Supplementary Data For Publication.

Broadening the versatility of lentiviral vectors as a tool in nucleic acid research via genetic code expansion

Yongxiang Zheng[1], Fei Yu[1], Yiming Wu[1], Longlong Si[1], Huan Xu[1], Chuanling Zhang[1], Qing Xia[1], Sulong Xiao[1], Qi Wang[1,2],Qiuchen He[1], Peng Chen[3], Jiangyun Wang[4], Lihe Zhang[1], Kazunari Taira[1] and Demin Zhou [1*]

[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China

[2]Department of Molecular and Cellular Pharmacology, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China

[3]Department of Chemical Biology, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

[4]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

* To Demin Zhou correspondence should be addressed. Tel: 86-10-8280-5857; Fax: 86-10-8280-5857 ; Email: deminzhou@bjmu.edu.cn

Present Address: [Demin Zhou], School of Pharmaceutical Sciences, Peking University, Beijing, 100191, China.

CONTENT

Supplementary Movie S1. Dynamic Tracking of an Alexa 488 labeled lentiviral vector in three dimensions. Lentiviral vectors were conjugated with Alexa 488 (green) via NAEK and then detected at 488 nm. The cell membrane was stained with DiI (red) and detected with 555-nm laser. The process of viral particles traveling from the cell membrane to the endoplasm was recorded using a spinning-disk confocal microscope equipped with a CO_2 online culture system (Perkin Elmer).

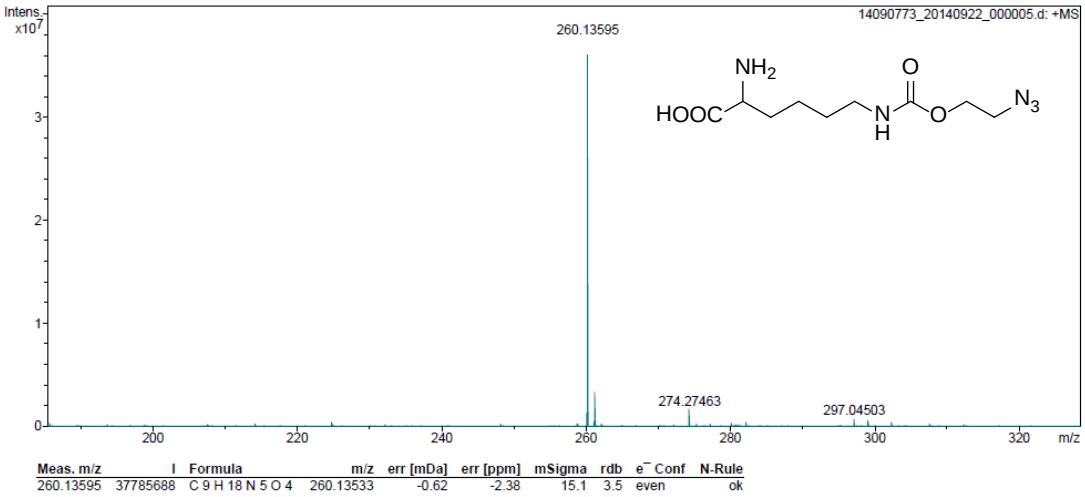
Supplementary Tables and Figures

Table S1. The primers for mutagenesis. The sequence of pCMV-VSVg plasmid is available at *http://www.ncbi.nlm.nih.gov/nuccore/AJ318514.1.*

Mutation Site	Primer	Sequence
N20	Forward	CTGGAAAAATGTTCCTTCTTAGTACCATTATTGCCCGTCAAGC
N20	Rearward	GCTTGACGGGCAATAATGGTACTAAGAAGGAACATTTTTCCAG
D35	Forward	GCTCAGATTTAAATTGGCATAATTAGTTAATAGGCACAGC
D35	Rearward	GCTGTGCCTATTAACTAATTATGCCAATTTAAATCTGAGC
I37	Forward	CAGATTTAAATTGGCATAATGACTTATAGGGCACAGCCTTACAAG
I37	Rearward	CTTGTAAGGCTGTGCCCTATAAGTCATTATGCCAATTTAAATCTG
Q42	Forward	GACTTAATAGGCACAGCCTTA TAG GTCAAAATGCCCAAGAG
Q42	Rearward	CTCTTGGGCATTTTGACCTATAAGGCTGTGCCTATTAAGTC
Y116	Forward	CTTCCCTCCTCAAAGTTGTGGATAGGCAACTGTGACGGATGCC
Y116	Rearward	GGCATCCGTCACA GTTGCCTATCCACA ACTTTGAGGAGGGAAG
D121	Forward	GTGGATATGCAACTGTGACGTAGGCCGAAGCAGTGATTGTCCAG
D121	Rearward	CTGGACAATCACTGCTTCGGCCTACGTCACAGTTGCATATCCAC
V161	Forward	GCAGCAATTACATATGCCCCACTTAGCATAACTCTACAACCTGG
V161	Rearward	CCAGGTTGTAGAGTTATGCTAAGTGGGGCATATGTAATTGCTGC
K172	Forward	CCTGGCATTCTGACTATTAGGTCAAAGGGCTATGTGATTC
K172	Rearward	GAATCACATAGCCCTTTGACCTAATAGTCAGAATGCCAGG
K174	Forward	CAACCTGGCATTCTGACTATAAGGTCTAGGGGCTATGTGATTC
K174	Rearward	GAATCACATAGCCCCTAGACCTTATAGTCAGAATGCCAGGTTG
D185	Forward	CTAACCTCATTTCCATGTAGATCACCTTCTTCTCAGAGGAC
D185	Rearward	GTCCTCTGAGAAGAAGGTGATCTACATGGAAATGAGGTTAG
D192	Forward	CACCTTCTTCTCAGAGTAGGGAGAGCTATCATCC
D192	Rearward	GGATGATAGCTCTCCCTACTCTGAGAAGAAGGTG
K200	Forward	GAGCTATCATCCCTGGGATAGGAGGGCACAGG
K200	Rearward	CCTGTGCCCTCCTATCCCAGGGATGATAGCTC
E201	Forward	CATCCCTGGGAAAGTAGGGCACAGGGTTCAG
E201	Rearward	CTGAACCCTGTGCCCTACTTTCCCAGGGATG

K225	Forward	CAAAATGCAATACTGCTAGCATTGGGGAGTCAGAC
K225	Rearward	GTCTGACTCCCCAATGCTAGCAGTATTGCATTTTG
A240	Forward	GTCTGGTTCGAGATGTAGGATAAGGATCTCTTTGC
A240	Rearward	GCAAAGAGATCCTTATCCTACATCTCGAACCAGAC
K242	Forward	CTGCAAAATGCAATACTGCTAGCATTGGGGAGTCAGAC
K242	Rearward	GTCTGACTCCCCAATGCTAGCAGTATTGCATTTTGCAG
D243	Forward	GGTTCGAGATGGCTGATAAGTAGCTCTTTGCTGCAG
D243	Rearward	CTGCAGCAAAGAGCTACTTATCAGCCATCTCGAACC
A246	Forward	GGCTGATAAGGATCTCTTTTAGGCAGCCAGATTCC
A246	Rearward	GGAATCTGGCTGCCTAAAAGAGATCCTTATCAGCC
A247	Forward	CTGATAAGGATCTCTTTGCTTAGGCCAGATTCCCTGAATG
A247	Rearward	CATTCAGGGAATCTGGCCTAAGCAAAGAGATCCTTATCAG
R249	Forward	GGATCTCTTTGCTGCAGCCTAGTTCCCTGAATGCCCAGAAGG
R249	Rearward	CCTTCTGGGCATTCAGGGAACTAGGCTGCAGCAAAGAGATCC
E252	Forward	CTGCAGCCAGATTCCCTTAGTGCCCAGAAGGGTCAAG
E252	Rearward	CTTGACCCTTCTGGGCACTAAGGGAATCTGGCTGCAG
Y281	Forward	GAGAGGATCTTGGATTAGTCCCTCTGCCAAGAAACC
Y281	Rearward	GGTTTCTTGGCAGAGGGACTAATCCAAGATCCTCTC
W288	Forward	CTCTGCCAAGAAACCTAGAGCAAAATCAGAGCG
W288	Rearward	CGCTCTGATTTTGCTCTAGGTTTCTTGGCAGAG
G321	Forward	GAACCGGTCCTGCTTTCACCATAATCAATTAGACCCTAAAATAC
G321	Rearward	GTATTTTAGGGTCTAATTGATTATGGTGAAAGCAGGACCGGTTC
R332	Forward	CTTTGAGACCAGATACATCTAGGTCGATATTGCTGCTCCAATCC
R332	Rearward	GGATTGGAGCAGCAATATCGACCTAGATGTATCTGGTCTCAAAG
I339	Forward	CAGAGTCGATATTGCTGCTCCATAGCTCTCAAGAATGGTCGG
I339	Rearward	CCGACCATTCTTGAGAGCTATGGAGCAGCAATATCGACTCTG
R342	Forward	GACCAGATACATCAGAGTCTAGATTGCTGCTCCAATCCTCTC
R342	Rearward	GAGAGGATTGGAGCAGCAATCTAGACTCTGATGTATCTGGTC
G345	Forward	CCAATCCTCTCAAGAATGGTCTAGATGATCAGTGGAACTACCAC
G345	Rearward	GTGGTAGTTCCACTGATCATCTAGACCATTCTTGAGAGGATTGG
M346	Forward	CAATCCTCTCAAGAATGGTCGGATAGATCAGTGGAACTACCACAG
M346	Rearward	CTGTGGTAGTTCCACTGATCTATCCGACCATTCTTGAGAGGATTG
P362	Forward	GGGATGACTGGGCATAGTATGAAGACGTGGAAATTGGACC
P362	Rearward	GGTCCAATTTCCACGTCTTCATACTATGCCCAGTCATCCC

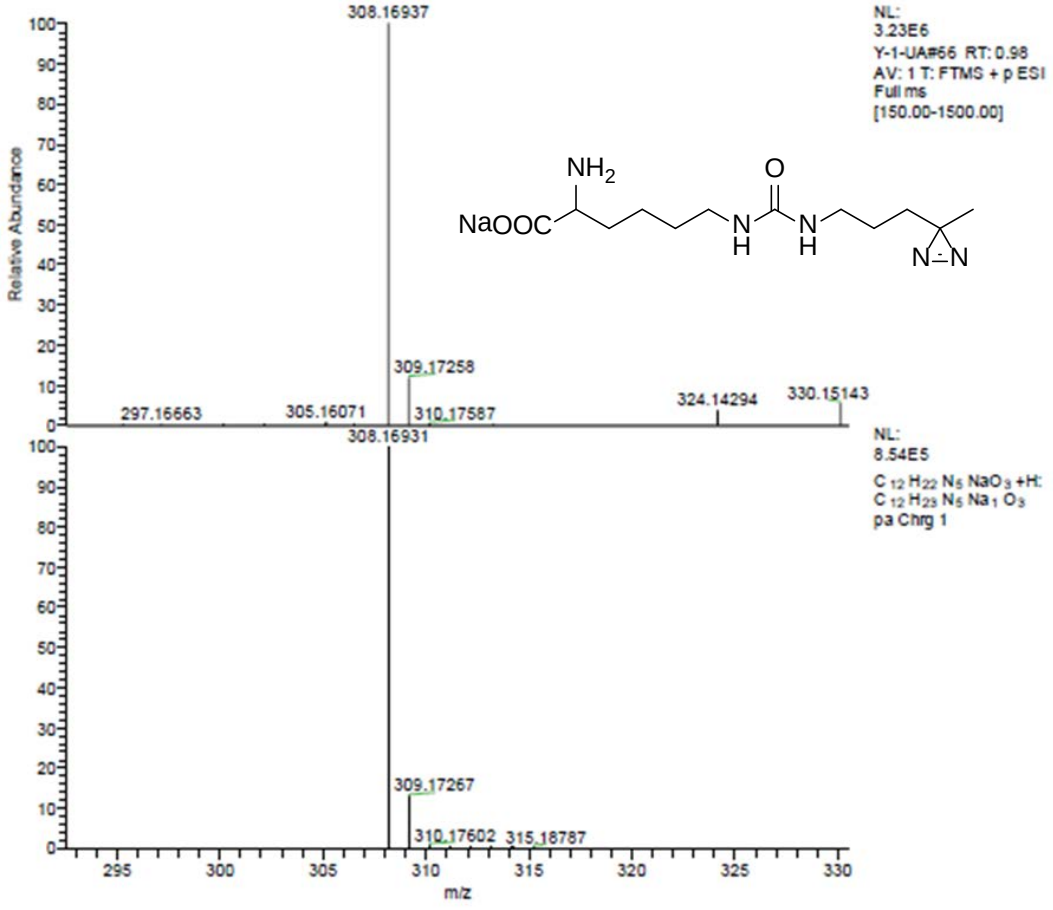
Supplementary Figure S1. Characterizations of unnatural amino acids NAEK and DiZPK synthesized in lab by high resolution mass spectrometry.



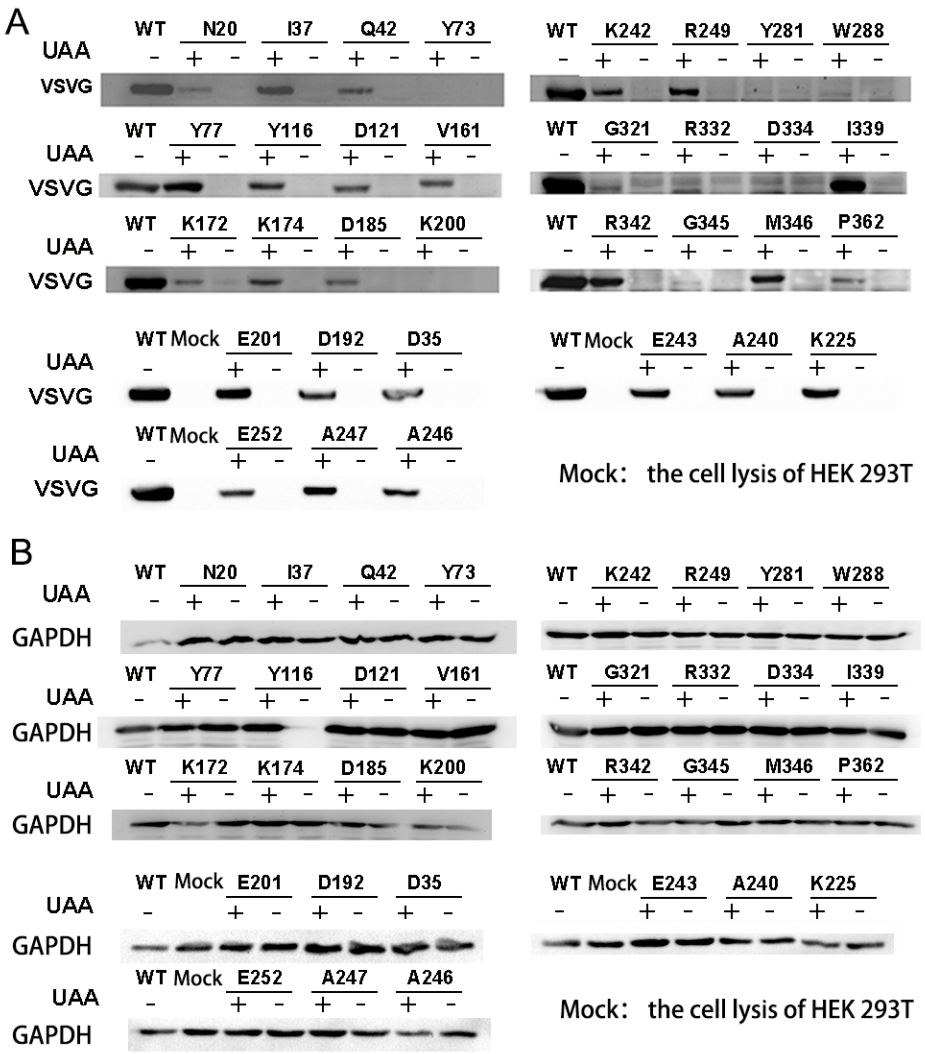
Meas. m/z	I	Formula	m/z	err [mDa]	err [ppm]	mSigma	rdb	e⁻ Conf	N-Rule
260.13595	37785688	C 9 H 18 N 5 O 4	260.13533	-0.62	-2.38	15.1	3.5	even	ok

HIGH RESLUTION MASS SPECTROMETRY REPORT

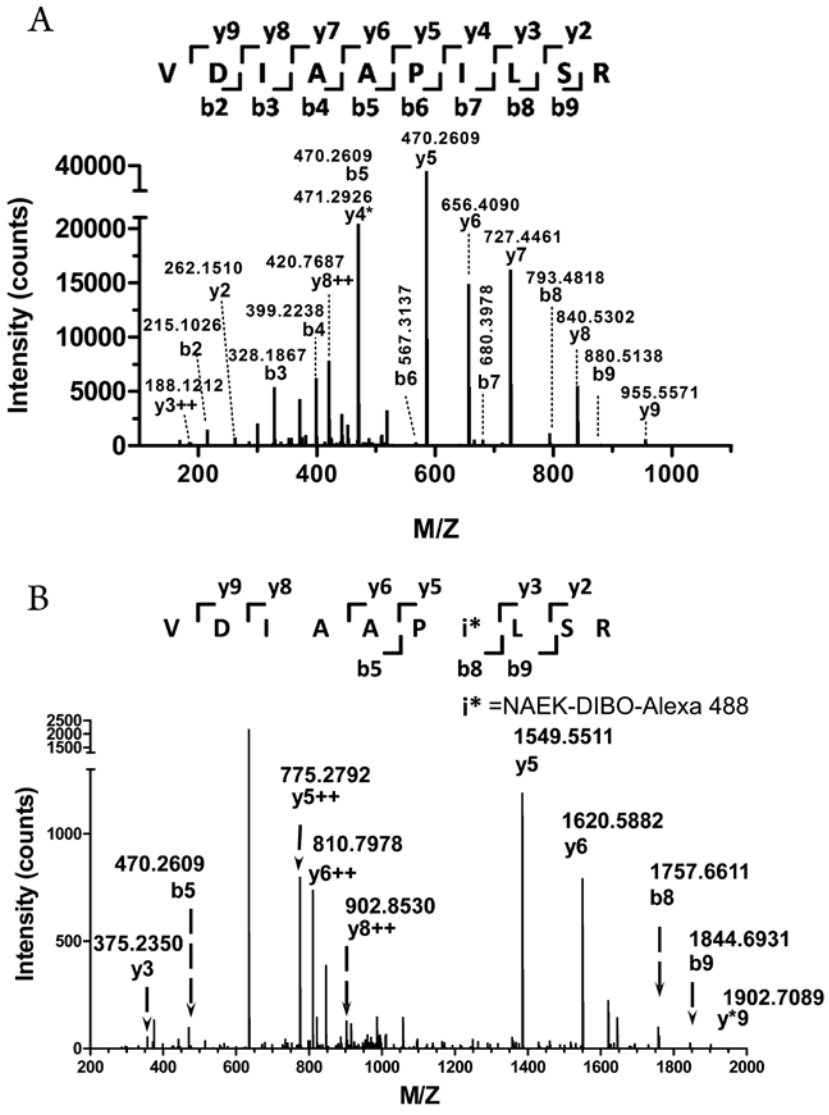
Sample No.	Measured Mass	Exact Mass	Error (m/z)	Error (ppm)	Description
Y-1-UA	308.16937	308.16931	0.00006	0.19470	



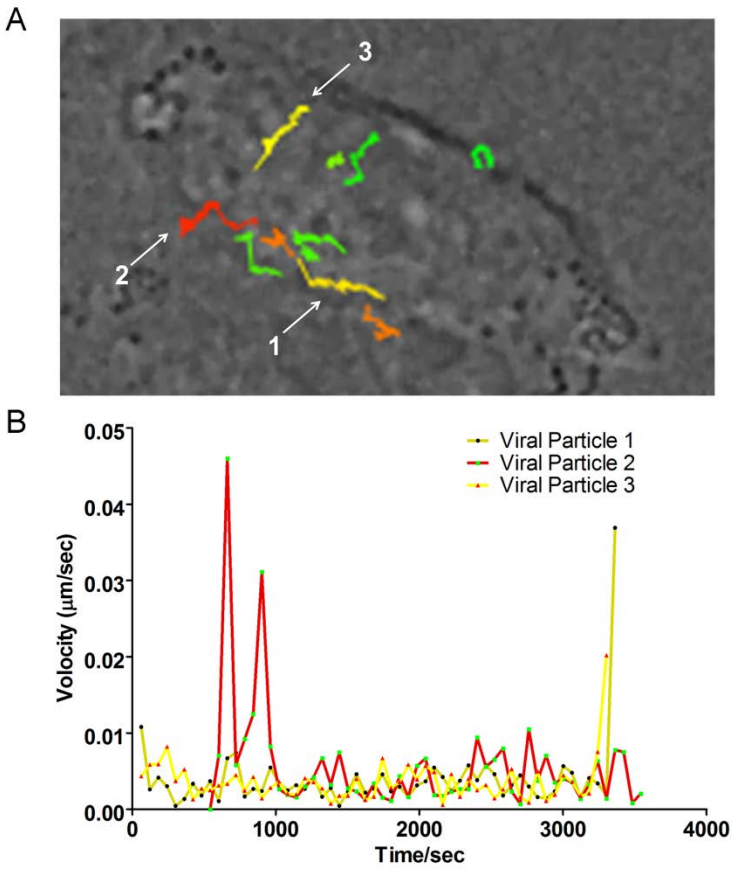
Supplementary Figure S2. Systematic exploitation of genetic code expansion-mediated unnatural amino acid incorporation into lentiviral envelope protein at different sites. (A) The expression level of VSVG was analyzed by SDS-PAGE and detected by western blotting using mouse monoclonal anti-VSVg (Sigma Aldrich). (B) The expression level of GAPDH was analyzed by SDS-PAGE and detected by western blotting using rabbit monoclonal anti-GAPDH as loading controls.



Supplementary Figure S3. LC-MS/MS peptide sequencing of VSVg protein. VSVg proteins with or without NAEK incorporated at different positions were purified by immunoprecipitation (IP), analyzed by SDS-PAGE and visualised with Coomassie blue staining. The corresponding protein bands were sliced and analyzed with LC-MS/MS. (**A**) The partial peptide sequencing of the wide type VSVg protein with each mass peak denoted by the corresponding residue. (**B**) The peptide sequencing of NAEK-containing VSVg protein conjugated with Alexa 488; each mass peak was denoted by the corresponding residue including the incorporated NAEK labeled as i*. As it was confirmed that NAEK was incorporated at I339, the mass difference [$y_5-y_3-m(pro)-m(NAEK)=1549.5511-375.2350-97.12-241.12=836.0761$] indicated the conjugation of DIBO- Alexa 488 to NAEK.

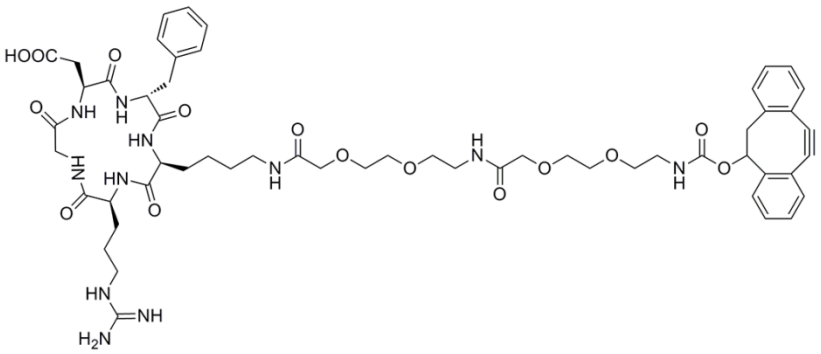


Supplementary Figure S4. Development of a NAEK-containing lentiviral vector as a tool for single-virus tracking. Monitoring (A) and quantitative analysis (B) of the viral motility of different single viruses in different pathways within the infected cell.



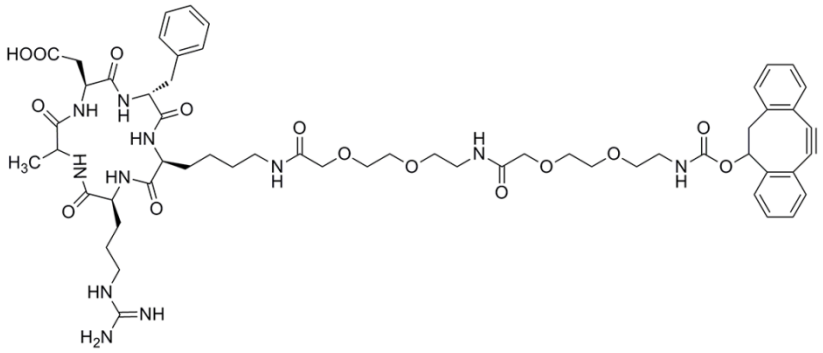
Supplementary Figure S5. Structural representatives of DIBO-RGD and DIBO-RAD used in this study.

c(RGDfK(PEG-PEG))



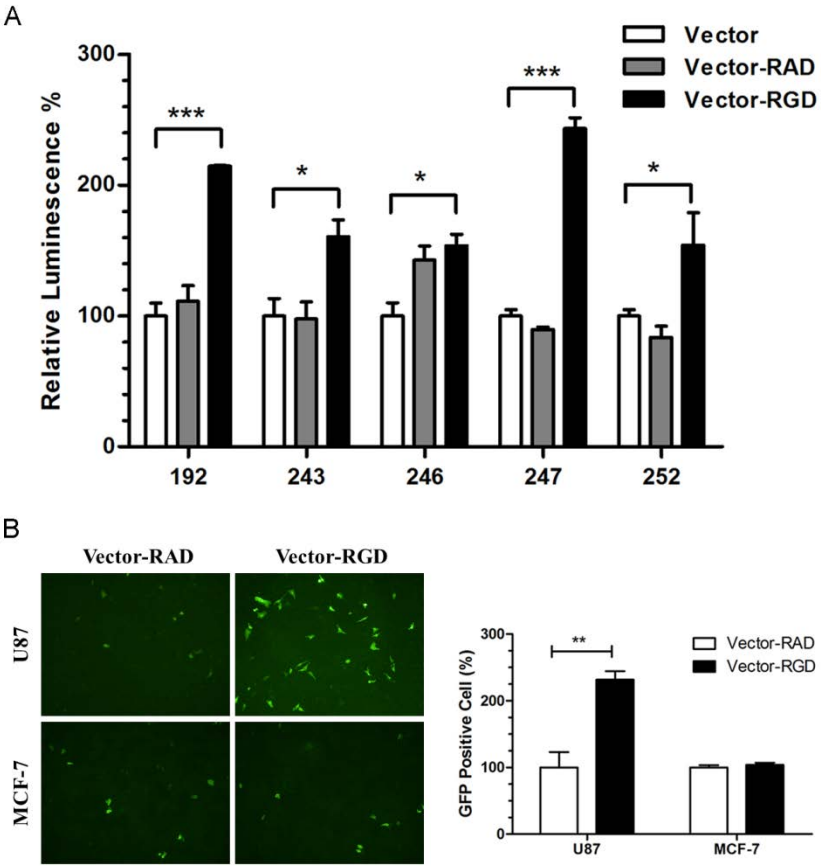
Chemical Formula: $C_{56}H_{73}N_{11}O_{15}$

c(RADfK(PEG-PEG))

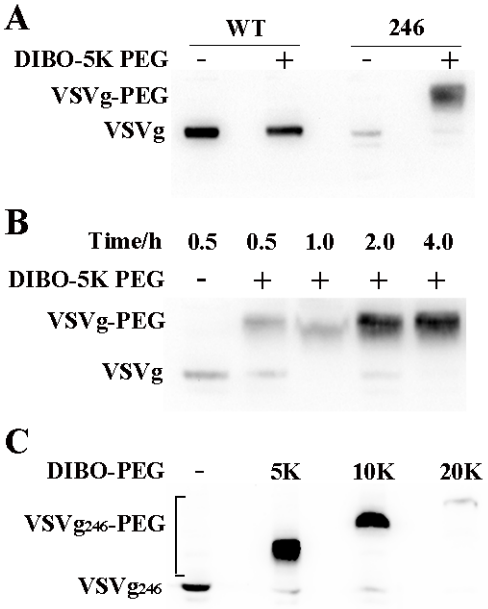


Chemical Formula: $C_{57}H_{75}N_{11}O_{15}$

Supplementary Figure S6. RGD conjugation to lentiviral vectors enhances the lentiviral vector tropism. (A) Characterization of the enhancedinfectivity in U87 cell upon RGD conjugation of lentiviral vectors at different sites, analyzed by one-way ANOVA test (*P < 0.05, *** P<0.001). (B) The RGD-conjugated lentiviral vector, which used green fluorescent protein (GFP) as reporter gene, was much more efficient than the corresponding RAD-conjugated vector in the delivery of these genes into U87 cells (** P<0.01).



Supplementary Figure S7. Site-specific PEGylation of the azide-containing lentiviral vector. (A) Conjugation of DIBO-5K PEG with the azide-containing lentiviral vector at site 246, while the wild type lentiviral vector as a control. The VSVg protein was analyzed by 9% Gel and detected by western blotting. (B) The time-dependent conjugation of DIBO-5K PEG with the azide-containing lentiviral vector; Conjugation occurred within 2 h. (C) Conjugation of the azide-containing lentiviral vector with different sized PEG moieties; the conjugated products were analyzed by western blotting.



Supplementary Figure S8. Display of DiZPK on lentiviral vector for detection of virus-host interactions. (A) Schematic representative of DiZPK-mediated conversion of physical protein-protein interaction to covalently bound protein complex via photo-crossing activation. (B) DiZPK-dependent expression of VSVg protein analyzed by western blotting using monoclonal anti-VSVg. The coding gene bears an amber codon at position Y77 or Y116. GAPDH acts as a loading control. (**C**) Comparisons of the effect of displaying DiZPK at site Y77 and Y116 on propagation and infectivity of lentiviral vectors, which was detected based on the luciferase assay. (D) Two undetermined cellular factors captured by DiZPK-bearing lentiviral vectors, which was analyzed by western blotting.