	+	50	Gag
ACCATGCTAAACACAGTGGGGGG	sequence:880	+	50	Gag
TCCCCCACTGTGTTTAGCATGG	sequence:881	-	55	Gag
TTCTGCAGCTTCTCATTGATGG	sequence:935	-	45	Gag
TCAATGAGGAAGCTGCAGAATGG	sequence:938	+	45	Gag
CAATGAGGAAGCTGCAGAATGGG	sequence:939	+	45	Gag
AGATTGCATCCAGTGCATGCCGG	sequence:964	+	50	Gag
GATTGCATCCAGTGCATGCCGGG	sequence:965	+	55	Gag
GCAATAGGCCCGCATGCACTGG	sequence:973	-	65	Gag

CATGCCGGGCCTATTGCACCAGG	sequence:979	+	65	Gag
CTGGCCTGGTGAATAGGCCCGG	sequence:983	-	65	Gag
CTCATCTGGCCTGGTGAATAGG	sequence:988	-	55	Gag
CTTGTTCTCTCATCTGGCCTGG	sequence:997	-	55	Gag
CAGGCCAGATGAGAGAACCAAGG	sequence:998	+	55	Gag
AGGCCAGATGAGAGAACCAAGGG	sequence:999	+	50	Gag
GGCCAGATGAGAGAACCAAGGGG	sequence:1000	+	55	Gag
TTCCCTTGGTTCTCTCATCTGG	sequence:1002	-	50	Gag
CCAAGGGGAAGTGACATAGCAGG	sequence:1015	+	55	Gag
AGGAACTACTAGTACCCTTCAGG	sequence:1035	+	45	Gag
AGTACCCTCAGGAACAAATAGG	sequence:1045	+	40	Gag
CCCTTCAGGAACAAATAGGATGG	sequence:1049	+	45	Gag
TCCATCCTATTTGTTCTGAAGG	sequence:1050	-	40	Gag
AATAATCCACCTATCCCAGTAGG	sequence:1078	+	40	Gag
ATTTCTCTACTGGGATAGGTGG	sequence:1084	-	45	Gag
GTATAGCCCTACCAGCATTCTGG	sequence:1152	+	50	Gag
TTATGTCCAGAATGCTGGTAGGG	sequence:1158	-	40	Gag
TTTATGTCCAGAATGCTGGTAGG	sequence:1159	-	40	Gag
GGACATAAAACAAGGACCAAAGG	sequence:1173	+	40	Gag
TAGTCTCTAAAGGGTTCCTTTGG	sequence:1189	-	40	Gag
AAGAGCTGAGCAAGCCTCACAGG	sequence:1236	+	55	Gag
TTGGATGACAGAAACCTTGTTGG	sequence:1269	+	40	Gag
GTTGCAATTTTGGACCAACAAGG	sequence:1283	-	40	Gag
ATTTCTTCTAGTGTAGCTGCTGG	sequence:1339	-	40	Gag
AGAAATGATGACCGCATGTCAGG	sequence:1356	+	45	Gag
GAAATGATGACCGCATGTCAGGG	sequence:1357	+	45	Gag
ATGACCGCATGTCAGGGAGTAGG	sequence:1363	+	55	Gag
TGACCGCATGTCAGGGAGTAGGG	sequence:1364	+	55	Gag
GACCGCATGTCAGGGAGTAGGGG	sequence:1365	+	60	Gag
ACCGCATGTCAGGGAGTAGGGGG	sequence:1366	+	60	Gag
TCCCCCTACTCCCTGACATGCGG	sequence:1367	-	60	Gag
TGTCAGGGAGTAGGGGGACCCGG	sequence:1372	+	65	Gag
AAAACCTCTCGCTTATGGCCGGG	sequence:1390	-	45	Gag
CAAACTCTCGCTTATGGCCGG	sequence:1391	-	45	Gag
CGGCCATAAAGCGAGAGTTTTGG	sequence:1392	+	50	Gag
CAGCCAAAACCTCTCGCTTATGG	sequence:1395	-	45	Gag
TAAAGCGAGAGTTTTGGCTGAGG	sequence:1398	+	45	Gag
TGGCAGGTGCACTTGTTACTTGG	sequence:1428	-	50	Gag
CCTGCCATAATGATGCAGAGAGG	sequence:1444	+	50	Gag
ATTGCCCTCTGCATCATTATGG	sequence:1448	-	40	Gag
GGCACATAGCCAGAAATTGCAGG	sequence:1520	+	50	Gag
GCACATAGCCAGAAATTGCAGGG	sequence:1521	+	45	Gag
TCTAGGGGCCCTGCAATTTCTGG	sequence:1529	-	55	Gag
ATTGCAGGGCCCCTAGAAAAAGG	sequence:1535	+	50	Gag
TTGCAGGGCCCCTAGAAAAAGGG	sequence:1536	+	50	Gag
GCCCCTAGAAAAAGGGCTGTTGG	sequence:1543	+	55	Gag
TCCAACAGCCCTTTTCTAGGGG	sequence:1544	-	45	Gag
TTCCAACAGCCCTTTTCTAGGG	sequence:1545	-	40	Gag
TTTCCAACAGCCCTTTTCTAGG	sequence:1546	-	40	Gag
GCTGTTGGAAATGTGAAAGAAGG	sequence:1558	+	40	Gag
GAAGATTTGGCCTTCCCACAAGG	sequence:1624	+	50	Gag Pol
AAGATTTGGCCTTCCCACAAGGG	sequence:1625	+	45	Gag Pol
AGATTTGGCCTTCCCACAAGGGG	sequence:1626	+	50	Gag Pol
TTTGGCCTTCCCACAAGGGGAGG	sequence:1629	+	60	Gag Pol
CCTTCCCACAAGGGGAGGCCAGG	sequence:1634	+	70	Gag Pol
CTTCCCACAAGGGGAGGCCAGGG	sequence:1635	+	65	Gag Pol
CTGCTCTGAAGAAAATCCCTGG	sequence:1652	-	45	Gag Pol
TGAGAAGGGGCTGTTGGCTCTGG	sequence:1676	-	60	Gag Pol
CTGCTCTGAGAAGGGGCTGTTGG	sequence:1682	-	60	Gag Pol
CTCTGGTCTGCTCTGAGAAGGGG	sequence:1689	-	55	Gag Pol
GCTCTGGTCTGCTCTGAGAAGGG	sequence:1690	-	55	Gag Pol
GGCTCTGGTCTGCTCTGAGAAGG	sequence:1691	-	60	Gag Pol
TCTGGTGGGGCTGTTGGCTCTGG	sequence:1706	-	65	Gag Pol
CTCTCTTCTGGTGGGGCTGTTGG	sequence:1712	-	60	Gag Pol
CCCACCAGAAGAGAGCTTCAGG	sequence:1719	+	60	Gag Pol
ACCTGAAGCTCTCTTCTGGTGGG	sequence:1720	-	50	Gag Pol
AACCTGAAGCTCTCTTCTGGTGG	sequence:1721	-	50	Gag Pol
CCAGAAGAGAGCTTCAGGTTTGG	sequence:1724	+	50	Gag Pol
CAGAAGAGAGCTTCAGGTTTGGG	sequence:1725	+	45	Gag Pol
AGAAGAGAGCTTCAGGTTTGGGG	sequence:1726	+	45	Gag Pol
AGACAACAACCTCCCTCTCAGAGG	sequence:1752	+	50	Gag Pol

AACAACCTCCCTCTCAGAGGCAGG	sequence:1756	+	55	Gag Pol
GTTGGCTCCTGCCTCTGAGAGGG	sequence:1763	-	60	Gag Pol
TGTTGGCTCCTGCCTCTGAGAGG	sequence:1764	-	60	Gag Pol
GAGGCAGGAGCCAACAGACAAGG	sequence:1771	+	60	Gag Pol
GGATACAGTTCTTGTCTGTTGG	sequence:1781	-	45	Gag Pol
AGTGATTTGAGGGAAGCTAAAGG	sequence:1802	-	40	Gag Pol
GCTTCCCTCAAATCACTCTTTGG	sequence:1808	+	45	Gag Pol
GCTGCCAAAGAGTGATTTGAGGG	sequence:1812	-	45	Gag Pol
AAAGATAGGGGGGCGACTAAAGG	sequence:1851	+	50	Pol
AAGGAAGCTCTATTAGACACAGG	sequence:1870	+	40	Pol
AAATGGATTTGCCAGGAAGATGG	sequence:1916	+	40	Pol
CCAAAAATGATAGGGGGAATTGG	sequence:1942	+	40	Pol
AAAATGATAGGGGGAATTGGAGG	sequence:1945	+	40	Pol
CCTACACCTGTCAACATAATTGG	sequence:2047	+	40	Pol
AGAAATCTGTTGACTCAGCTTGG	sequence:2071	+	40	Pol
ACTTTTGGGCCATCCATTCCTGG	sequence:2149	-	50	Pol
ATGGCCAAAAGTCAAACAATGG	sequence:2159	+	40	Pol
TGTACAGAACTGGAAGGAAGG	sequence:2221	+	40	Pol
GACTTCTGGGAAGTTCAATTAGG	sequence:2365	+	40	Pol
ATTAGGAATACCACACCCTGCGG	sequence:2382	+	45	Pol
TTAGGAATACCACACCCTGCGGG	sequence:2383	+	50	Pol
TAGGAATACCACACCCTGCGGGG	sequence:2384	+	55	Pol
TTTTTTAACCCCGCAGGGTGTGG	sequence:2392	-	50	Pol
ATCAGTAACAGTACTGGACGTGG	sequence:2421	+	45	Pol
TCAGTAACAGTACTGGACGTGGG	sequence:2422	+	45	Pol
GGTGTCTCATTGTTTGTACTAGG	sequence:2506	-	40	Pol
AGTACAAACAATGAGACACCAGG	sequence:2509	+	40	Pol
GTACAAACAATGAGACACCAGGG	sequence:2510	+	40	Pol
TCAGTATAATGTGCTTCCACAGG	sequence:2541	+	40	Pol
CAGTATAATGTGCTTCCACAGGG	sequence:2542	+	40	Pol
ATAATGTGCTTCCACAGGGATGG	sequence:2546	+	45	Pol
GTGCTTCCACAGGGATGGAAAGG	sequence:2551	+	55	Pol
GGTGATCCTTCCATCCCTGTGG	sequence:2557	-	55	Pol
AGAAAGAACCCCATTCCTTTGG	sequence:2774	+	45	Pol
AGAACCCCATTCCTTTGGATGG	sequence:2778	+	50	Pol
GAACCCCATTCCTTTGGATGGG	sequence:2779	+	50	Pol
TAACCCATCCAAAGGAATGGGGG	sequence:2782	-	45	Pol
ATAACCCATCCAAAGGAATGGGG	sequence:2783	-	40	Pol
CATAACCCATCCAAAGGAATGGG	sequence:2784	-	40	Pol
TCATAACCCATCCAAAGGAATGG	sequence:2785	-	40	Pol
GGAGTTCATAACCCATCCAAAGG	sequence:2790	-	45	Pol
GTAATGTCCATTTATCAGGATGG	sequence:2811	-	40	Pol
GGCTGTACTGTCCATTTATCAGG	sequence:2815	-	45	Pol
TCTTTTTCTGGCAGCACTATAGG	sequence:2836	-	40	Pol
GTGCTGCCAGAAAAAGACCCTGG	sequence:2842	+	55	Pol
TGACAGCCAGGGTCTTTTTCTGG	sequence:2848	-	50	Pol
CTGTATGTCATTGACAGCCAGGG	sequence:2859	-	45	Pol
TCTGTATGTCATTGACAGCCAGG	sequence:2860	-	45	Pol
TGGGCAAGTCAGATTTATCCAGG	sequence:2902	+	45	Pol
GGGCAAGTCAGATTTATCCAGGG	sequence:2903	+	45	Pol
TTAGTGCTTTAGTCCCTAAGG	sequence:2952	-	40	Pol
AGCTCTGCTTCTGCTGTTAGTGG	sequence:2986	-	50	Pol
GAAATACAGAAGCAGGGACAAGG	sequence:3088	+	45	Pol
AGAAGCAGGGACAAGGCCAATGG	sequence:3095	+	55	Pol
CAGGAAAGTATGCAAAAGTGAGG	sequence:3158	+	40	Pol
GGAAAGTATGCAAAAGTGAGGGG	sequence:3160	+	40	Pol
AAGAAACATGGGAAGCATGGTGG	sequence:3290	+	45	Pol
AAGCATGGTGGACAGAGTATTGG	sequence:3302	+	45	Pol
CAGAGTATTGGCAAGCCACCTGG	sequence:3314	+	55	Pol
AAGCCACCTGGATTCCTGAGTGG	sequence:3326	+	55	Pol
AGCCACCTGGATTCCTGAGTGGG	sequence:3327	+	55	Pol
AACTCCCACTCAGGAATCCAGG	sequence:3332	-	50	Pol
GTATTGACAAACTCCCACTCAGG	sequence:3340	-	45	Pol
CCCCTCCCTTAGTGAAATTATGG	sequence:3362	+	45	Pol
CTATGGGTTCTTTTCTAACTGG	sequence:3387	-	40	Pol
GAAAGTTTCTGCTCCTACTATGG	sequence:3404	-	40	Pol
ATGTAGATGGGGCAGCTAATAGG	sequence:3428	+	45	Pol
TGTAGATGGGGCAGCTAATAGGG	sequence:3429	+	45	Pol
GCAGGATATGTTACTGACAGAGG	sequence:3469	+	45	Pol
CAGGATATGTTACTGACAGAGGG	sequence:3470	+	40	Pol
CATCTAGCTTTGCAAGATTCAGG	sequence:3553	+	40	Pol

AGGAAAAGGTCTACCTGGCATGG	sequence:3689	+	50	Pol
GGAAAAGGTCTACCTGGCATGGG	sequence:3690	+	50	Pol
TGTGTGCTGGCACCCATGCCAGG	sequence:3702	-	65	Pol
GCATGGGTGCCAGCACACAAAGG	sequence:3706	+	60	Pol
CCAGCACACAAAGGAATTGGAGG	sequence:3715	+	50	Pol
TCACAGTAATTGGAGAGCAATGG	sequence:3831	+	40	Pol
CTTTTGCTACTACAGGTGGCAGG	sequence:3867	-	50	Pol
CTAAAAGGAGAAGCCATGCATGG	sequence:3919	+	45	Pol
GGACAAGTAGACTGTAGTCCAGG	sequence:3940	+	50	Pol
TAGACTGTAGTCCAGGAATATGG	sequence:3947	+	40	Pol
CAATCTAGCTGCCATATTCCTGG	sequence:3958	-	45	Pol
TGGCTACATGAACTGCTACCAGG	sequence:4005	-	50	Pol
GTAGCAGTTCATGTAGCCAGTGG	sequence:4009	+	50	Pol
GAAGTTATTCCAGCAGAAACAGG	sequence:4045	+	40	Pol
GGGGCCATTGTCTGTATGTATGG	sequence:4121	-	50	Pol
CTGCAGTACTGATGAAATTGGGG	sequence:4140	-	40	Pol
CTGCAGTTAAGGCCGCTGTTGG	sequence:4157	+	60	Pol
CAGTTAAGGCCGCTGTTGGTGG	sequence:4160	+	60	Pol
AGTTAAGGCCGCTGTTGGTGGG	sequence:4161	+	55	Pol
TAAGGCCGCTGTTGGTGGGCGG	sequence:4164	+	65	Pol
AAGGCCGCTGTTGGTGGGCGGG	sequence:4165	+	70	Pol
GATCCCCGCCACCAACAGGCGG	sequence:4169	-	70	Pol
CTTGATCCCCGCCACCAACAGG	sequence:4172	-	65	Pol
TTGGTGGGCGGGGATCAAGCAGG	sequence:4176	+	65	Pol
CCCTACAATCCCCAAAGTCAAGG	sequence:4210	+	50	Pol
TCCTTGACTTTGGGGATTGTAGG	sequence:4211	-	45	Pol
TCTACTACTCCTTGACTTTGGGG	sequence:4219	-	40	Pol
TATAGGACAGGTAAGAGAGCAGG	sequence:4269	+	45	Pol
TCTCAAGACAGCAGTACAAATGG	sequence:4299	+	40	Pol
TAAAAGAAAAGGGGGGATTGGGG	sequence:4341	+	40	Pol cPPT
GGGATTGGGGAGTATAGTGCAGG	sequence:4354	+	55	Pol cPPT
GGATTGGGGAGTATAGTGCAGGG	sequence:4355	+	50	Pol
GATTGGGGAGTATAGTGCAGGGG	sequence:4356	+	50	Pol
GGGACAGCAGAGATCCACTTTGG	sequence:4469	+	55	Pol
AGCAGAGATCCACTTTGGAAAGG	sequence:4474	+	45	Pol
TTTGCTGGTCCTTTCCAAAGTGG	sequence:4483	-	45	Pol
AAGGACCAGCAAAGCTTCTCTGG	sequence:4493	+	50	Pol
CCAGCAAAGCTTCTCTGGAAAGG	sequence:4498	+	50	Pol
AAGCTTCTCTGGAAAGGTGAAGG	sequence:4504	+	45	Pol
AGCTTCTCTGGAAAGGTGAAGGG	sequence:4505	+	45	Pol
GATTATGGAAAACAGATGGCAGG	sequence:4594	+	40	Pol
TGATTGTGTGGCAAGTAGACAGG	sequence:4620	+	45	Vif
TGTGGCAAGTAGACAGGATGAGG	sequence:4626	+	50	Vif
ACAGGATGAGGATTAGAACATGG	sequence:4638	+	40	Vif
CTTCCCTGACCTATACATATGG	sequence:4678	-	40	Vif
GGTCAGGGAAAGCTAAAAAATGG	sequence:4689	+	40	Vif
TCAGAAGTACACATCCCCTAGG	sequence:4754	+	45	Vif
CAGAAGTACACATCCCCTAGGG	sequence:4755	+	45	Vif
AGAAGTACACATCCCCTAGGGG	sequence:4756	+	45	Vif
AGTACACATCCCCTAGGGGAGG	sequence:4759	+	55	Vif
CTAATTTAGCCTCCCCTAGTGGG	sequence:4768	-	45	Vif
ACTAATTTAGCCTCCCCTAGTGG	sequence:4769	-	45	Vif
ACATATTGGGGTCTGCATACAGG	sequence:4799	+	45	Vif
TGCATACAGGAGAAAGAGACTGG	sequence:4812	+	45	Vif
GTCAGGGAGTCTCCATAGAATGG	sequence:4842	+	50	Vif
AGGGAGTCTCCATAGAATGGAGG	sequence:4845	+	50	Vif
GGGAGTCTCCATAGAATGGAGGG	sequence:4846	+	50	Vif
TCTCCATAGAATGGAGGGAAAGG	sequence:4851	+	45	Vif
TCTCCTTCCCTCCATTCTATGG	sequence:4854	-	45	Vif
TAGGTCAGGGTCTACTTGTGTGG	sequence:4881	-	50	Vif
TTAGTTGGTCTGCTAGGTCAGGG	sequence:4894	-	45	Vif
ATTAGTTGGTCTGCTAGGTCAGG	sequence:4895	-	45	Vif
CATGAATTAGTTGGTCTGCTAGG	sequence:4900	-	40	Vif
TAGGACGTTTAGTTAGCCCTAGG	sequence:4971	+	45	Vif
CTGCTTGATATTCACACCTAGGG	sequence:4987	-	40	Vif
CCTAGGTGTGAATATCAAGCAGG	sequence:4988	+	45	Vif
CAAGCAGGACATAACAAGGTAGG	sequence:5003	+	45	Vif
AAGCAGGACATAACAAGGTAGGG	sequence:5004	+	40	Vif
GGTAGGGTCTCTACAGTACTTGG	sequence:5020	+	50	Vif
AGCCACCTTTGCCTAGTGTTAGG	sequence:5076	+	50	Vif
TTCCTAACACTAGGCCAAAGGTGG	sequence:5078	-	45	Vif

AGTTTCCTAACACTAGGCAAAGG	sequence:5081	-	40	Vif
TCTGTCAGTTTCCTAACACTAGG	sequence:5087	-	40	Vif
GGAAACTGACAGAGGATAGATGG	sequence:5097	+	45	Vif
AGGATAGATGGAACAAGCCCCGG	sequence:5109	+	50	Vif
GAACAAGCCCCGGAAGACCAAGG	sequence:5119	+	60	Vif
AACAAGCCCCGGAAGACCAAGGG	sequence:5120	+	55	Vif
CTGTGGCCCTTGGTCTTCCGGGG	sequence:5126	-	65	Vif
TCTGTGGCCCTTGGTCTTCCGGG	sequence:5127	-	60	Vif
CTCTGTGGCCCTTGGTCTTCCGG	sequence:5128	-	60	Vif
ATGGCTCTCTGTGGCCCTTGG	sequence:5136	-	60	Vif Vpr
GCATTGTATGGCTCTCTGTGG	sequence:5143	-	50	Vif Vpr
AGAGAGAGCCATAACAATGCATGG	sequence:5147	+	45	Vif Vpr
TCTAGTGCCATGCATTGTATGG	sequence:5155	-	40	Vif Vpr
ATGGACTAGAGCTTCTAGAGG	sequence:5166	+	45	Vif Vpr
GCTGTTAGACATTTTCTAGGGG	sequence:5203	+	40	Vpr
TTAGACATTTTCTAGGGGATGG	sequence:5207	+	40	Vpr
TTTCTAGGGGATGGCTCCATGG	sequence:5215	+	55	Vpr
AAGCCATGGAGCCATCCCCTAGG	sequence:5218	-	60	Vpr
AGATATGTTGCTAAGCCATGG	sequence:5232	-	40	Vpr
ACTTATGGGGATACTTGGGCAGG	sequence:5260	+	50	Vpr
TGGGGATACTTGGGCAGGAGTGG	sequence:5265	+	60	Vpr
GGATGTCAACATAGCAGAATAGG	sequence:5338	+	40	Vpr
AGAATAGGCATTACTIONGACGAGG	sequence:5353	+	45	Vpr
GAATAGGCATTACTIONGACGAGGG	sequence:5354	+	45	Vpr
TTACTCGACGAGGGAGAGTAAGG	sequence:5363	+	50	Vpr
CGACGAGGGAGAGTAAGGAATGG	sequence:5368	+	55	Vpr
GGCTCTAGTCTAGGGTCTACTGG	sequence:5393	-	55	Vpr
TAGACCCTAGACTAGAGCCCTGG	sequence:5397	+	55	Vpr 5' Tat
GCTTCCAGGGCTCTAGTCTAGGG	sequence:5401	-	55	Vpr 5' Tat
TGCTTCCAGGGCTCTAGTCTAGG	sequence:5402	-	55	Vpr 5' Tat
CTAGAGCCCTGGAAGCATCCAGG	sequence:5408	+	60	Vpr 5' Tat
TGACTTCTGGATGCTTCCAGGG	sequence:5414	-	50	Vpr 5' Tat
CTGACTTCTGGATGCTTCCAGG	sequence:5415	-	55	5' Tat
GCAGTCTTAGGCTGACTTCTGG	sequence:5426	-	55	5' Tat
AAAGGCTTAGGCATCTCCTATGG	sequence:5507	+	45	5' Tat
GCTTAGGCATCTCCTATGGCAGG	sequence:5511	+	55	5' Tat
TCTCCTATGGCAGGAAGAAGCGG	sequence:5520	+	50	5' Tat
TCTCCGCTTCTTCTGCCATAGG	sequence:5523	-	55	5' Tat
ACAGCGACGAAGAACTCCTCAGG	sequence:5545	+	55	5' Tat 5' Rev
TGATGAGTCTGACTGTCCTGAGG	sequence:5561	-	50	5' Tat 5' Rev
AGAGAAAGAGCAGAAGACAGTGG	sequence:5756	+	45	Vpu
CAGTGGCAATGAAAGTGAAGGGG	sequence:5773	+	45	Vpu
CAATGAAAGTGAAGGGGATCAGG	sequence:5779	+	45	Vpu
GGAAGAATTATCAGCTCTTGTGG	sequence:5800	+	40	Vpu
AGATGGGGCATCATGCTCCTTGG	sequence:5823	+	55	Vpu
GATGGGGCATCATGCTCCTTGGG	sequence:5824	+	55	Vpu
TACAGATCATCAACGTCCCAAGG	sequence:5840	-	45	Vpu
GGGAGCAGCAGGAAGCACTATGG	sequence:7346	+	60	RRE
GGAGCAGCAGGAAGCACTATGGG	sequence:7347	+	55	RRE
CACTATGGGCGCAGCGTCAATGG	sequence:7361	+	60	RRE
CGCAGCGTCAATGGCGCTGACGG	sequence:7370	+	65	RRE
GTCAATGGCGCTGACGGTACAGG	sequence:7376	+	60	RRE
CAGGCCAGACAATTATTGTCTGG	sequence:7395	+	45	RRE
CAATTTGCTGAGGGCTATTGAGG	sequence:7436	+	45	RRE
ATCTGCTGCAACTCACAGTCTGG	sequence:7468	+	50	RRE
TCTGCTGCAACTCACAGTCTGGG	sequence:7469	+	50	RRE
CTGCTGCAACTCACAGTCTGGGG	sequence:7470	+	55	RRE
CTGGGGCATTAAAGCAGCTCCAGG	sequence:7487	+	60	RRE
GCAGCTCCAGGCAAGAGTCCTGG	sequence:7499	+	65	RRE
CTATAGCCAGGACTCTTGCCTGG	sequence:7505	-	55	RRE
GTAGGTATCTTTCTATAGCCAGG	sequence:7517	-	40	RRE
GGCTATAGAAAGATACCTACAGG	sequence:7520	+	40	RRE
TGTCGGGTCCCCTCGGAGCTGGG	sequence:7940	-	70	3' Tat 3' Rev
TCGATTCCTTCGGGCCTGTCGGG	sequence:7956	-	60	3' Tat 3' Rev
TTCGATTCCTTCGGGCCTGTCGG	sequence:7957	-	55	3' Tat 3' Rev
CCGAAGGAATCGAAGAAGAAGG	sequence:7965	+	50	3' Tat 3' Rev
ACCTTCTTCTTCGATTCCTTCGG	sequence:7966	-	40	3' Tat 3' Rev
GAAGGAATCGAAGAAGAAGGTGG	sequence:7968	+	45	3' Tat 3' Rev
GAGAGAGACAGAGACGGATCCGG	sequence:7992	+	55	3' Tat 3' Rev
AGAGAGACAGAGACGGATCCGGG	sequence:7993	+	55	3' Tat 3' Rev
GGATCCGGGCCATTAGTGAATGG	sequence:8007	+	55	3' Tat 3' Rev

GAATCCATTCACTAATGGCCCGG	sequence:8011	-	45	3' Tat 3' Rev
CAATCATCTGGGTCGACCTGCGG	sequence:8038	+	55	3' Rev
TGAAGAGGCACAGGCTCCGCAGG	sequence:8054	-	65	3' Rev
GGTGGTAGCTGAAGAGGCACAGG	sequence:8063	-	60	3' Rev
TCAAGCGGTGGTAGCTGAAGAGG	sequence:8069	-	55	3' Rev
AGAGTAAGTCTCTCAAGCGGTGG	sequence:8081	-	50	3' Rev
TCAAGAGTAAGTCTCTCAAGCGG	sequence:8084	-	40	3' Rev
CTTGATTGTAGTGAGGATTGTGG	sequence:8102	+	40	3' Rev
AGTGAGGATTGTGGAACCTTCTGG	sequence:8111	+	45	3' Rev
GTGAGGATTGTGGAACCTTCTGGG	sequence:8112	+	45	3' Rev
TTGTGGAACCTTCTGGGACGCAGG	sequence:8119	+	55	3' Rev
TGTGGAACCTTCTGGGACGCAGGG	sequence:8120	+	55	3' Rev
GTGGAACCTTCTGGGACGCAGGGG	sequence:8121	+	60	3' Rev
TGGAACCTTCTGGGACGCAGGGGG	sequence:8122	+	60	3' Rev
ACTTCTGGGACGCAGGGGGTGGG	sequence:8126	+	65	3' Rev
GGTGGGAAGCCCTCAAATACTGG	sequence:8143	+	55	3' Rev
GGGAAGCCCTCAAATACTGGTGG	sequence:8146	+	55	3' Rev
GAGATTCCACCAGTATTTGAGGG	sequence:8152	-	40	3' Rev
GGAGATTCCACCAGTATTTGAGG	sequence:8153	-	45	3' Rev
GGTGGAAATCTCCTGCAATATTGG	sequence:8164	+	45	3' Rev
TCTCCTGCAATATTGGATTGAGG	sequence:8171	+	40	3' Rev
GTTCTGAAATCCAATATTGCAGG	sequence:8174	-	40	3' Rev
AAGTGGTCAAAACATAAGGGGGG	sequence:8359	+	40	Nef
TGGTCAAAACATAAGGGGGGTGG	sequence:8362	+	50	Nef
CAAAACATAAGGGGGGTGGATGG	sequence:8366	+	50	Nef
TAAGGGGGGTGGATGGCCTGCGG	sequence:8373	+	65	Nef
GGGTGGATGGCCTGCGGTGAGGG	sequence:8379	+	70	Nef
TGCATTCTTTCCCTCACCGCAGG	sequence:8389	-	55	Nef
CGAGCTGAGCCAGCAGCAGACGG	sequence:8413	+	65	Nef
GAGCTGAGCCAGCAGCAGACGGG	sequence:8414	+	65	Nef
AGCTGAGCCAGCAGCAGACGGGG	sequence:8415	+	65	Nef
TGAGCCAGCAGCAGACGGGGTGG	sequence:8418	+	70	Nef
GAGCCAGCAGCAGACGGGGTGGG	sequence:8419	+	70	Nef
GGGACCAGTATCTCGAGACTTGG	sequence:8439	+	55	Nef
TTTTCCAAGTCTCGAGATACTGG	sequence:8443	-	40	Nef
TCTCGAGACTTGGAAAACATGG	sequence:8449	+	40	Nef
CTAATAATGCTGCTTGTGTCTGG	sequence:8492	+	40	Nef
TGTCTGGCTAGAAGCACAAAGAGG	sequence:8508	+	50	Nef
CTGGCTAGAAGCACAAAGAGGAGG	sequence:8511	+	55	Nef
GCTAGAAGCACAAAGAGGAGGAGG	sequence:8514	+	55	Nef
GGGTTTTCCAGTCAGACCTCAGG	sequence:8541	+	55	Nef
AAAGGTACCTGAGGTCTGACTGG	sequence:8548	-	50	Nef
ATTGGTCTTAAAGGTACCTGAGG	sequence:8557	-	40	Nef
AAAGCTCCCTTGTAAAGTCATTGG	sequence:8575	-	40	Nef
AATGACTTACAAGGGAGCTTTGG	sequence:8577	+	40	Nef
AAAGAAAAGGGGGGACTGGAAGG	sequence:8617	+	50	Nef
AAGAAAAGGGGGGACTGGAAGGG	sequence:8618	+	50	Nef
GACAAGATATCCTTGATCTGTGG	sequence:8660	+	40	LTR U3 Nef
TGTGGTAGACCCACAGATCAAGG	sequence:8670	-	50	LTR U3 Nef
CTGTGGGTCTACCACACACAAGG	sequence:8677	+	55	LTR U3 Nef
CAGGGAAGTAGCCTTGTGTGTGG	sequence:8688	-	55	LTR U3 Nef
CACAAGGCTACTTCCCTGATTGG	sequence:8693	+	50	LTR U3 Nef
ACAAGGCTACTTCCCTGATTGGG	sequence:8694	+	45	LTR U3 Nef
GTGTGTAGTTATCCCAATCAGGG	sequence:8706	-	40	LTR U3 Nef
GGTGTGTAGTTATCCCAATCAGG	sequence:8707	-	45	LTR U3 Nef
GATTGGGATAACTACACACCAGG	sequence:8710	+	45	LTR U3 Nef
ATTGGGATAACTACACACCAGGG	sequence:8711	+	40	LTR U3 Nef
GATAACTACACACCAGGGCCAGG	sequence:8716	+	55	LTR U3 Nef
GGATATCTGATCCCTGGCCCTGG	sequence:8728	-	60	LTR U3 Nef
ATCAGTGGATATCTGATCCCTGG	sequence:8734	-	45	LTR U3 Nef
GATATCCACTGATCTTTGGATGG	sequence:8744	+	40	LTR U3 Nef
AAGCACCATCCAAAGATCAGTGG	sequence:8749	-	45	LTR U3 Nef
AGTACCAGTTGATCCAGACCAGG	sequence:8778	+	50	LTR U3 Nef
TCTACCTGGTCTGGATCAACTGG	sequence:8782	-	50	LTR U3 Nef
TGATCCAGACCAGGTAGAAAAGG	sequence:8787	+	45	LTR U3 Nef
TTGGCCTTTTCTACCTGGTCTGG	sequence:8791	-	50	LTR U3 Nef
CTTCATTGGCCTTTTCTACCTGG	sequence:8796	-	45	LTR U3 Nef
CAGGTAGAAAAGGCCAATGAAGG	sequence:8797	+	45	LTR U3 Nef
TTACACCCTATGAGCCAACATGG	sequence:8836	+	45	LTR U3 Nef
TACACCCTATGAGCCAACATGGG	sequence:8837	+	45	LTR U3 Nef
CTATCCCATGTTGGCTCATAGGG	sequence:8841	-	45	LTR U3 Nef

TCTATCCCATGTTGGCTCATAGG	sequence:8842	-	45	LTR U3 Nef
CCAACATGGGATAGATGACCCGG	sequence:8850	+	50	LTR U3 Nef
CGGAGAAAGAAGTATTAGTGTGG	sequence:8870	+	40	LTR U3 Nef
CCGCCTAGCATTTCATCACATGG	sequence:8904	+	50	LTR U3 Nef
GGGCCATGTGATGAAATGCTAGG	sequence:8907	-	50	LTR U3 Nef
CATGGCCCGAGAGATACATCCGG	sequence:8922	+	55	LTR U3 Nef
GTACTCCGGATGTATCTCTCGGG	sequence:8927	-	50	LTR U3 Nef
AGTACTCCGGATGTATCTCTCGG	sequence:8928	-	45	LTR U3 Nef
CAGCAGTTCTTGTAGTACTCCGG	sequence:8941	-	45	LTR U3 Nef
ACTACAAGAAGTCTGCTGACACCGG	sequence:8948	+	45	LTR U3 Nef
CAGCATTCTTGTAGTACTCCGG	sequence:8967	-	40	LTR U3 Nef
GCTGACACCGAGCTTCTACAAGG	sequence:8986	+	55	LTR U3
CTGACACCGAGCTTCTACAAGG	sequence:8987	+	50	LTR U3
GAAAGTCCCTTGTAGAAGCTCGG	sequence:8993	-	45	LTR U3
TTCTACAAGGGACTTTCCGCTGG	sequence:8999	+	50	LTR U3
TCTACAAGGGACTTTCCGCTGGG	sequence:9000	+	50	LTR U3
CTACAAGGGACTTTCCGCTGGGG	sequence:9001	+	55	LTR U3
CTTTCCGCTGGGGACTTTCCAGG	sequence:9011	+	60	LTR U3
TTTCCGCTGGGGACTTTCCAGGG	sequence:9012	+	55	LTR U3
CCGCTGGGGACTTTCCAGGGAGG	sequence:9015	+	70	LTR U3
GGGGACTTTCCAGGGAGGTGTGG	sequence:9020	+	65	LTR U3
TTTCCAGGGAGGTGTGGCCTGGG	sequence:9026	+	60	LTR U3
AGGTGTGGCCTGGGCGGAACAGG	sequence:9035	+	70	LTR U3
GGTGTGGCCTGGGCGGAACAGGG	sequence:9036	+	70	LTR U3
GTGTGGCCTGGGCGGAACAGGGG	sequence:9037	+	70	LTR U3
ACCACTCCCCTGTTCCGCCAGG	sequence:9043	-	70	LTR U3
TGCTTATATGCAGCATCTGAGGG	sequence:9069	-	40	LTR U3
CTGCTTATATGCAGCATCTGAGG	sequence:9070	-	45	LTR U3
CAGCTGCTTTTTGCCTGTACTGG	sequence:9090	+	50	LTR U3
AGCTGCTTTTTGCCTGTACTGGG	sequence:9091	+	45	LTR U3
TAAGTGGAGAGACCCAGTACAGG	sequence:9103	-	50	LTR U3
AGGCTCAGATCTGGTCTAACTGG	sequence:9119	-	50	LTR R
AGTTAGACCAGATCTGAGCCTGG	sequence:9121	+	50	LTR R
GTTAGACCAGATCTGAGCCTGGG	sequence:9122	+	50	LTR R
AGAGCTCCCAGGCTCAGATCTGG	sequence:9128	-	60	LTR R
ATCTGAGCCTGGGAGCTCTCTGG	sequence:9132	+	60	LTR R
CAGCTAGCCAGAGAGCTCCCAGG	sequence:9139	-	65	LTR R
CTGGGAGCTCTCTGGCTAGCTGG	sequence:9140	+	65	LTR R
TGGGAGCTCTCTGGCTAGCTGGG	sequence:9141	+	60	LTR R
GGGAGCTCTCTGGCTAGCTGGGG	sequence:9142	+	65	LTR R
TAGATTGAGGCTTAAGCAGTGGG	sequence:9167	-	40	LTR R
ATAGATTGAGGCTTAAGCAGTGG	sequence:9168	-	40	LTR R
TTAGGTGACCCTATAGATTGAGG	sequence:9180	-	40	LTR R
GGTCACCTAAATACACTGTTTGG	sequence:9193	+	40	LTR R
TTGATCCAGACCAGGTAGAAAGG	sequence:9239	+	45	LTR U5
ATTGGCCTTTCTACCTGGTCTGG	sequence:9244	-	50	LTR U5
CCAGGTAGAAAGGCCAATGAAGG	sequence:9249	+	50	LTR U5
CAGGTAGAAAGGCCAATGAAGGG	sequence:9250	+	45	LTR U5
AACAGTTGTTCTCCCTTCATTGG	sequence:9262	-	40	LTR U5
GTTACCCCTATGAGCCAACATGG	sequence:9286	+	50	LTR U5
TTCACCCTATGAGCCAACATGGG	sequence:9287	+	45	LTR U5
GCCAACATGGGATAGATACCCGG	sequence:9299	+	50	LTR U5
TCCGGGTATCTATCCCATGTTGG	sequence:9300	-	50	LTR U5

HIV-1 1058 Target sequence	5' end of gRNA target site	Strand	% GC content	HIV-1 target
GCGGGGGAAAATTAGATACATGG	sequence:14	+	45	Gag
CGGGGGAAAATTAGATACATGGG	sequence:15	+	40	Gag
CATGGGAGAAAATTCGGTTAAGG	sequence:32	+	40	Gag
GAGAAAATTCGGTTAAGGCCAGG	sequence:37	+	45	Gag
AGAAAATTCGGTTAAGGCCAGGG	sequence:38	+	40	Gag
GAAAATTCGGTTAAGGCCAGGGG	sequence:39	+	45	Gag
AAAATTCGGTTAAGGCCAGGGGG	sequence:40	+	45	Gag
AACATATAGTATGGGCAAGCAGG	sequence:83	+	40	Gag
ACATATAGTATGGGCAAGCAGGG	sequence:84	+	40	Gag
GAACGATTTGACTTAACCCTGG	sequence:112	+	45	Gag
CTGCTGTTTCTAACAGGCCAGGG	sequence:129	-	50	Gag
TCTGCTGTTTCTAACAGGCCAGG	sequence:130	-	50	Gag
AGCCTTCTGCTGTTTCTAACAGG	sequence:135	-	45	Gag
CTACAGCCATCCCTTCAAACAGG	sequence:178	+	50	Gag
TCTGATCCTGTTTGAAGGGATGG	sequence:184	-	45	Gag
GTTCTTCTGATCCTGTTTGAAGG	sequence:189	-	40	Gag
CCCTCTATTGTGTACATCAAAGG	sequence:239	+	40	Gag
AGAGGTAAGAGACACCAAAGAGG	sequence:264	+	45	Gag
AGAGGCTTTAGACAAGATAGAGG	sequence:282	+	40	Gag
TGACACAGGAAACAGCAGCCAGG	sequence:354	+	55	Gag
TAGGGTAATTTGGCTGACCTGG	sequence:372	-	45	Gag
CTTGAGGTTCTGCACTATAGGG	sequence:390	-	45	Gag
CCTATAGTGCAGAACCTCCAAGG	sequence:391	+	50	Gag
GCAGAACCTCCAAGGGCAAATGG	sequence:399	+	55	Gag
GATGTACCATTTGCCCTTGGAGG	sequence:405	-	50	Gag
CCAAGGGCAAATGGTACATCAGG	sequence:408	+	50	Gag
GCAAACATGGGTATTACTTCTGG	sequence:490	-	40	Gag
GGTGTTTAAATCTTGTGGGGTGG	sequence:530	-	45	Gag
AAACACCATGCTAAACACAGTGG	sequence:546	+	40	Gag
AACACCATGCTAAACACAGTGGG	sequence:547	+	40	Gag
ACACCATGCTAAACACAGTGGGG	sequence:548	+	45	Gag
CACCATGCTAAACACAGTGGGGG	sequence:549	+	50	Gag
ACCATGCTAAACACAGTGGGGGG	sequence:550	+	50	Gag
TCCCCCACTGTGTTTAGCATGG	sequence:551	-	55	Gag
TTCTGCAGCTTCTCATTGATGG	sequence:605	-	45	Gag
TCAATGAGGAAGCTGCAGAATGG	sequence:608	+	45	Gag
CAATGAGGAAGCTGCAGAATGGG	sequence:609	+	45	Gag
AGAATACATCCAGCGCAAGCAGG	sequence:634	+	50	Gag
GAATACATCCAGCGCAAGCAGGG	sequence:635	+	50	Gag
GCTATAGGCCCTGCTTGCCTGG	sequence:643	-	65	Gag
CAAGCAGGGCCTATAGCACCAGG	sequence:649	+	60	Gag
CTTGGGTCTCTTATCTGGCCTGG	sequence:667	-	55	Gag
CAGGCCAGATAAGAGACCCAAGG	sequence:668	+	55	Gag
AGGCCAGATAAGAGACCCAAGGG	sequence:669	+	50	Gag
GGCCAGATAAGAGACCCAAGGGG	sequence:670	+	55	Gag
TTCCCTTGGGTCTCTTATCTGG	sequence:672	-	50	Gag
CTGCTATGTCACCTCCCCTTGGG	sequence:684	-	50	Gag
CTTATCTGGCCTGGTGTATAGG	sequence:658	-	50	Gag
CCAAGGGGAAGTGACATAGCAGG	sequence:685	+	55	Gag
AGGAACTACTAGTACCCTTCAGG	sequence:705	+	45	Gag
CCCTTCAGGAACAAATAACATGG	sequence:719	+	40	Gag
AATAATCCACCTATCCCAGTAGG	sequence:748	+	40	Gag
ATTTCTCCTACTGGGATAGGTGG	sequence:754	-	45	Gag
TAGATTTCTCCTACTGGGATAGG	sequence:757	-	40	Gag
GTATAGTCCCTACCAGCATTCTGG	sequence:822	+	45	Gag
CTTATGTCCAGAATGCTGGTAGG	sequence:829	-	45	Gag
AGCATTCTGGACATAAGACAAGG	sequence:835	+	40	Gag
GGACATAAGACAAGGACCAAAGG	sequence:843	+	45	Gag
AAGGACCAAAGGAACCCTTTAGG	sequence:854	+	45	Gag
AGGACCAAAGGAACCCTTTAGGG	sequence:855	+	45	Gag
TAGTCCCTAAAGGGTTCCTTTGG	sequence:859	-	45	Gag
CGGTCTACATAGTCCCTAAAGGG	sequence:868	-	45	Gag
CCTTTAGGACTATGTAGACCGG	sequence:869	+	45	Gag
AAGAGCCGAGCAAGCTTCACAGG	sequence:906	+	55	Gag
AGCCGAGCAAGCTTCACAGGAGG	sequence:909	+	60	Gag
TACCTCCTGTGAAGCTTGCTCGG	sequence:911	-	50	Gag
TTGGATGACAGAAACCTTGTTGG	sequence:939	+	40	Gag
GTTGCGTTTTGGACCAACAAGG	sequence:953	-	50	Gag
TACAATCTGGGTTGCGTTTTGG	sequence:963	-	45	Gag
AGGACCAGCAGCTACACTAGAGG	sequence:1005	+	55	Gag

ATTCCTCTAGTGTAGCTGCTGG	sequence:1009	-	45	Gag
GGAAATGATGACAGCATGTCAGG	sequence:1026	+	45	Gag
GAAATGATGACAGCATGTCAGGG	sequence:1027	+	40	Gag
GATGACAGCATGTCAGGGAGTGG	sequence:1032	+	55	Gag
ATGACAGCATGTCAGGGAGTGGG	sequence:1033	+	50	Gag
ACAGCATGTCAGGGAGTGGGAGG	sequence:1036	+	60	Gag
TGTCAGGGAGTGGGAGGACCTGG	sequence:1042	+	65	Gag
AAAACCTTGCTTTATGGCCAGG	sequence:1060	-	40	Gag
TGGCCATAAAGCAAGAGTTTTGG	sequence:1062	+	40	Gag
CAGCCAAAACCTTGCTTTATGG	sequence:1065	-	40	Gag
TGGCACCTGAATTTGTTACTTGG	sequence:1098	-	40	Gag
GGTGCCATAATGATGCAGAAAGG	sequence:1114	+	45	Gag
GAAAGGCAATTTTAGGAACCAGG	sequence:1131	+	40	Gag
GGCACATAGCCAAAAATTGCAGG	sequence:1184	+	45	Gag
GCACATAGCCAAAAATTGCAGGG	sequence:1185	+	40	Gag
CCAAAAATTGCAGGGCCCCTAGG	sequence:1193	+	55	Gag
TTGCAGGGCCCCTAGGAAGAAGG	sequence:1200	+	60	Gag
TGCAGGGCCCCTAGGAAGAAGGG	sequence:1201	+	60	Gag
TCCAACAGCCCTTCTCCTAGGG	sequence:1209	-	50	Gag
TTCCAACAGCCCTTCTCCTAGG	sequence:1210	-	50	Gag
AAGAAGGGCTGTTGGAAATGTGG	sequence:1216	+	45	Gag
GAAAGATTGTGATCAGAGACAGG	sequence:1257	+	40	Gag
AAGATCTGGCCTTCCCACAAAGG	sequence:1294	+	50	Gag Pol
TCTGGCCTTCCCACAAAGGAAGG	sequence:1298	+	55	Gag Pol
CCTTCCCACAAAGGAAGGCCAGG	sequence:1303	+	60	Gag Pol
AATTCCCTGGCCTTCTTTGTGG	sequence:1308	-	50	Gag Pol
CTGCTCTGAAGAAAATTCCCTGG	sequence:1321	-	45	Gag Pol
GCTGGTGGGGCTGTTGGCTCTGG	sequence:1345	-	70	Gag Pol
CTCTCTGCTGGTGGGGCTGTTGG	sequence:1351	-	65	Gag Pol
CCCACCAGCAGAGAGCTTCAGG	sequence:1358	+	65	Gag Pol
CCAGCAGAGAGCTTCAGGTTTGG	sequence:1363	+	55	Gag Pol
CAGCAGAGAGCTTCAGGTTTGGG	sequence:1364	+	50	Gag Pol
AGCAGAGAGCTTCAGGTTTGGGG	sequence:1365	+	50	Gag Pol
AGAGAGCTTCAGGTTTGGGGAGG	sequence:1368	+	55	Gag Pol
AACAACCTCCCTCAGAAGCAGG	sequence:1395	+	50	Gag Pol
TGTGGCTCCTGCTTCTGAGAGGG	sequence:1402	-	55	Gag Pol
TTGTGGCTCCTGCTTCTGAGAGG	sequence:1403	-	55	Gag Pol
AGTGATCTGAGGGAAGCTAAAGG	sequence:1432	-	45	Gag Pol
GTTCCCTCAGATCACTTTTGG	sequence:1438	+	50	Gag Pol
GTTGCCAAAGAGTGATCTGAGGG	sequence:1442	-	45	Gag Pol
CGTTGCCAAAGAGTGATCTGAGG	sequence:1443	-	50	Gag Pol
CCCCTCGTCACAATAAAGATAGG	sequence:1467	+	45	Gag Pol
CCCTCGTCACAATAAAGATAGGG	sequence:1468	+	40	Gag Pol
CCTCGTCACAATAAAGATAGGGG	sequence:1469	+	40	Gag Pol
CTCGTCACAATAAAGATAGGGGG	sequence:1470	+	40	Gag Pol
TCGTACAATAAAGATAGGGGGG	sequence:1471	+	40	Gag Pol
AAAGATAGGGGGGCGAGCTAAAGG	sequence:1481	+	50	Pol
AAATGAATTTGCCCGAAGATGG	sequence:1546	+	40	Pol
CCAAAAATGATAGGGGGAATTGG	sequence:1572	+	40	Pol
AAAATGATAGGGGGAATTGGAGG	sequence:1575	+	40	Pol
CCTACACCTGTCAACATAATTGG	sequence:1677	+	40	Pol
GTACCAGTAAAATTGAAGCCAGG	sequence:1761	+	40	Pol
AAATTGAAGCCAGGAATGGATGG	sequence:1770	+	40	Pol
ACTTTTGGACCATCCATTCCTGG	sequence:1779	-	45	Pol
TGTGCAGAAATGGAAAAGGAAGG	sequence:1851	+	40	Pol
GTGCAGAAATGGAAAAGGAAGGG	sequence:1852	+	40	Pol
GACTTCTGGGAAGTTCAATTAGG	sequence:1995	+	40	Pol
TTAGGAATACCACATCCCAGG	sequence:2013	+	50	Pol
TAGGAATACCACATCCCAGGG	sequence:2014	+	50	Pol
TTTTTTAACCTGCGGGATGTGG	sequence:2022	-	45	Pol
ATCAGTAACAGTACTGGATGTGG	sequence:2051	+	40	Pol
TCAGTAACAGTACTGGATGTGGG	sequence:2052	+	40	Pol
CTCGTTGTTTACACTAGGTATGG	sequence:2131	-	40	Pol
GGTGTCTCGTTGTTTACACTAGG	sequence:2136	-	45	Pol
AGTGTAACAACGAGACACCAGG	sequence:2139	+	45	Pol
GTGTAACAACGAGACACCAGGG	sequence:2140	+	45	Pol
TCAGTATAATGTGCTTCCACAGG	sequence:2171	+	40	Pol
CAGTATAATGTGCTTCCACAGGG	sequence:2172	+	40	Pol
ATAATGTGCTTCCACAGGGATGG	sequence:2176	+	45	Pol
GTGCTTCCACAGGGATGGAAAGG	sequence:2181	+	55	Pol
GGTGATCCTTCCATCCCTGTGG	sequence:2187	-	55	Pol

TGAGACAACATCTGTTGAAGTGG	sequence:2353	+	40	Pol
GAGACAACATCTGTTGAAGTGGG	sequence:2354	+	40	Pol
AGACAACATCTGTTGAAGTGGGG	sequence:2355	+	40	Pol
AGAAAGAACCTCCATTCTTTGG	sequence:2404	+	40	Pol
AGAACCTCCATTCTTTGGATGG	sequence:2408	+	45	Pol
GAACCTCCATTCTTTGGATGGG	sequence:2409	+	45	Pol
AACCTCCATTCTTTGGATGGGG	sequence:2410	+	45	Pol
TACCCCATCAAAGGAATGGAGG	sequence:2412	-	50	Pol
TCATACCCCATCAAAGGAATGG	sequence:2415	-	45	Pol
GGAGTTCATACCCCATCAAAGG	sequence:2420	-	50	Pol
GTA CTGTCCATTATCAGGATGG	sequence:2441	-	40	Pol
GGCTGTACTGTCCATTTATCAGG	sequence:2445	-	45	Pol
TCTTTCTCTGGCAGCACTATAGG	sequence:2466	-	45	Pol
TGCTGCCAGAGAAAGACACCTGG	sequence:2473	+	55	Pol
ACAGTCCAGGTGTCTTTCTCTGG	sequence:2478	-	50	Pol
TGGGCAAGTCAGATTTATGCAGG	sequence:2535	+	45	Pol
GGGCAAGTCAGATTTATGCAGGG	sequence:2536	+	45	Pol
GGCACTTATGTAAACTCCTTAGG	sequence:2569	+	40	Pol
TTAGTGCTTTGGTTCCCCTAAGG	sequence:2585	-	45	Pol
TAGCAGAAATACAGAAGCAGGGG	sequence:2716	+	40	Pol
GAAATACAGAAGCAGGGGCAAGG	sequence:2721	+	50	Pol
AGAAGCAGGGGCAAGGCCAATGG	sequence:2728	+	60	Pol
AGGAAACATGGGAATCATGGTGG	sequence:2923	+	45	Pol
CAGATTATTGGCAAGCCACCTGG	sequence:2947	+	50	Pol
AAGCCACCTGGATTCTGAGTGG	sequence:2959	+	55	Pol
AGCCACCTGGATTCTGAGTGGG	sequence:2960	+	55	Pol
AAATCCCCTCAGGAATCCAGG	sequence:2965	-	45	Pol
GTATTGACAAATTCCCCTCAGG	sequence:2973	-	40	Pol
CCCCTCCCTTAGTGAAATTATGG	sequence:2995	+	45	Pol
CTATGGGTTCTTTCTCTAACTGG	sequence:3020	-	40	Pol
AAGGTTTCTGCTCCTACTATGGG	sequence:3036	-	40	Pol
GAAGGTTTCTGCTCCTACTATGG	sequence:3037	-	45	Pol
GCAGAAACCTTCTACGTAGATGG	sequence:3048	+	45	Pol
CAGAAACCTTCTACGTAGATGGG	sequence:3049	+	40	Pol
AGAAACCTTCTACGTAGATGGGG	sequence:3050	+	40	Pol
AGCTGCCCATCTACGTAGAAGG	sequence:3055	-	55	Pol
ACGTAGATGGGGCAGCTAACAGG	sequence:3061	+	55	Pol
CGTAGATGGGGCAGCTAACAGGG	sequence:3062	+	55	Pol
GCTAACAGGGAGACTAACTAGG	sequence:3075	+	45	Pol
GAGACTAACTAGGAAAAGCAGG	sequence:3084	+	40	Pol
GCAGGATATGTTACCGACAGAGG	sequence:3102	+	50	Pol
TACCGACAGAGGAAGACAAAAGG	sequence:3113	+	45	Pol
AACTTTTGTCTTCTCTGTCGG	sequence:3115	-	40	Pol
CTGATTTGTTGTGTCAGTCAGGG	sequence:3142	-	40	Pol
AGCAATTCATCTAGCACTGCAGG	sequence:3179	+	45	Pol
CATCTAGCACTGCAGGATTCAGG	sequence:3186	+	50	Pol
AGGAAAAGGTCTACCTGACATGG	sequence:3322	+	45	Pol
TGTGTGCTGGTACCCATGTCAGG	sequence:3335	-	55	Pol
ACATGGGTACCAGCACACAAAGG	sequence:3339	+	50	Pol
CCAGCACACAAAGGAATTGGGGG	sequence:3348	+	50	Pol
AATTAGTCAGTGCTGGGATCAGG	sequence:3388	+	45	Pol
AGATGGAATAGATAAGGCCCAGG	sequence:3425	+	45	Pol
CTTTTGCTACTATAGGTGGCAGG	sequence:3500	-	45	Pol
TAGCTGACATTTATCACAGCTGG	sequence:3532	-	40	Pol
CTAAAAGGGGAAGCCATGCATGG	sequence:3552	+	50	Pol
GGACAAGTAGACTGTAGTCCAGG	sequence:3573	+	50	Pol
TAGACTGTAGTCCAGGAATATGG	sequence:3580	+	40	Pol
CAATCTAGTTGCCATATTCCTGG	sequence:3591	-	40	Pol
TGGCTACATGAACTGCTACCAGG	sequence:3638	-	50	Pol
GTAGCAGTTCATGTAGCCAGTGG	sequence:3642	+	50	Pol
TAGCAGTTCATGTAGCCAGTGGG	sequence:3643	+	45	Pol
TTCTGCTTCTATATACCCACTGG	sequence:3658	-	40	Pol
GAAGTGATCCCAGCAGAGACAGG	sequence:3678	+	55	Pol
AAGTGATCCCAGCAGAGACAGGG	sequence:3679	+	50	Pol
TTTCTGCCCTGTCTCTGCTGGG	sequence:3686	-	55	Pol
GTTTCTGCCCTGTCTCTGCTGG	sequence:3687	-	60	Pol
CAGCAACTTCACAGTAGTGCGG	sequence:3773	+	50	Pol
CTTACCAGTAGTGCGGTCAAGG	sequence:3779	+	55	Pol
GGCGGCCTTGACCGCACTACTGG	sequence:3784	-	70	Pol
GTGCGGTCAAGGCCGCCTGTTGG	sequence:3790	+	70	Pol
CGGTCAAGGCCGCCTGTTGGTGG	sequence:3793	+	70	Pol

GGTCAAGGCCCGCTGTTGGTGGG	sequence:3794	+	65	Pol
CAAGGCCGCCTGTTGGTGGGCGG	sequence:3797	+	70	Pol
AAGGCCGCCTGTTGGTGGGCGGG	sequence:3798	+	70	Pol
CTTGACCCCCGCCACCAACAGG	sequence:3805	-	70	Pol
TTGGTGGGCGGGGGTCAAGCAGG	sequence:3809	+	70	Pol
GCGGGGGTCAAGCAGGAATTTGG	sequence:3816	+	60	Pol
CCCTACAATCCCCAAAGTCAAGG	sequence:3843	+	50	Pol
TCCTTGACTTTGGGGATTGTAGG	sequence:3844	-	45	Pol
TCTACTACTCCTTGACTTTGGGG	sequence:3852	-	40	Pol
TATAGGACAGGTAAGAGATCAGG	sequence:3902	+	40	Pol
TTTAAAAGAAGAGGGGGGATTGG	sequence:3972	+	40	Pol cPPT
TTAAAAGAAGAGGGGGGATTGGG	sequence:3973	+	40	Pol cPPT
TAAAAGAAGAGGGGGGATTGGGG	sequence:3974	+	45	Pol cPPT
GGGATTGGGGACTATAGTGCAGG	sequence:3987	+	55	Pol cPPT
GGATTGGGGACTATAGTGCAGGG	sequence:3988	+	50	Pol cPPT
GATTGGGGACTATAGTGCAGGGG	sequence:3989	+	50	Pol cPPT
GGGACAGCAGAGATCCACTTTGG	sequence:4102	+	55	Pol
AGCAGAGATCCACTTTGGAAAGG	sequence:4107	+	45	Pol
TTTGCTGGTCCTTTCCAAAGTGG	sequence:4116	-	45	Pol
AAGGACCAGCAAAGCTCCTCTGG	sequence:4126	+	55	Pol
CCAGCAAAGCTCCTCTGGAAAGG	sequence:4131	+	55	Pol
AAGCTCCTCTGGAAAGGTGAAGG	sequence:4137	+	50	Pol
AGCTCCTCTGGAAAGGTGAAGGG	sequence:4138	+	50	Pol
GCTCCTCTGGAAAGGTGAAGGGG	sequence:4139	+	55	Pol
CTGCCCTTCACCTTTCCAGAGG	sequence:4142	-	60	Pol
GATTATGGAAAACAGATGGCAGG	sequence:4227	+	40	Pol
GATGGCAGGTGATGATTGTATGG	sequence:4241	+	45	Pol Vif
TGATTGTATGGCAAGTAGACAGG	sequence:4253	+	40	Pol Vif
TATGGCAAGTAGACAGGATGAGG	sequence:4259	+	45	Pol Vif
ACAGGATGAGGATTAGAACATGG	sequence:4271	+	40	Pol Vif
TATCAAGGAAAGCTAAGGACTGG	sequence:4322	+	40	Vif
AGAAGTACACATCCCCTAGAGG	sequence:4389	+	45	Vif
CCCCTAGAGGATGCTGAATTGG	sequence:4401	+	50	Vif
ACCAATTCAGCATCCTCTAGTGG	sequence:4402	-	45	Vif
ACATATTGGGGTCTGCATACAGG	sequence:4432	+	45	Vif
TGCATACAGGAGAAAGAGAATGG	sequence:4445	+	40	Vif
AGGAGAAAGAGAATGGCATTGG	sequence:4452	+	40	Vif
GTCAGGGAGTCTCCATAGAATGG	sequence:4475	+	50	Vif
AGGGAGTCTCCATAGAATGGAGG	sequence:4478	+	50	Vif
TTAGTTGGTCTGCTAGGTCAGGG	sequence:4527	-	45	Vif
ATTAGTTGGTCTGCTAGGTCAGG	sequence:4528	-	45	Vif
GATGAATTAGTTGGTCTGCTAGG	sequence:4533	-	40	Vif
GCCATAAGAAATGCCCTGTTAGG	sequence:4585	+	45	Vif
TCCTAACAGGGCATTCTTATGG	sequence:4586	-	40	Vif
CATAGGTGTGAATATCAAGCAGG	sequence:4621	+	40	Vif
CAAGCAGGACATAACAAGGTAGG	sequence:4636	+	45	Vif
GGTAGGATCTTTACAGTACTTGG	sequence:4653	+	40	Vif
AAAGGTGGCTTTCTCCTTTTGG	sequence:4696	-	40	Vif
TTCGCAACACTAGGCAAAGGTGG	sequence:4711	-	50	Vif
AGTTTCGCAACACTAGGCAAAGG	sequence:4714	-	45	Vif
GCCTAGTGTTCGAAACTGACGG	sequence:4719	+	50	Vif
TCCGTCAAGTTTCGCAACACTAGG	sequence:4720	-	50	Vif
TAGTGTTCGAAACTGACGGAGG	sequence:4722	+	50	Vif
CGAAACTGACGGAGGACAGATGG	sequence:4730	+	55	Vif
GAACAAGTCCCAGCAGACCAAGG	sequence:4752	+	55	Vif Vpr
AACAAGTCCCAGCAGACCAAGGG	sequence:4753	+	50	Vif Vpr
TCTGTGGCCCTTGGTCTGCTGGG	sequence:4760	-	60	Vif Vpr
AGCAGACCAAGGGCCACAGAGGG	sequence:4763	+	60	Vif Vpr
ATGGCTCCCTCTGTGGCCCTTGG	sequence:4769	-	65	Vif Vpr
TCATTGTATGGCTCCCTCTGTGG	sequence:4776	-	50	Vif Vpr
AGAGGGAGCCATACAATGAATGG	sequence:4780	+	45	Vif Vpr
TGAGACACTTTCCAGGATATGG	sequence:4840	+	45	Vpr
AAGTTGTGAAGCCATATCCTGGG	sequence:4851	-	40	Vpr
TAAGTTGTGAAGCCATATCCTGG	sequence:4852	-	40	Vpr
AGGATATGGCTTCAACTTAGG	sequence:4854	+	40	Vpr
ACTTATGGGGATACTTGGGTAGG	sequence:4893	+	45	Vpr
TGGGGATACTTGGGTAGGAGTGG	sequence:4898	+	55	Vpr
GGATGTCAACATAGCAGAATAGG	sequence:4971	+	40	Vpr
GAATAGGCATTATTCGACAGAGG	sequence:4987	+	40	Vpr
CGACAGAGGAGAGCAAGAAATGG	sequence:5001	+	50	Vpr
GGCTCTAGTTTAGGGTCTACTGG	sequence:5026	-	50	Vpr

TAGACCCTAAACTAGAGCCCTGG	sequence:5030	+	50	Vpr 5' Tat
GCTTCCAGGGCTCTAGTTTAGGG	sequence:5034	-	50	Vpr 5' Tat
TGCTTCCAGGGCTCTAGTTTAGG	sequence:5035	-	50	Vpr 5' Tat
CTAGAGCCCTGGAAGCATCCAGG	sequence:5041	+	60	5' Tat
TGACTTCTGGATGCTTCCAGGG	sequence:5047	-	50	5' Tat
CTGACTTCTGGATGCTTCCAGG	sequence:5048	-	55	5' Tat
AGCATCCAGGAAGTCAGCCTAGG	sequence:5054	+	55	5' Tat
GCAGTCCTAGGCTGACTTCCTGG	sequence:5059	-	60	5' Tat
CAATTGGTACAAGCAGTCCTAGG	sequence:5071	-	45	5' Tat
AAAGGCTTAGGCATCTCCTATGG	sequence:5140	+	45	5' Tat
GCTTAGGCATCTCCTATGGCAGG	sequence:5144	+	55	5' Tat
TCTCCTATGGCAGGAAGAAGCGG	sequence:5153	+	50	5' Tat
TCTCCGCTTCTTCTGCCATAGG	sequence:5156	-	55	5' Tat
TGATCAATCTGACTGTCTGGAGG	sequence:5194	-	45	5' Tat 5' Rev
GCTTGATCAATCTGACTGTCTGG	sequence:5197	-	45	5' Tat 5' Rev
ACAGGTAATACTACTGCTTTGG	sequence:5227	-	45	5' Tat 5' Rev
GATAGTGGCAATGAGAGTGAAGG	sequence:5404	+	45	Vpu
ATAGTGGCAATGAGAGTGAAGGG	sequence:5405	+	40	Vpu
TAGTGGCAATGAGAGTGAAGGGG	sequence:5406	+	45	Vpu
CAATGAGAGTGAAGGGGATCAGG	sequence:5412	+	50	Vpu
TGAGAGTGAAGGGGATCAGGAGG	sequence:5415	+	55	Vpu
GGAGGAATTGTCAGCACTCGTGG	sequence:5433	+	55	Vpu
ATTGTCAGCACTCGTGGAGATGG	sequence:5439	+	50	Vpu
TTGTCAGCACTCGTGGAGATGGG	sequence:5440	+	50	Vpu
TGTCAGCACTCGTGGAGATGGGG	sequence:5441	+	55	Vpu
ACTACGCTACCATGCTCCTTGG	sequence:5465	+	55	Vpu
CATCAATATTCCAAGGAGCATGG	sequence:5475	-	40	Vpu
GTAGAGCTGCAGAACAGTTGTGG	sequence:5502	+	50	Vpu
GGGAGCAGCAGGAAGCACTATGG	sequence:6982	+	60	RRE
GGAGCAGCAGGAAGCACTATGGG	sequence:6983	+	55	RRE
CGCAGCGTCAATGACGCTGACGG	sequence:7006	+	60	RRE
GTCAATGACGCTGACGGTACAGG	sequence:7012	+	55	RRE
CAGGCCAGACTATTATTGTCTGG	sequence:7031	+	45	RRE
AACAGCAGAACAACCTTGCTGAGG	sequence:7062	+	45	RRE
ACAGCAGAACAACCTTGCTGAGGG	sequence:7063	+	45	RRE
CAACTTGCTGAGGGCTATTGAGG	sequence:7072	+	50	RRE
ACAACATCTGTTGCAACTCACGG	sequence:7099	+	40	RRE
ATCTGTTGCAACTCACGGTCTGG	sequence:7104	+	50	RRE
TCTGTTGCAACTCACGGTCTGGG	sequence:7105	+	50	RRE
CTGTTGCAACTCACGGTCTGGGG	sequence:7106	+	55	RRE
CTGGGGCATCAAGCAGCTCCAGG	sequence:7123	+	65	RRE
GCAGCTCCAGGCAAGAGTCCTGG	sequence:7135	+	65	RRE
CCAGGCAAGAGTCCTGGCTGTGG	sequence:7141	+	65	RRE
CTAGGTATCTTCCACAGCCAGG	sequence:7153	-	50	RRE
GGCTGTGGAAAGATACCTAGAGG	sequence:7156	+	50	RRE
ACCCGCCTCCCAACCCAGAGGGG	sequence:7568	+	70	3' Tat 3' Rev
TCCCCTCTGGGTTGGGAGGCGGG	sequence:7569	-	70	3' Tat 3' Rev
GTCCCCTCTGGGTTGGGAGGCGG	sequence:7570	-	70	3' Tat 3' Rev
TGTCGGGTCCCCTCTGGGTTGGG	sequence:7576	-	65	3' Tat 3' Rev
CTGTCGGGTCCCCTCTGGGTTGG	sequence:7577	-	70	3' Tat 3' Rev
CAACCCAGAGGGGACCCGACAGG	sequence:7578	+	70	3' Tat 3' Rev
TCTGTTCTTCGGGCCTGTCGGG	sequence:7592	-	60	3' Tat 3' Rev
TTCTGTTCTTCGGGCCTGTCGG	sequence:7593	-	55	3' Tat 3' Rev
CCCGAAGGAACAGAAGAAGAAGG	sequence:7601	+	50	3' Tat 3' Rev
ACCTTCTTCTTCTGTTCTTCGG	sequence:7602	-	40	3' Tat 3' Rev
GAGAGAGACAGAGACAGATCCGG	sequence:7628	+	50	3' Tat 3' Rev
AGACAGATCCGGACCATTAGTGG	sequence:7639	+	50	3' Tat 3' Rev
AGATCCGGACCATTAGTGGATGG	sequence:7643	+	50	3' Tat 3' Rev
GAATCCATCCACTAATGGTCCGG	sequence:7647	-	45	3' Tat 3' Rev
GCTAAGAATCCATCCACTAATGG	sequence:7652	-	40	3' Tat 3' Rev
CAATCATCTGGGTCGATCTACGG	sequence:7674	+	45	3' Rev
AGTCTCTCAATCGGTGGTAGAGG	sequence:7711	-	50	3' Rev
AGAGTAAGTCTCTCAATCGGTGG	sequence:7717	-	45	3' Rev
ACTTACTCTTGATTGTGACGAGG	sequence:7731	+	40	3' Rev
CTTGATTGTGACGAGGACTCTGG	sequence:7738	+	50	3' Rev
GACGAGGACTCTGGAACCTTCTGG	sequence:7747	+	55	3' Rev
ACGAGGACTCTGGAACCTTCTGGG	sequence:7748	+	50	3' Rev
CTCTGGAACCTTCTGGGACGCAGG	sequence:7755	+	60	3' Rev
TCTGGAACCTTCTGGGACGCAGGG	sequence:7756	+	55	3' Rev
CTGGAACCTTCTGGGACGCAGGGG	sequence:7757	+	60	3' Rev
TGGAACCTTCTGGGACGCAGGGGG	sequence:7758	+	60	3' Rev

ACTTCTGGGACGCAGGGGGTGGG	sequence:7762	+	65	3' Rev
GGTGGGAGATCCTCAAATATTGG	sequence:7779	+	45	3' Rev
GGGAGATCCTCAAATATTGGTGG	sequence:7782	+	45	3' Rev
GGAGATTCCACCAATATTTGAGG	sequence:7789	-	40	3' Rev
TCTCCTACAGTATTGGAGTCAGG	sequence:7807	+	45	3' Rev
GTTCTGACTCCAATACTGTAGG	sequence:7810	-	45	3' Rev
GTGGTCAAACGATAGTATTGTGG	sequence:7997	+	40	Nef
CAAACGATAGTATTGTGGGATGG	sequence:8002	+	40	Nef
TGCTGGCTCAGTTGCTGGTGG	sequence:8053	-	60	Nef
TCTGCTGCTGGCTCAGTTGCTGG	sequence:8058	-	60	Nef
GCAACTGAGCCAGCAGCAGAAGG	sequence:8061	+	60	Nef
CAACTGAGCCAGCAGCAGAAGGG	sequence:8062	+	55	Nef
TGAGCCAGCAGCAGAAGGGGTGG	sequence:8066	+	65	Nef
GAGCCAGCAGCAGAAGGGGTGGG	sequence:8067	+	65	Nef
GCTCCCACCCCTTCTGCTGCTGG	sequence:8070	-	70	Nef
GGGAGCAGCATCTCGAGACCTGG	sequence:8087	+	65	Nef
TCTCGAGACCTGGAAAGACATGG	sequence:8097	+	50	Nef
TGAGTGCTCCATGTCTTTCCAGG	sequence:8105	-	50	Nef
CTAACAATGCTGCTTGTGCCTGG	sequence:8146	+	50	Nef
TGCCTGGCTAGAAGCACAAAGAGG	sequence:8162	+	55	Nef
ATCCTCTTGTGCTTCTAGCCAGG	sequence:8164	-	50	Nef
GCTAGAAGCACAAAGAGGATGAGG	sequence:8168	+	50	Nef
AGAAGCACAAAGAGGATGAGGAGG	sequence:8171	+	50	Nef
AGCACAAAGAGGATGAGGAGGTGG	sequence:8174	+	55	Nef
GCACAAGAGGATGAGGAGGTGGG	sequence:8175	+	55	Nef
GGGTTTTCCAGTCAGACCTCAGG	sequence:8195	+	55	Nef
AAAGGCACCTGAGGTCTGACTGG	sequence:8202	-	55	Nef
ATTGGTCTTAAAGGCACCTGAGG	sequence:8211	-	45	Nef
AAAGAAAAGGGGGGACTGGAAGG	sequence:8271	+	50	Nef
AAGAAAAGGGGGGACTGGAAGGG	sequence:8272	+	50	Nef
AATTTACTCCAGCAAAGACAGG	sequence:8297	+	40	LTR U3 Nef
AAGGATATCCTGTCTTTGCTGGG	sequence:8305	-	40	LTR U3 Nef
CAAGGATATCCTGTCTTTGCTGG	sequence:8306	-	45	LTR U3 Nef
GACAGGATATCCTTGATCTGTGG	sequence:8314	+	45	LTR U3 Nef
ACAGGATATCCTTGATCTGTGGG	sequence:8315	+	40	LTR U3 Nef
TGTGGTAGACCCACAGATCAAGG	sequence:8324	-	50	LTR U3 Nef
CTGTGGGTCTACCACACAAGG	sequence:8331	+	55	LTR U3 Nef
CAGGGAAGTAGCCTTGTGTGTGG	sequence:8342	-	55	LTR U3 Nef
CACAAGGCTACTTCCCTGATTGG	sequence:8347	+	50	LTR U3 Nef
GTGTGTAATTCTGCCAATCAGGG	sequence:8360	-	40	LTR U3 Nef
GGTGTGTAATTCTGCCAATCAGG	sequence:8361	-	45	LTR U3 Nef
GATTGGCAGAATTACACACCAGG	sequence:8364	+	45	LTR U3 Nef
ATTGGCAGAATTACACACCAGGG	sequence:8365	+	40	LTR U3 Nef
CAGAATTACACACCAGGGCCAGG	sequence:8370	+	55	LTR U3 Nef
AGAATTACACACCAGGGCCAGGG	sequence:8371	+	50	LTR U3 Nef
GGATATCTGGTCCCTGGCCCTGG	sequence:8382	-	65	LTR U3 Nef
GTCAGGGGATATCTGGTCCCTGG	sequence:8388	-	60	LTR U3 Nef
ACCAGATATCCCCTGACCTTTGG	sequence:8394	+	50	LTR U3 Nef
TCCAAAGGTCAGGGGATATCTGG	sequence:8395	-	50	LTR U3 Nef
GATATCCCCTGACCTTTGGATGG	sequence:8398	+	50	LTR U3 Nef
AAGCACCATCCAAAGGTCAGGGG	sequence:8403	-	50	LTR U3 Nef
AAAGCACCATCCAAAGGTCAGGG	sequence:8404	-	45	LTR U3 Nef
TAAAGCACCATCCAAAGGTCAGG	sequence:8405	-	45	LTR U3 Nef
TAGCTTAAAGCACCATCCAAAGG	sequence:8410	-	40	LTR U3 Nef
TCTATCTTCTCTGGTTCAACTGG	sequence:8436	-	40	LTR U3 Nef
GTGGCCTCTTCTATCTTCTCTGG	sequence:8445	-	50	LTR U3 Nef
AAGATAGAAGAGGCCACTGAAGG	sequence:8451	+	45	LTR U3 Nef
GCAGTTGTTTTCTCTCAGTGG	sequence:8464	-	45	LTR U3 Nef
TTACACCCGATAAACCAGCATGG	sequence:8490	+	45	LTR U3 Nef
TACACCCGATAAACCAGCATGGG	sequence:8491	+	45	LTR U3 Nef
CCCGATAAACCAGCATGGGATGG	sequence:8495	+	55	LTR U3 Nef
TCCATCCCATGCTGGTTATCGG	sequence:8496	-	45	LTR U3 Nef
CCAGCATGGGATGGACGACCCGG	sequence:8504	+	65	LTR U3 Nef
ACACTAACACTTCTCTCCGGG	sequence:8522	-	45	LTR U3 Nef
CACACTAACACTTCTCTCCGG	sequence:8523	-	45	LTR U3 Nef
CGGAGAGAGAAGTGTTAGTGTGG	sequence:8524	+	50	LTR U3 Nef
GTCGCTAGCACTCCGTCACGTGG	sequence:8557	+	65	LTR U3 Nef
GCATCTCTCGGGCCACGTGACGG	sequence:8569	-	65	LTR U3 Nef
CGTGGCCCGAGAGATGCATCCGG	sequence:8575	+	65	LTR U3 Nef
GTACTCCGGATGCATCTCTCGGG	sequence:8580	-	55	LTR U3 Nef
AGTACTCCGGATGCATCTCTCGG	sequence:8581	-	50	LTR U3 Nef

GATGCATCCGGAGTACTACAAGG	sequence:8587	+	50	LTR U3 Nef
CAGCAGTCCTTGTAGTACTCCGG	sequence:8594	-	50	LTR U3 Nef
CTGACATCGAGATTTCTACAAGG	sequence:8614	+	40	LTR U3 Nef
TTCTACAAGGGACTTTCCACTGG	sequence:8627	+	45	LTR U3
TCTACAAGGGACTTTCCACTGGG	sequence:8628	+	45	LTR U3
CTACAAGGGACTTTCCACTGGGG	sequence:8629	+	50	LTR U3
ACTTTCCACTGGGGACTTTCCGG	sequence:8638	+	50	LTR U3
CTTTCCACTGGGGACTTTCCGGG	sequence:8639	+	55	LTR U3
TTTCCACTGGGGACTTTCCGGGG	sequence:8640	+	55	LTR U3
CCACTGGGGACTTTCCGGGGAGG	sequence:8643	+	70	LTR U3
TTTCCGGGGAGGCGTGGCCTGGG	sequence:8654	+	70	LTR U3
TGCTTATATGCAGCATCTGAGGG	sequence:8697	-	40	LTR U3
CTGCTTATATGCAGCATCTGAGG	sequence:8698	-	45	LTR U3
CAGCTGCTTTTTGCCTGTACTGG	sequence:8718	+	50	LTR U3
AGCTGCTTTTTGCCTGTACTGGG	sequence:8719	+	45	LTR U3
TCTACAGAGAGACCCAGTACAGG	sequence:8731	-	50	LTR U3
CTGTAGACCAGATCTGAGCCTGG	sequence:8747	+	55	LTR R
TGTAGACCAGATCTGAGCCTGGG	sequence:8748	+	50	LTR R
AGAGCTCCCAGGCTCAGATCTGG	sequence:8754	-	60	LTR R
ATCTGAGCCTGGGAGCTCTCTGG	sequence:8758	+	60	LTR R
TAGTTAGCCAGAGAGCTCCCAGG	sequence:8765	-	55	LTR R
TGGGAGCTCTCTGGCTAACTAGG	sequence:8767	+	55	LTR R
GGGAGCTCTCTGGCTAACTAGGG	sequence:8768	+	55	LTR R
CTTTATTGAGGCTTAAGCAGTGG	sequence:8794	-	40	LTR R
ACTCAAGGCAAGCTTTATTGAGG	sequence:8806	-	40	LTR R