Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell Type Specific Gene Editing

Romain Rouet[a,b], Benjamin A. Thuma[c], Marc D. Roy[d], Nathanael G. Lintner[e], David M. Rubitski[e], James E. Finley[e], Hanna M. Wisniewska[c], Rima Mendonsa[b,i], Ariana Hirsh[b,i], Lorena de Oñate[b,i], Joan Compte Barrón[b,i], Thomas J. McLellan[c], Justin Bellenger[c], Xidong Feng[c], Alison Varghese[c], Boris A. Chrunyk[c], Kris Borzilleri[c], Kevin D. Hesp[c], Kaihong Zhou[a,b], Nannan Ma[a,b], Meihua Tu[d], Robert Dullea[f], Kim F. McClure[d], Ross C. Wilson[b,i], Spiros Liras[d,1,3], Vincent Mascitti[c,1,2] & Jennifer A. Doudna[a,b,g-j,1,2,3]

[a]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. [b]California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720, USA. [c]Pfizer Medicine Design, Groton, CT 06340, USA. [d]Pfizer Drug Safety R&D, Groton, CT 06340, USA. [e]Pfizer Medicine Design, Cambridge, MA 02139, USA. [f]Pfizer CVMET Biology, Cambridge, MA 02139, USA. [g]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA. [h]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [i]Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA. [j]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[1]Correspondence should be addressed to J.A.D. (doudna@berkeley.edu), V.M. (vincent.mascitti@pfizer.com), or S.L. (spiros.liras@pfizer.com).

[2]Research project leaders

[3]Collaboration leaders

Supporting information (SI)

Table of Contents

Abbreviations

shorthand	*molecule*	*description*
lig	ASGPrL	ASGP receptor ligand; conjugated to Cas9 constructs
AFr	Alexafluor647	Red fluorophore moiety; conjugated to Cas9 constructs
AFg	Alexafluor532	Green fluorophore moiety; conjugated to Cas9 constructs
mCh	mCherry protein	Fluorescent red protein; fused to Cas9 in some constructs

1. Materials and Methods

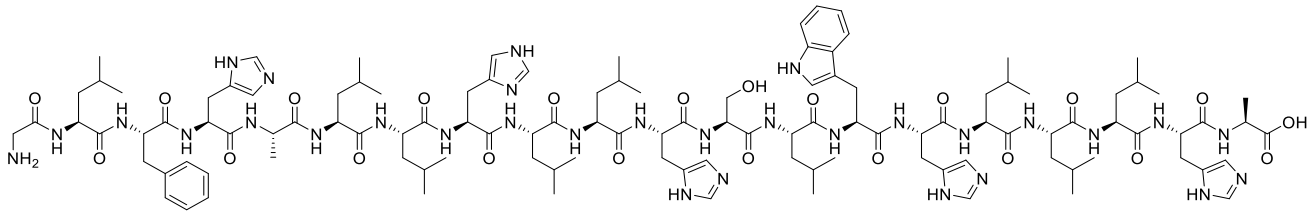
All starting materials not synthesized below were purchased from Sigma Aldrich, Thermo-Fisher (Alexa Fluor™ 647 NHS Ester (Succinimidyl Ester) cat. Number = A37566 and Alexa Fluor™ 532 NHS Ester (Succinimidyl Ester) Cat. Number = A20001), BroadPharm, Quanta biodesign Limited, and Chem-Impex International, Inc. All solvents for small molecule synthesis were reagent grade. Conjugations utilized Zeba™ spin desalting columns and EMD Millipore Amicon™ Ultra Centrifugal Filter Units purchased from Theromo-Fisher and used according to the manufacturer's instructions. **SI 1**, **SI 2** (1) and **ppTG21** (2) were prepared using previously published methods.

Mass spectrometry: An Agilent 6200 series TOF/6500 series Q-TOF or Thermo-LCQ Advantage system was used to collect mass data based on ESI.

2. ppTG21 Synthesis

Method B: Pure peptides were analyzed using a HP1090 system coupled to a Phenomenex C18 (2) (5 microns, 100 Å, 4.6 mm × 150 mm) reversed phase HPLC column eluting with a solvent gradient of A:C, where A = 0.1% TFA in water and C = 0.09%TFA in acetonitrile:water (4:1) over 20 min at a flow rate of 1.0 mL/min.

ppTG21: GLFHALLHLLHSLWHLLLHA**:**



ppTG21: GLFHALLHLLHSLWHLLLHA-OH

Fmoc-Ala-Wang resin (5.0 mmol, 10 g) was placed in a peptide reactor, and the resin was swelled in DMF for 2 h. Then, the Fmoc group was removed by addition of a 20% (by volume) solution of piperidine in DMF (150 mL) followed by 1 min of agitation. This treatment was repeated 5 times. A Kaiser ninhydrin test was performed to demonstrate complete deprotection. A solution of Fmoc-His(Trt)-OH (15 mmol, 9.3 g) and HBTU (14 mmol, 5.4 g) in DMF (~40 mL) was treated with *N*-methylmorpholine (30 mmol, 3.3 mL) at 0 °C, and the mixture was kept at 0 °C for 15 min. This solution was then added to the H-Ala-Wang resin, and mixture was stirred at 25 °C for 1 h, at which point the Kaiser ninhydrin test indicated the reaction was complete. The mixture was filtered, and the solid was washed with DMF (5 × 150 mL). The resulting Fmoc-His(Trt)-Ala-Wang resin product was used in subsequent step without further treatment.

After Fmoc deprotection of the peptidyl resin, Fmoc-amino acids were coupled to the resin bound peptide sequentially using the standard amide coupling/FMOC cleavage method described above to deliver H-Gly-Leu-Phe-His(Trt)-Ala-Leu-Leu-His(Trt)-Leu-Leu-His(Trt)-Ser(tBu)-Leu-Trp(Boc)-His(Trt)-Leu-Leu-Leu-His(Trt)-Ala-Wang resin. The peptidyl resin was washed with MeOH (2 × 150 mL), dichloromethane (2 × 150 mL) and MeOH (2 × 150 mL). The resin was dried under vacuum overnight. A solution of TFA: thioanisole: phenol: EDT: H_2O (87.5: 5: 2.5: 2.5: 2.5, 650 mL) was added to the peptidyl resin, and the resulting suspension was shaken for 2.5 h and filtered. Ether (5 L) was added to the filtrate which afforded a solid. The mixture was centrifuged, and the ether layer was decanted. The resulting solid was washed with ether (3×) and dried *in vacuo* overnight. The resulting crude material was then purified via reverse phase HPLC, like fractions were combined, and lyophilized to deliver 5.2 g of the desired peptide (6TFA salt) as a white solid. UV purity (220

nm) = 95.4% (Method B, retention time = 9.22 min, solvent gradient A:B, 24:76–14:86), ESI (*m/z*) 2341.3430 $(M+H)^+$.

Small molecule LCMS and HRMS conditions

Note: Structures of all alexafluor647-containing reagents are not disclosed by the vendor; as such, any compounds containing this group do not have a predicted mass. Predicted mass of protein constructs is based on the observed masses for the starting ligands.

HRMS Method A: The sample analysis was carried out on an Agilent 6530 QTof mass spectrometer equipped with a Dual AJS electrospray source operated in negative ion mode. The sample was diluted to 2.5 µM with 0.1% formic acid in 50:50 water:acetonitrile. The sample was then infused directly into the instrument. Raw mass spectra were viewed using MassHunter (version B.07.00 Service Pack 2, Agilent).

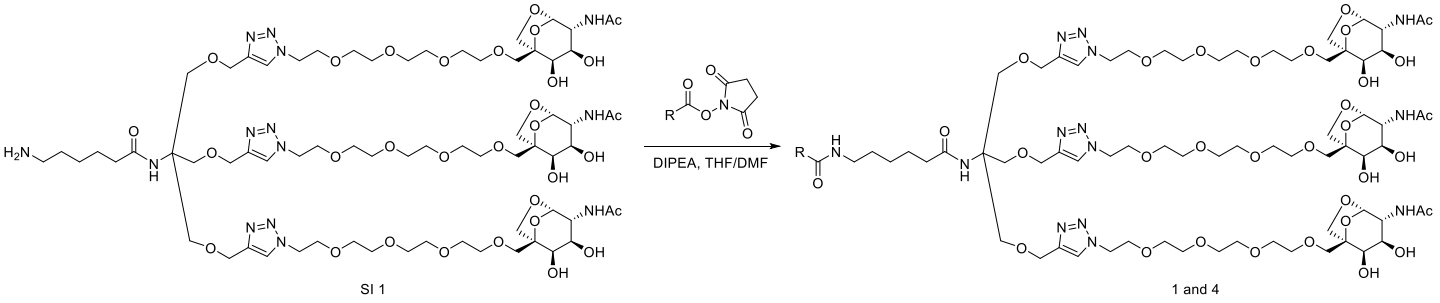
HRMS Method B: The sample analysis was carried out on an Agilent 6530 QTof mass spectrometer equipped with a Dual AJS electrospray source operated in positive ion mode. The mass spectrometer was interfaced with an Agilent 1290 UPLC system. The Agilent 1290 autosampler injected 10 µL aliquots of sample which was diluted to 5 mM in MilliQ water just prior to analysis. The material was separated using a Agilent PLRP-S 100 Å 50 × 2.1 mm with 3.0 µm particles column (part no. PL1912-1300). The mobile phases were: A) 0.1% formic acid in water and B) 0.1% formic acid in acetonitrile. Raw mass spectra were viewed using MassHunter (version B.07.00 Service Pack 2, Agilent).

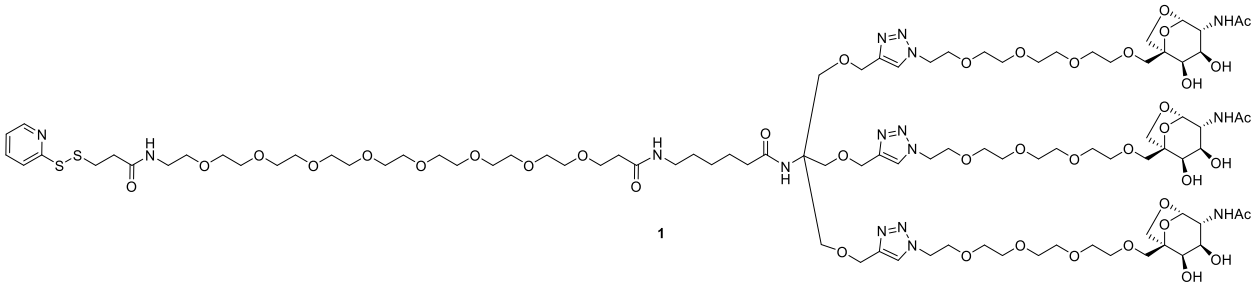
Method C 1.5 min LRMS (low resolution mass spectroscopy): Waters Acquity HSS T3, 2.1 mm × 50 mm, C18, 1.7 µm; Mobile Phase: A: 0.1% formic acid in water (v/v); Mobile phase B: 0.1% formic acid in acetonitrile (v/v); Flow-1.25 ml/min; Initial conditions: A-95%:B-5%; hold at initial from 0.0–0.1 min; Linear Ramp to A-5%:B-95% over 0.1–1.0 min; hold at A-5%:B-95% from 1.0–1.1 min; return to initial conditions 1.1–1.5 min.

Method C 3.0 min LRMS (low resolution mass spectroscopy): Waters Acquity HSS T3, 2.1mmx50mm, C18, 1.7µm; Mobile Phase: A: 0.1% formic acid in water (v/v); Mobile phase B: 0.1% formic acid in acetonitrile (v/v); Flow-1.25 ml/min; Initial conditions: A-95%:B-5%; hold at initial from 0.0–0.1 min; Linear Ramp to A-5%:B-95% over 0.1–2.6 min; hold at A-5%:B-95% from 2.6–2.95 min; return to initial conditions 2.95–3.0 min.

3. Ligand Synthesis

Scheme 1. Synthesis of Compounds 1 and 4





1

N-(1,3-bis[(1-{1-[(1S,2R,3R,4R,5S)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1H-1,2,3-triazol-4-yl)methoxy]-2-{[(1-{1-[(1S,2R,3R,4R,5S)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1H-1,2,3-triazol-4-yl)methoxy]methyl}propan-2-yl)-3,31-dioxo-1-(pyridin-2-yldisulfanyl)-7,10,13,16,19,22,25,28-octaoxa-4,32-diazaoctatriacontan-38-amide (**1**)

To a solution of 6-amino-N-(1,3-bis[(1-{1-[(1S,2R,3R,4R,5S)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1H-1,2,3-triazol-4-yl)methoxy]-2-{[(1-{1-[(1S,2R,3R,4R,5S)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1H-1,2,3-triazol-4-yl)methoxy]methyl}propan-2-yl)hexanamide acetate salt **(SI 1)** (70 mg, 0.041 mmol) and N-{27-[(2,5-dioxopyrrolidin-1-yl)oxy]-27-oxo-3,6,9,12,15,18,21,24-octaoxaheptacos-1-yl}-3-(pyridin-2-yldisulfanyl)propanamide (QuantaBiodesign LTD, CAS# = 1252257-56-9, 30 mg, 0.041 mmol) in *N*,*N*-dimethylformamide (0.6 mL) and tetrahydrofuran (0.6 mL) was added *N*,*N*-diisopropylethylamine (0.029 mL, 0.16 mmol). The reaction was allowed to stir at room temperature. After 18 h, the reaction mixture was concentrated under reduced pressure. The crude material was purified using reverse-phase chromatography using the conditions below yielding the title compound **(1)** as a gum (59 mg, 64%).

^{1}H NMR (METHANOL-d$_4$) δ: 8.47 (d, *J*=5.1 Hz, 1H), 8.01 (s, 3H), 7.92 (d, *J*=3.5 Hz, 2H), 7.39–7.30 (m, 1H), 5.21 (s, 3H), 4.62–4.57 (m, 6H), 4.57 (s, 6H), 3.99–3.92 (m, 6H), 3.92–3.86 (m, 9H), 3.77 (s, 9H), 3.74–3.69 (m, 6H), 3.68–3.50 (m, 73H), 3.14 (t, *J*=7.0 Hz, 2H), 3.10 (t, *J*=7.0 Hz, 2H), 2.64 (t, *J*=6.8 Hz, 2H), 2.43 (t, *J*=6.0 Hz, 2H), 2.17 (t, *J*=7.4 Hz, 2H), 1.99 (s, 9H), 1.63–1.53 (m, 2H), 1.52–1.42 (m, 2H), 1.32 (dt, *J*=15.1, 7.5 Hz, 2H); HRMS method A in positive ion mode (ESI) calcd for $C_{97}H_{162}N_{16}O_{41}S_2$ (*m/z*) $[M + 2H]^{2+}$ 1,136.525, found 1136.532

Purification Conditions

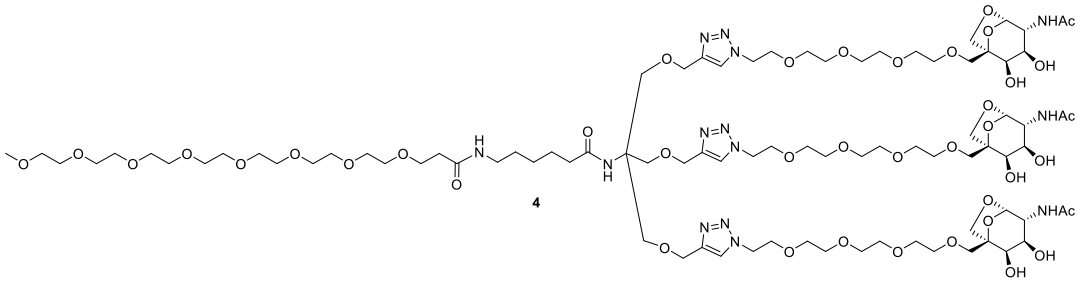
The residue was dissolved in dimethyl sulfoxide (1 mL) and purified by reversed-phase HPLC Column: Waters Sunfire C18 19 × 100 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v);

Gradient: 80.0% H_2O/20.0% Acetonitrile linear to 70% H_2O/30% Acetonitrile in 8.5 min to 0% H_2O/100% MeCN to 9.0min, HOLD at 0% H_2O / 100% Acetonitrile from 9.0 to 10.0min. Flow: 25 mL/min.

QC conditions

Column: Waters Atlantis dC18 4.6 × 50 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v);

Gradient: 95.0% H_2O/5.0% Acetonitrile linear to 5% H_2O/95% Acetonitrile in 4.0 min, HOLD at 5% H_2O/95% Acetonitrile to 5.0 min. Flow: 2 mL/min. Retention time = 1.85 min



N-(1,3-bis[(1-{1-[(1S,2R,3R,4R,5S)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1*H*-1,2,3-triazol-4-yl)methoxy]-2-{[(1-{1-[(1S,2R,3R,4R,5S)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1*H*-1,2,3-triazol-4-yl)methoxy]methyl}propan-2-yl)-26-oxo-2,5,8,11,14,17,20,23-octaoxa-27-azatritriacontan-33-amide (**4**)

To a solution of 6-amino-N-(1,3-bis[(1-{1-[(1*S*,2*R*,3*R*,4*R*,5*S*)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1H-1,2,3-triazol-4-yl)methoxy]-2-{[(1-{1-[(1*S*,2*R*,3*R*,4*R*,5*S*)-4-(acetylamino)-2,3-dihydroxy-6,8-dioxabicyclo[3.2.1]oct-1-yl]-2,5,8,11-tetraoxatridecan-13-yl}-1H-1,2,3-triazol-4-yl)methoxy]methyl}propan-2-yl)hexanamide **(SI 1)** (70 mg, 0.042 mmol) and 1-[(26-oxo-2,5,8,11,14,17,20,23-octaoxahexacosan-26-yl)oxy]pyrrolidine-2,5-dione (31 mg, 0.061 mmol, BroadPharm, CAS# = 756525-90-3) in *N*,*N*-dimethylformamide (1 mL) and tetrahydrofuran (1 mL) was added *N*,*N*-diisopropylethylamine (0.04 mL, 0.21 mmol). After 18 h, the reaction mixture was concentrated under reduced pressure. The crude material was purified using reverse-phase chromatography using the conditions below yielding the title compound (**4**) as a gum (23 mg, 27%).

Method C 3.0 min LRMS (ESI) calcd for $C_{88}H_{152}N_{14}O_{40}$ (*m/z*) $[M + 2H]^{2+}$1,023.5, found 1,023.2

retention time = 1.01 min; ^{1}H NMR (METHANOL-d_4) δ: 8.00 (s, 3H), 5.21 (s, 3H), 4.66–4.46 (m, 12H), 3.99–3.92 (m, 6H), 3.89 (dd, *J*=11.3, 4.7 Hz, 9H), 3.80–3.74 (m, 9H), 3.74–3.68 (m, 6H), 3.68–3.49 (m, 69H), 3.36 (s, 3H), 3.15 (t, *J*=7.0 Hz, 2H), 2.43 (t, *J*=6.2 Hz, 2H), 2.18 (t, *J*=7.4 Hz, 2H), 1.99 (s, 9H), 1.62–1.53 (m, 2H), 1.52–1.44 (m, 2H), 1.39–1.27 (m, 2H); HRMS method A:in positive mode calcd for $C_{88}H_{152}N_{14}O_{40}$ $[M + 2H]^{2+}$ 1,023.5225, found 1,023.5184

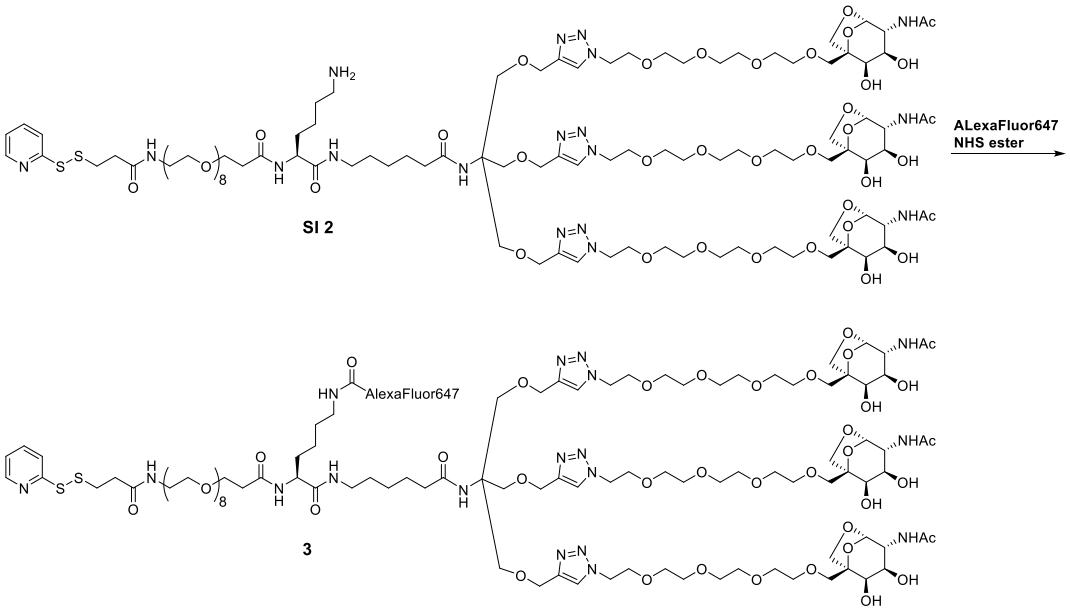
Purification Conditions

The residue was dissolved in dimethyl sulfoxide (1 mL) and purified by reversed-phase HPLC (Column: Waters Sunfire C18 19×100 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v); 80.0% H_2O/20.0% Acetonitrile hold to 80% H_2O/20% Acetonitrile in 10.5 min, 80% H_2O/20% Acetonitrile linear to 0% H_2O/100% MeCN in 0.5 min, HOLD at 0% H_2O/100% Acetonitrile from 11.0–12.0 min. Flow: 25 mL/min.

QC conditions

Column: Waters Atlantis dC18 4.6 × 50 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v); 95.0% H_2O/5.0% Acetonitrile linear to 5% H_2O/95% Acetonitrile in 4.0 min, HOLD at 5% H_2O/95% Acetonitrile to 5.0 min. Flow: 2 mL/min. Retention time = 1.71 min

Scheme 2. Synthesis of Compounds 3



Branched ASGPr /AlexaFluor647 Ligand (**3**)

A solution of **SI 2** (23 mg, 9.1 µmol, 1.0 equiv) in water/DMSO (0.46 mL, 1:1) was added to a solution of AlexaFluor647 NHS ester (12 mg, 9.5 µmol, 1.1 equiv) in DMSO (0.32 mL). Diisopropylamine (16 µL, 91 µmol, 10 equiv) was then added and the reaction was stirred at room temperature protected from light. After 1 h the reaction was concentrated and the resultant residue purified by reverse-phase chromatography using the conditions below yielding the title compound as a deep blue glassy solid (14 mg, 44%).

HRMS Method A in negative Ion mode (ESI) found $[M - 3H]^{3-}$ = 1078.7616

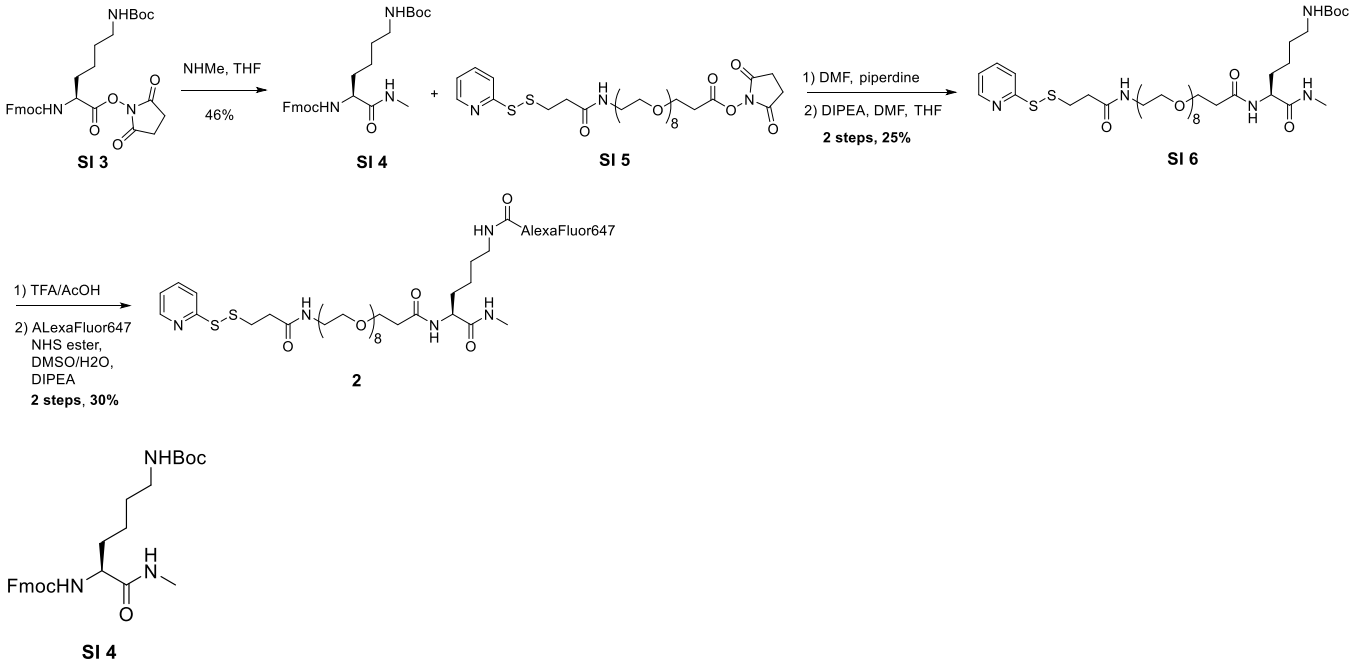
Purification Conditions

The residue was dissolved in dimethyl sulfoxide (1 mL) and purified by reversed-phase HPLC (Column: Waters Sunfire C18 19 × 100 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v); 80.0% H_20/20.0% Acetonitrile linear to 70% H_20/30% Acetonitrile in 8.5 min, 70% H_2O/30% Acetonitrile linear to 0% H_2O/100% MeCN in 0.5 min, HOLD at 0% H_20/100% Acetonitrile to 10.0 min. Flow: 25 mL/min.

QC conditions

Column: Waters Atlantis dC18 4.6 × 50 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v); 95.0% H_20/5.0% Acetonitrile linear to 5% H_20/95% Acetonitrile in 4.0 min, HOLD at 5% H_20/95% Acetonitrile to 5.0 min. Flow: 2mL/min. Retention time = 2.1 min

Scheme 3. Synthesis of Compounds 2



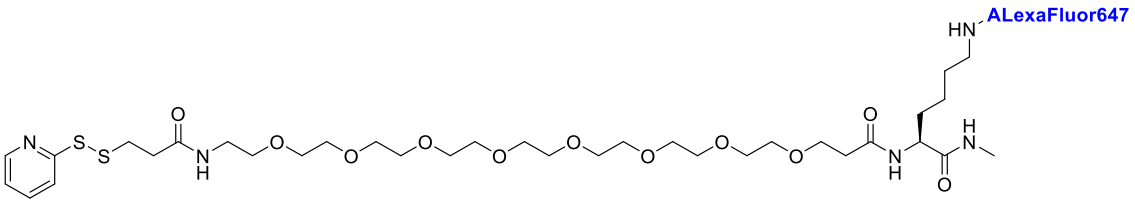
9H-fluoren-9-ylmethyl [(2S)-6-[(tert-butoxycarbonyl)amino]-1-(methylamino)-1-oxohexan-2-yl]carbamate (**SI 4**).

N,*N*-diisopropylethylamine (0.50 g, 3.9 mmol, 0.70 mL) was added to a solution of 2,5-dioxopyrrolidin-1-yl N~6~-(tert-butoxycarbonyl)-N~2~-[(9H-fluoren-9-ylmethoxy)carbonyl]-L-lysinate **SI 3** (Chem-Impex International, Inc. CAS# 132307-50-7, 0.44 g, 0.77 mmol) and methylamine hydrochloride (52 mg, 0.77 mmol) in tetrahydrofuran (5 mL) and the colorless, homogenous solution was stirred at room temperature. After 20 h, the reaction mixture was concentrated under reduced pressure and the crude material was purified by flash chromatography (ISCO, RediSepGold 12 g column, 0–20% MeOH in DCM) to afford the title compound **(SI 4)** as an oil (0.17 g, 46%). ^{1}H NMR (METHANOL-d_4) δ: 7.80 (d, *J*=7.8 Hz, 2H), 7.67 (t, *J*=6.2 Hz, 2H), 7.44–7.36 (m, 2H), 7.35–7.27 (m, 2H), 4.48–4.32 (m, 2H), 4.17–4.26 (m, 1H), 4.00 (dd, *J*=9.0, 5.1 Hz, 1H), 3.03 (q, *J*=6.5 Hz, 2H), 2.72 (s, 3H), 1.84–1.68 (m, 1H), 1.66–1.54 (m, 1H), 1.42 (s, 9H), 1.51–1.39 (m, 2H), 1.38–1.27 (m, 2H); Method C 1.5 min LRMS (ESI) calcd for $C_{27}H_{35}N_3O_5$ (*m/z*) $[M + H]^+$ 482.3, found 482.5, retention time = 0.95 min.

N~6~-(tert-butoxycarbonyl)-N-methyl-N~2~-[29-oxo-31-(pyridin-2-yldisulfanyl)-4,7,10,13,16,19,22,25-octaoxa-28-azahentriacontan-1-oyl]-L-lysinamide (**SI 6**).

To a solution of compound **SI 4** (0.16 g, 0.33 mmol) in tetrahydrofuran (15 mL) was added piperidine (1.5 mL, 15 mmol) and the reaction was stirred at room temperature overnight. Water was added to the reaction mixture to azeotrope the piperidine and the reaction was concentrated. This was repeated 3 times. Toluene was then added to the reaction mixture to azeotrope the water and the reaction was concentrated. This was repeated 3 times. The crude material was used in the subsequent step without purification. A solution of the resulting crude material (84 mg, 0.33 mmol) in *N*,*N*-dimethylformamide (1.0 mL) was prepared and added to a solution of *N*-{27-[(2,5-dioxopyrrolidin-1-yl)oxy]-27-oxo-3,6,9,12,15,18,21,24-octaoxaheptacos-1-yl}-3-(pyridin-2-yldisulfanyl)propanamide (**SI 5**, MFCD13185003, 0.36 g, 0.49 mmol) in tetrahydrofuran (10 mL) and *N*,*N*-

diisopropylethylamine (0.30 mL, 1.6 mmol) and the reaction was stirred at room temperature for 1.5 h. The sample was concentrated under reduced pressure. The resulting crude material was purified by flash chromatography (ISCO, RediSepGold 12 g column, 0–20% MeOH in DCM) to afford the title compound as an oil (71 mg, 25%). Method C 1.5 min LRMS (ESI) calcd for $C_{39}H_{69}N_5O_{13}S_2$ (*m/z*) $[M + H]^+$ 880.441, found 880.8, retention time = 0.78 min; ^1H NMR (400MHz, METHANOL-d_4) δ = 8.41 (d, *J*=4.7 Hz, 1H), 7.88–7.78 (m, 2H), 7.23 (ddd, *J*=1.6, 5.0, 6.7 Hz, 1H), 4.27 (dd, *J*=5.1, 9.0 Hz, 1H), 3.78–3.71 (m, 2H), 3.67–3.58 (m, 27H), 3.58–3.51 (m, 2H), 3.37 (t, *J*=5.3 Hz, 2H), 3.11–2.97 (m, 4H), 2.68 (s, 3H), 2.63 (t, *J*=7.0 Hz, 2H), 2.51 (t, *J*=6.0 Hz, 2H), 1.87–1.73 (m, 1H), 1.62 (dtd, *J*=4.9, 9.3, 13.9 Hz, 1H), 1.52–1.45 (m, 3H), 1.43 (s, 9H), 1.40–1.24 (m, 2H); HRMS Method B (ESI) calcd for $C_{39}H_{69}N_5O_{13}S_2$ (*m/z*) $[M + H]^+$ 880.441, found 880.440.



Dithiopyridine-dPEG-Lys-N-6-AlexaFluor647 (**2**)

SI 6 (32 mg, 0.036 mmol) was dissolved in a mixture of trifluoroacetic acid (0.45 mL, 6.0 mmol) and acetic acid (0.85 mL) at room temperature. After stirring for 6 h, the reaction was concentrated under reduced pressure. A portion of the crude material was carried forward to the next step. The conversion of the Boc deprotection is quantitative by LCMS. To a solution of crude (3.1 mg, 3.5 µmol) in water/DMSO (1:1, 2 mL) was added *N*,*N*-diisopropylethylamine (5.6 µL, 32 µmol) followed by the addition of a solution of AlexaFluor647 NHS ester (4.0 mg, 3.0 µmol, 0.11 mL, 30 mM) in DMSO. The reaction was stirred at room temperature protected from light for 1 h. The reaction was concentrated under reduced pressure and resultant residue purified by reverse-phase chromatography using the conditions below yielding the title compound as a deep blue gum (1.5 mg, 23% over 2 steps, where Boc deprotection is assumed to be quantitative). HRMS method B: (ESI) found $[M + 2H]^{2+}$ = 810.7808

Purification Conditions

The residue was dissolved in dimethyl sulfoxide (1 mL) and purified by reversed-phase HPLC (Column: Waters XBridge C4 19 × 100 mm, 5 µm; Mobile phase A: 0.05% TEAA in water (v/v); Mobile phase B: 0.05% TEAA in acetonitrile (v/v); 90.0% H_2O/10.0% Acetonitrile linear to 60% H_2O/40% Acetonitrile in 8.5 min, HOLD at 0% H_2O/100% Acetonitrile to 10.0 min. Flow: 25 mL/min.

Note: TEAA (Triethylammonium Acetate) Collection triggered by mass at 811.39 m/z

QC conditions

Column: Waters Atlantis dc18 4.6 × 50 mm, 5 µm; Mobile phase A: 0.05% TFA in water (v/v); Mobile phase B: 0.05% TFA in acetonitrile (v/v); 90.0% H_2O/10.0% Acetonitrile HOLD for 1 min, then linear to 5.00% H_2O/95.0% Acetonitrile from 1.0– 4.0 min, HOLD at 5.0% H_2O/95.0% Acetonitrile to 5.0 min. Flow: 2 mL/min. Retention time = 1.52 min.

4. Protein Conjugation

Buffer preparation

Sodium chloride Gel filtration (GF) buffer

Glycerol (12.6 g, 1.37 mol, 100 mL, MFCD00004722) was added to a 1 L bottle equipped with a magnetic stir bar and to which was added water 700 mL (BioWhittaker water for cell culture application, (Lonza)), followed by the addition of 5 M solution saline (30 mL, 5 M) and HEPES (4.77 g, 20.0 mmol) with magnetic stirring. Water ((BioWhittaker water for cell culture application, (Lonza))) was added to produce a 1 L solution. The pH of the solution was adjusted using 15 M sodium hydroxide solution to a pH = 7.45. The solution was filtered through a 0.2 µm filter.

Potassium chloride Gel filtration (GF) buffer

Glycerol (12.6 g, 1.37 mol, 100 mL, MFCD00004722) was added to a 1 L bottle equipped with a magnetic stir bar and to which was added water 700 mL (BioWhittaker water for cell culture application, (Lonza)), followed by the addition of potassium chloride (MFCD00011360, bioultra for molecular biology >99.5% 11 mg, 0.15 mmol) and HEPES (4.77 g, 20.0 mmol) with magnetic stirring. Water ((BioWhittaker water for cell culture application, (Lonza))) was added to produce a 1 L solution. The pH of the solution was adjusted using 10 M potassium hydroxide solution to a pH = 7.45. The solution was filtered through a 0.2 µm filter.

Protein sequences

Cas9 (M1C/C80S) mCherry

SNAT**C**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRISYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY
VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE**C**FDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV
DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGDAYPYDVPDYASLGSGSPKKKRKVEDPKKKRKVDGIGSGSNGSSGSVSKGEEDNM
AIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIP
DYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYP
EDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGG
MDELYKGSPKKKRKVE

Cas9 (M1C/C80S) 1NLS

SNAT**C**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRISYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY
VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE**C**FDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV
DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD
NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGDAYPYDVPDYASLGSGSPKKKRKVD

Expression and purification of Cas9-mCherry and Cas9-1NLS

Cas9 mCherry expression and purification

Cas9-mCherry (Cas9-mCh) fusion protein was expressed and purified as previously described (3), with modifications to remove endotoxins. After capture of His-MBP tagged Cas9 onto Ni-NTA resin (Qiagen), the resin was washed with 20 column volumes of 20 mM HEPES pH 7.5 with 500 mM KCl, 20 mM imidazole, 0.1% Triton X-114 and 10% glycerol, followed by 20 column volumes of 20 mM HEPES pH 7.5 with 500 mM KCl, 20 mM imidazole, and 10% glycerol. Ion-exchange chromatography (HiTrap Heparin, GE Healthcare) was carried out identically; however, size exclusion chromatography (HiLoad Superdex 200 16/60, GE Healthcare) was performed in 20 mM HEPES buffer pH 7.5 with 150 mM KCl and 10% glycerol, and the column was cleaned with 0.5 M NaOH between purifications to remove endotoxin. Protein was sterile filtered, flash frozen, and stored at –80 °C.

Cas9-1NLS expression and purification

Cas9-1NLS protein was purified as above, but with the following changes. In place of Ni-NTA resin, 5 mL HisTrap FF columns were used (GE Healthcare). While the protein was bound to the column, the Triton X-114 wash was performed at room temperature. Size exclusion chromatography was performed identically but using 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, 10% (v/v) glycerol, and the eluted protein samples flash frozen in this buffer before being stored at –80 °C.

Conjugation Procedure

Preparative CEX was typically performed on a Waters H-Class Acquity UPLC using a Sepax SCN-NP5-5µm column. Mobile Phase A: 20 mM Hepes,150 mM NaCl, 10% glycerol (pH = 7.4) and Mobile Phase B: 20 mM HEPES, 1.0 M NaCl, 10% glycerol (pH = 7.4).
The flow rate was 0.7–1.0 mL/min at room temperature over 40–55 min. UV triggered Fraction Collection at 280 nm.

General Procedure for Protein Ligation

The starting protein was passed through a 0.22 µm cellulose acetate spin-X centrifuge tube filter and centrifuged at 15000 × g for 2 min. The protein was buffer exchanged using a Zeba spin filter (40 K MWCO) equilibrating with either KCl GF buffer or NaCl GF buffer using the manufacturer's instructions to remove the TCEP. After buffer exchange, ligand was added in a 10% DMSO/GF buffer solution (10–40 equiv., 3.0 mM) with gentle mixing. The resulting reaction mixture was allowed to incubate for an appropriate time at room temperature, then desalted using Zeba spin filters (40 K MWCO) equilibrating with either KCl or NaCl GF buffer using the manufacturer's instructions to remove excess ligand. The reaction was either submitted to purification without further manipulations or concentrated using EMD Millipore Amicon™ Ultra Centrifugal Filter Units (50 K MWCO) to ~9–10 mg/mL.

After cationic exchange chromatography, isolated samples were diluted with GF buffer (KCl or NaCl) and added to a pre-washed EMD Millipore Amicon™ Ultra Centrifugal Filter Units (50 K MWCO) and concentrated (1000 to 1500 × g for 5 min intervals, pipetting gently to mix after each interval to an appropriate concentration. Final products could then be buffer exchanged using a Zeba spin filter (40 K MWCO) equilibrating with NaCl GF buffer using the manufacturer's instructions.

<u>Conjugate A: Cas9-2lig-1NLS</u>

Conjugate A was prepared in multiple batches with a combined total of 50.6 mg starting protein. A representative example of a batch is listed below.

The starting protein (12 mg) was treated with a solution of **1** (3.0 mM, 40 equiv.) and incubated for 2 h. The crude material was purified using preparative CEX using a Sepax SCN-NP5 7.8 × 250 mm 5 µm column on a Waters H-Class Acquity UPLC. Mobile Phase A: 20 mM HEPES, 150 mM KCl, 10% glycerol, pH = 7.4;

Mobile Phase B: 20 mM HEPES, 1.0 M KCl, 10% glycerol, pH =7.4. Injection volumes = 3 × 800 µL; purification gradient: 10% A hold for 14 min, linear to 13% A from 14–22 min, linear to 45% from 22–28 min. Flow rate 0.7 ml/min. Fractions were collected between 13.25–17.25 min. The pooled fractions containing the desired product were concentrated yielding 345 µL. The concentration was determined via UV absorbance (56 µM, 9.3 mg/mL) yielding 3.2 mg (25%) of the desired conjugate. Combined isolated conjugate A (14mg) was buffer exchanged into NaCl GF buffer yielding 1.3 mL. The concentration was determined via UV absorbance (50 µM, 8.3 mg/mL) yielding 10 mg (20%) of the desired conjugate.

Conjugate B: Cas9-AFr-1NLS

Conjugate B was prepared in multiple batches with a combined total of 41 mg starting protein. A representative example of a batch is listed below.

The starting protein (20 mg) was treated with a crude solution of **2** (3.0 mM, 20 equiv.) and incubated for 2 h. The crude material was purified using preparative CEX using a Sepax SCN-NP5 10 × 250 mm 5 µm column on a Waters H-Class Acquity UPLC. Mobile Phase A: 20 mM HEPES, 150 mM NaCl, 10% glycerol, pH = 7.4; Mobile Phase B: 20mM HEPES, 1.0M NaCl, 10% glycerol, pH =7.4. Injection volumes = 3 × 800 µL; purification gradient: 21% A hold for 18 min, linear to 26% A from 18–27.5 min, linear to 45% from 27.5–35 min. Flow rate 1.0 ml/min. Fractions were collected between 14–20 min. The pooled fractions containing the desired product were concentrated to 908 µL and the concentration was determined via UV absorbance (29 µM, 4.8 mg/mL, Degree Of Ligation [DOL] = 1.9) yielding 4.4 mg (21%) of the desired conjugate. Combined isolated conjugate B (14 mg) was buffer exchanged into NaCl GF buffer yielding 1.73 mL. The concentration was determined via UV absorbance (25 µM, 4.1 mg/mL, DOL = 1.8) yielding 7.1 mg (17%) of the desired conjugate.

Conjugate C: Cas9-2lig-AFr-1NLS

The starting protein (9.1 mg) was treated with a solution of **3** (3.0 mM, 38 equiv.) and incubated for 1.5 h. The crude material was purified using preparative CEX using a Sepax SCN-NP5 7.8 × 250 mm 5 µm column on a Waters H-Class Acquity UPLC. Mobile Phase A: 20 mM HEPES, 150 mM NaCl, 10% glycerol, pH = 7.4; Mobile Phase B: 20 mM HEPES, 1.0 M NaCl, 10% glycerol, pH =7.4. Injection volumes = 2 × 990 µL; purification gradient: 20% A hold for 14 min, linear to 25% A from 14–22 min, linear to 45% from 22–28 min. Flow rate 0.7 ml/min. Fractions were collected between 13–16 min. The pooled fractions containing the desired product were concentrated and buffer exchanged into NaCl GF buffer yielding 308 µL. The concentration was determined via UV absorbance (20 µM, 3.3 mg/mL, DOL = 1.9) yielding 1.0 mg (11%) of the desired conjugate.

Conjugate G: Cas9-2lig-mCh

Compound (**1**) was dissolved at 8 mM in DMSO. Cas9-mCh was filtered through a 0.2 µm filter prior to conjugation. Conjugation was carried out using a 20:1 molar ratio of (**1**) (diluted 10-fold in 20 mM HEPES buffer pH 7.5 with 150 mM KCl and 10% glycerol) to Cas9 at 4 °C over-night. The bioconjugate was further purified by size-exclusion chromatography in 20 mM HEPES buffer pH 7.5 with 150 mM KCl and 10% glycerol to remove unreacted (**1**).

Conjugation of construct (A) or (C) with Alexa Fluor® 532 (AFg)

General procedure: The starting protein was buffer exchanged using a Zeba spin filter (40 K MWCO) equilibrating with NaCl GF buffer, pH = 8.3) using the manufacturer's instructions. After buffer exchange, a 3.0 mM solution of Alexa Fluor® 532 NHS Ester (Succinimidyl Ester, ThermoFisher, CAS# = 271795-14-3, 1.5 equiv.) in DMSO was added with gentle mixing. The resulting reaction mixture was allowed to incubate for an appropriate time at room temperature, then desalted using Zeba spin filters (40 K MWCO) and equilibrating with NaCl GF buffer (pH = 8.3) using the manufacturer's instructions to remove excess ligand. The reaction was submitted to purification without further manipulations. Isolated samples after CEX purification were

added to a pre-washed EMD Millipore Amicon™ Ultra Centrifugal Filter Units (50 K MWCO) and concentrated (1500 × g for 5 min intervals, pipetting gently to mix after each interval) to an appropriate concentration then diluted with NaCl GF buffer (10×) and concentrated (1500 × g for 5 min intervals, pipetting gently to mix after each interval) to an appropriate concentration.

Conjugate E: AFg-Cas9-2lig-1NLS

The starting protein (A; 1.1 mg) was treated with a 3.0 mM DMSO solution of Alexa Fluor® 532 NHS Ester (Succinimidyl Ester) (1.5 equiv.) and incubated for 2 h at room temperature. The crude material was purified using preparative CEX using a Sepax SCN-NP5 7.8 × 250 mm 5 µm column on a Waters H-Class Acquity UPLC. Mobile Phase A: 20 mM HEPES, 150 mM NaCl, 10% glycerol, pH = 7.4; Mobile Phase B: 20 mM HEPES, 1.0 M NaCl, 10% glycerol, pH =7.4. Injection volumes = 183 µL; purification gradient: 28% A hold for 14 min, linear to 38% A from 14–22 min, linear to 45% from 22–28 min. Flow rate 0.7 ml/min. Fractions were collected between 12.3–22.3 min. The pooled fractions containing the desired product were concentrated yielding 103 µL. The concentration was determined via UV absorbance (11 µM, 1.8 mg/mL, DOL by UV absorbance = 2.7) yielding 0.19 mg (17%) of the desired conjugate.

Conjugate F: AFg-Cas9-2lig-AFr-1NLS

The starting protein (C; 1.25 mg) was treated with a 3.0 mM DMSO solution of Alexa Fluor® 532 NHS Ester (Succinimidyl Ester) (1.5 equiv.) and incubated for 2 h at room temperature. The crude material was purified using preparative CEX using a Sepax SCN-NP5 7.8 × 250 mm 5 µm column on a Waters H-Class Acquity UPLC. Mobile Phase A: 20 mM HEPES, 150 mM NaCl, 10% glycerol, pH = 7.4; Mobile Phase B: 20 mM HEPES, 1.0 M NaCl, 10% glycerol, pH =7.4. Injection volume = 366 µL; purification gradient: 29% A hold for 14 min, linear to 38% A from 14–22 min, linear to 45% from 22–28 min. Flow rate 1 ml/min. Fractions were collected between 13.1–19.1 min. The pooled fractions containing the desired product were concentrated yielding 106 µL. The concentration was determined via UV absorbance (7.7 µM, 1.3 mg/mL, DOL by UV absorbance = 2.2) yielding 0.14 mg (11%) of the desired conjugate.

Conjugate H: Cas9-2AFr (thioether)-1NLS

The starting protein Cas9 (M1C/C80S) 1 NLS was passed through a 0.22µm cellulose acetate spin-X centrifuge tube filter and centrifuged at 15000×g for 2 minutes. The Cas9 (M1C/C80S) 1 NLS endotoxin free (9.28 mg) was buffer exchanged using a Zeba 5 mL spin desalting column (Cat# = 87770, 40K MWCO) equilibrating with NaCl GF buffer using the manufacturer's instructions. After buffer exchange, 10 equivalents of Alexa Fluor® 647 C_2 Maleimide (catalog number =A20347,ThermoFisher, 3.0 mM, 10% DMSO/GF buffer solution (576 nmol, 192 µL)) was added with gentle mixing. The resulting reaction mixture was allowed to incubate for 1.5 h at room temperature. After incubation, the reaction was desalted using a Zeba 5 mL spin desalting column (Cat# = 87770, 40K MWCO) equilibrating with NaCl GF buffer using the manufacturer's instructions to remove excess ligand yielding 1600 µL of solution. The crude material was purified using preparative CEX using a Sepax SCN-NP5 7.8 × 250 mm 5 µm column on a Waters H-Class Acquity UPLC. Mobile Phase A: 20mM HEPES, 150mM NaCl, 10% glycerol, pH = 7.4; Mobile Phase B: 20mM HEPES, 1.0M NaCl, 10% glycerol, pH =7.4. Injection volumes = 600 µL ×2, 400 µL; purification gradient: 22% A hold for 14 minutes, linear to 24% A from 14–22 minutes, linear to 45% from 22-28 minutes. Fractions were collected between 18.9–23.2 minutes. The pooled fractions containing the desired product (10.5 mL) were diluted with NaCl GF buffer (10.5 mL) and added to a pre-washed EMD Millipore Amicon™ Ultra-15 Centrifugal Filter Units (UFC905024, 50K MWCO) and concentrated (15000×g for 5 minute intervals, pipetting gently to mix after each interval) to a final volume of 335 µL (buffer concentration 0.02 M HEPES, 0.22 M NaCl, 10% glycerol, pH = 7.4). The concentration was determined via UV absorbance (15.5 µM, 2.53 mg/mL, DOL = 1.98) yielding 0.85 mg (9%) of the desired conjugate.

Intact Protein Mass Spectrometry

Cas-1NLS constructs

The sample analysis was carried out on an Agilent 6530 QTof mass spectrometer equipped with a Dual AJS electrospray source operated in positive ion mode. The mass spectrometer was interfaced with an Agilent 1290 UPLC system. The Agilent 1290 autosampler injected 10 µL aliquots of sample which was diluted to 0.1 mg/mL in MilliQ water just prior to analysis. The material was separated using a Agilent PLRP-S 1000Å 50 x 2.1 mm with 5.0 µm particles column (part no. PL1912-1502). The mobile phases were: A) 0.1% formic acid in water and B) 0.1% formic acid in acetonitrile. Raw mass spectra were viewed using MassHunter (version B.07.00 Service Pack 2, Agilent) and mass spectral deconvolution was performed using BioConfirm (B.07.00, Agilent).

A = Cas9-2lig-1NLS

B = Cas9-AFr-1NLS

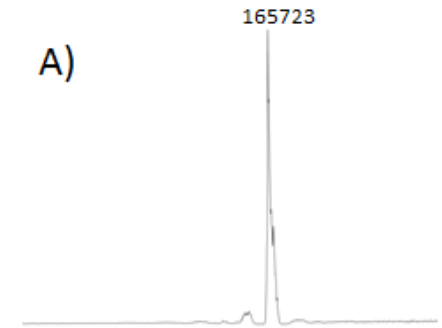
C = Cas9-2lig-AFr-1NLS

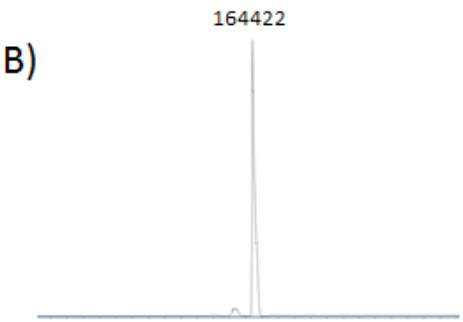
D = Cas9-1NLS

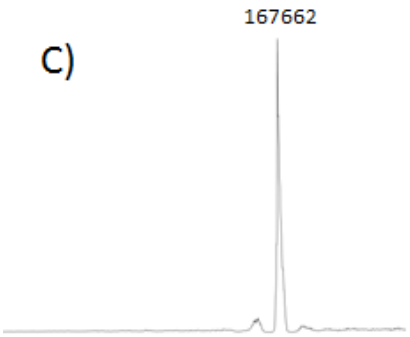
E = AFg-Cas9-2lig-1NLS

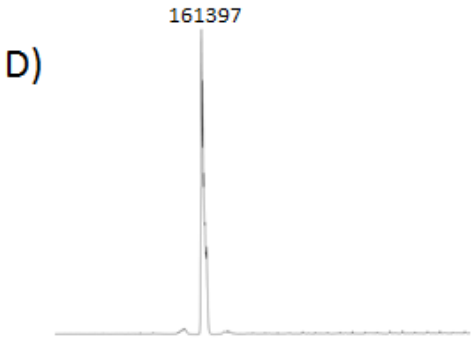
F = AFg-Cas9-2lig-AFr-1NLS

H = Cas9-2AFr (thioether)-1NLS (analytical data are combined in a distinct section, page S23)



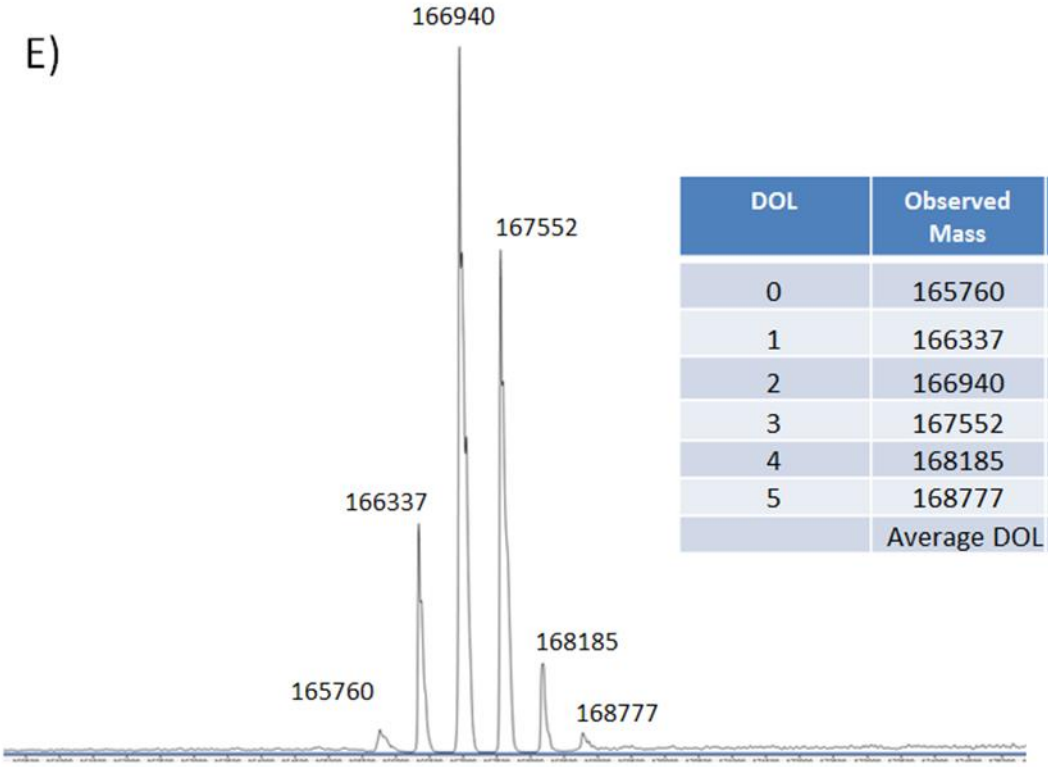






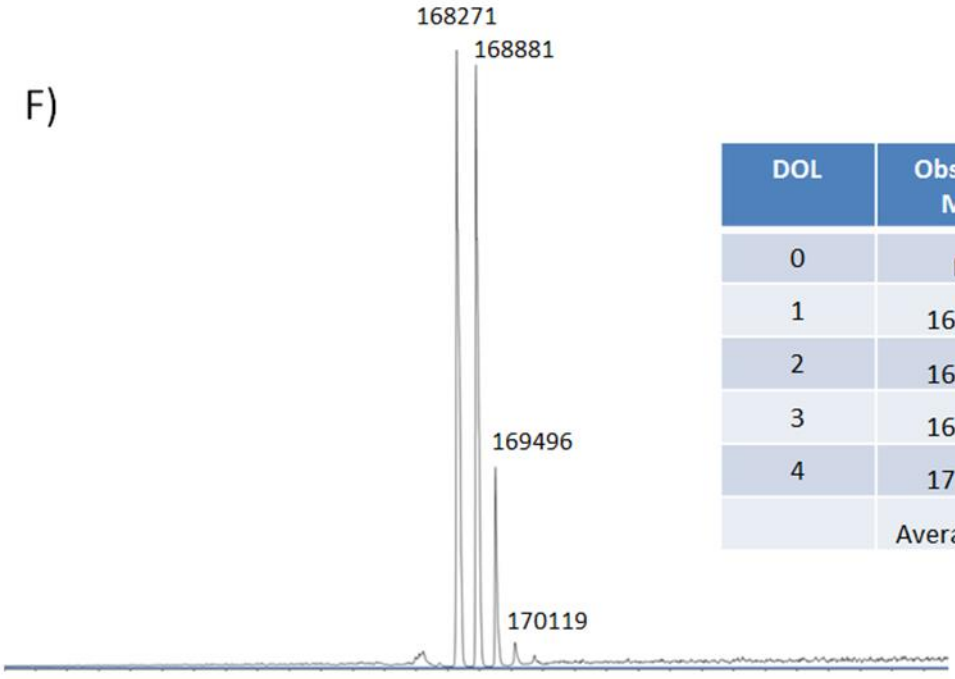
Sample	Observed Mass	Expected Mass
A	165723	165719
B	164422	164415
C	167662	167655
D	161397	161395

E)



DOL	Observed Mass	Expected Mass
0	165760	165719
1	166337	166328
2	166940	166937
3	167552	167546
4	168185	168156
5	168777	168765
	Average DOL	2.3

F)

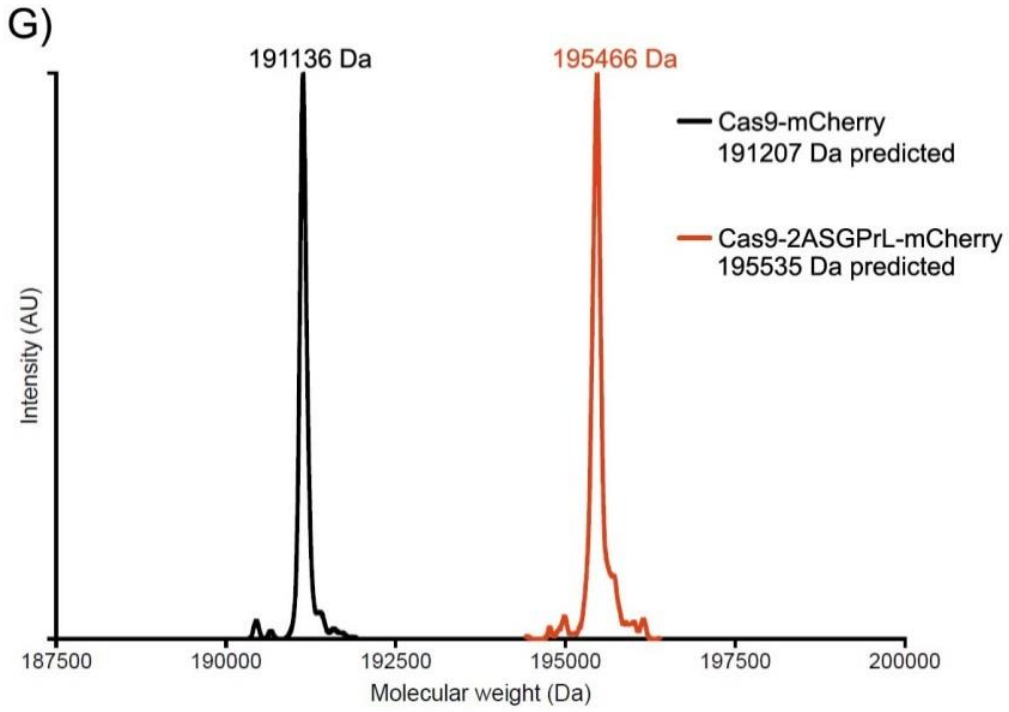


DOL	Observed Mass	Expected Mass
0	ND	167655
1	168271	168264
2	168881	168873
3	169496	169482
4	170119	170092
	Average DOL	1.7

The deconvoluted spectra showing A) Cas9-1NLS bisligated to **1** (-thioPyridyl); B) Cas9-1NLS bisligated to **2** (-thioPyridyl); C) Cas9-1NLS bisligated to **3** (-thioPyridyl); D) Cas9-1NLS starting material, E) Cas9-2lig-1NLS ligated to AF532, F) Cas9-2lig-AFr-1NLS ligated to AF532 (AFg).

Cas9-mCh constructs

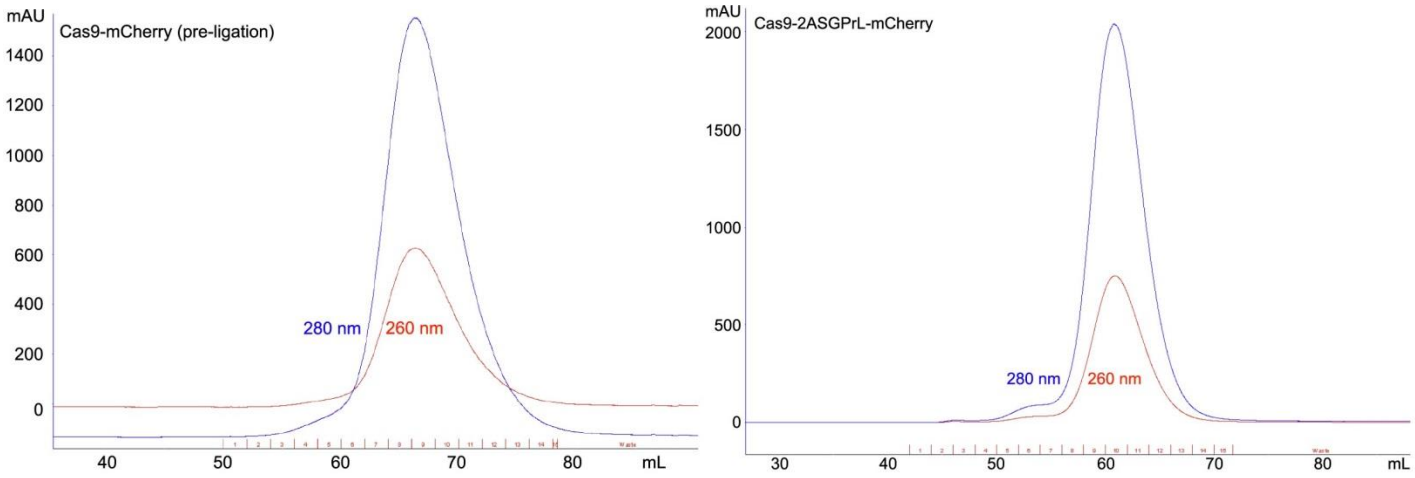
500 pmol of intact protein was buffer exchange to 1 M ammonium acetate pH 7.5 and directly injected in a Thermo LTQ-Orbitrap-XL mass spectrometer equipped with an electrospray ionization (ESI) source. Raw mass spectra were viewed using Xcalibur software (version 2.0.7, Thermo) and mass spectral deconvolution was performed using ProMass software (version 2.5 SR-1, Novatia). Data also shown in Figure S1A.



Size exclusion chromatography (SEC)

Cas9-mCh constructs

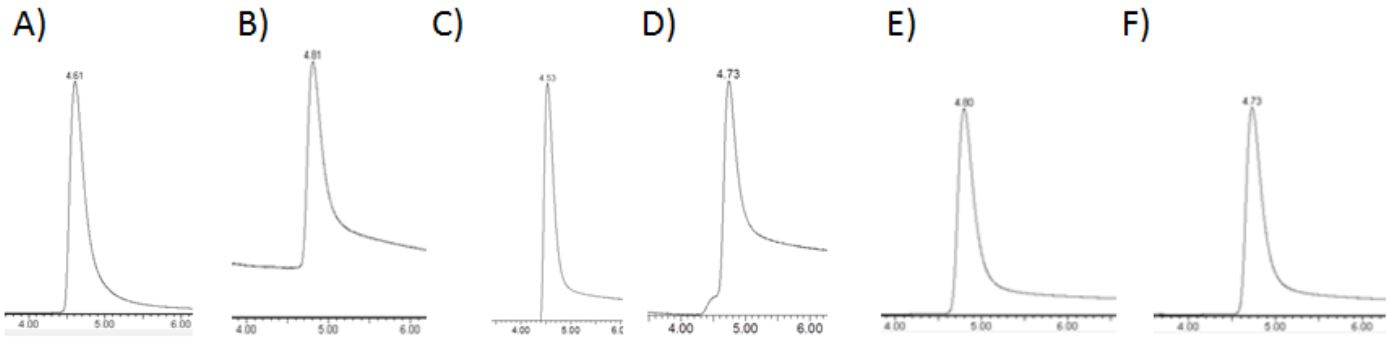
SEC was performed on a Akta Purifier using a HiLoad 16/60 S200 superdex column with gel filtration buffer (20 mM HEPES pH 7.5, 150 mM KCl, 10% (v/v) glycerol) with a flow rate of 1 mL/min. Protein was loaded in volumes no greater than 1 mL.



Note: the ligated mCherry construct used for experiments in Figure S2 (Cas9-2lig-mCh) were conjugated according to the protocol above but were subjected to a simplified clean-up. Instead of the SEC step, the conjugation reaction was passed through a 0.5 mL capacity, 40 kDa MWCO Zeba desalting column (Life Technologies) to remove unreacted ligand. This would not be anticipated to ensure complete removal of the unreacted ligand, but we do not anticipate low levels of contaminating ligand to interfere with the qualitative result (see ligand competition data in Figure S11).

Cas9-1NLS constructs

Analytical SEC was performed on a Waters H-Class Acquity UPLC using a Waters BEH200 SEC 1.8µm 4.6 × 150 mm ID column. An isocratic gradient was used wherein the mobile phase was sodium chloride GF buffer (20 mM HEPES, 150 mM NaCl, 10% glycerol at pH = 7.4). The flow rate was 0.25 mL/min at room temperature over 12 min. Typical injection size is 10–30 µg of protein.



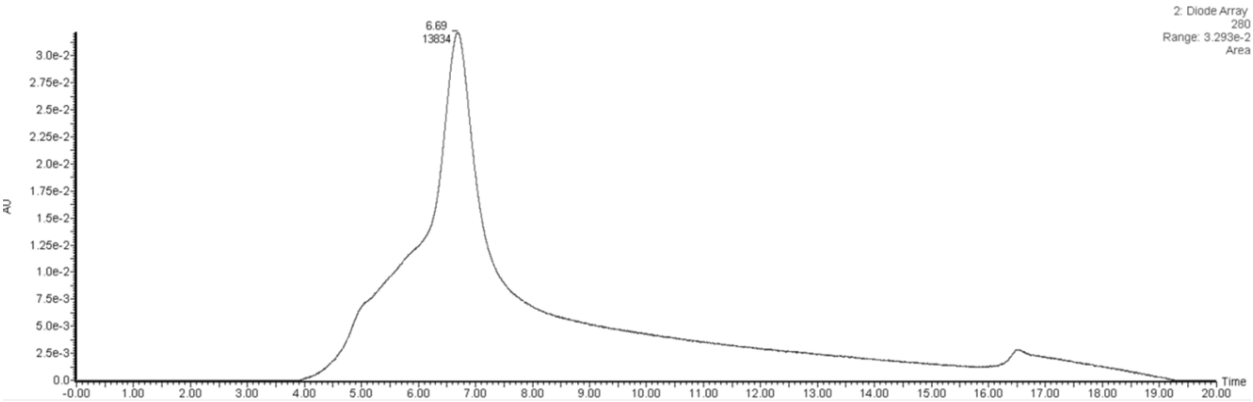
Retention times: Cas9-2lig-1NLS (**A**) = 4.61 min.; Cas9-AFr-1NLS (**B**) = 4.81 min.; Cas9-2lig-AFr-1NLS (**C**) = 4.53 min.; Cas9 1NLS starting material (**D**) = 4.73 min.; AFg-Cas9-2lig-1NLS (**E**) = 4.80 min.; AFg-Cas9-2lig-AFr-1NLS (**F**) = 4.73 min.

Cationic exchange chromatography (CEX)

Analytical CEX was performed on a Waters H-Class Acquity UPLC using a Sepax SCN-NP5 4.6 × 250 mm 5 µm. Mobile Phase A: 20 mM Hepes,150 mM NaCl, 10% glycerol (pH = 7.4) and Mobile Phase B: 20 mM HEPES, 1.0 M NaCl, 10% glycerol (pH = 7.4).
The flow rate was 0.6 mL/min at room temperature over 20 min. Typical injection size is 10–30 µg of protein. Broadened peak shape due to initial conditions being isocratic gradient necessary to observe DOL differences.

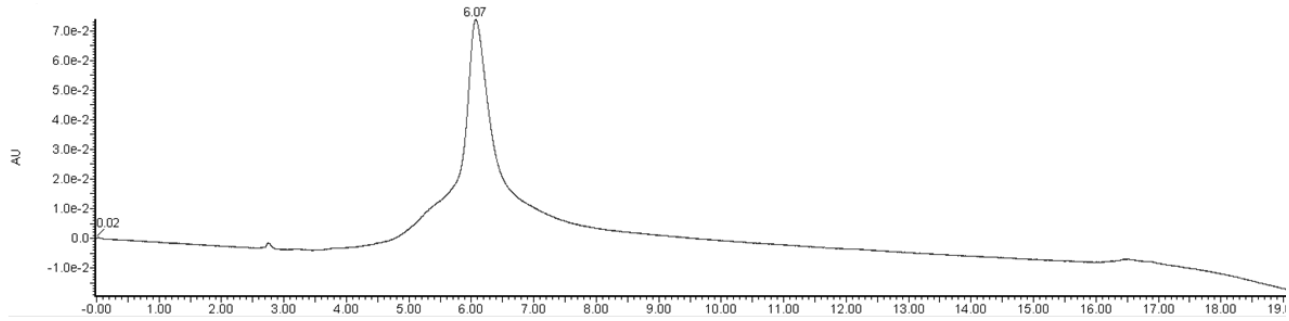
A = Cas9-2lig-1NLS

Column: Sepax SCN-NP5 4.6 × 250mm 5 µm
Mobile Phase A: 20 mM Hepes + 150 mM KCl, Mobile Phase B: 20 mM Hepes + 1.0M KCl
10% A Hold for 6 min, linear to 13% A from 6–11 min, linear to 45% from 11–14 min, column wash at 100% A and re-equilibration to 20 min. Retention time = 6.69 min



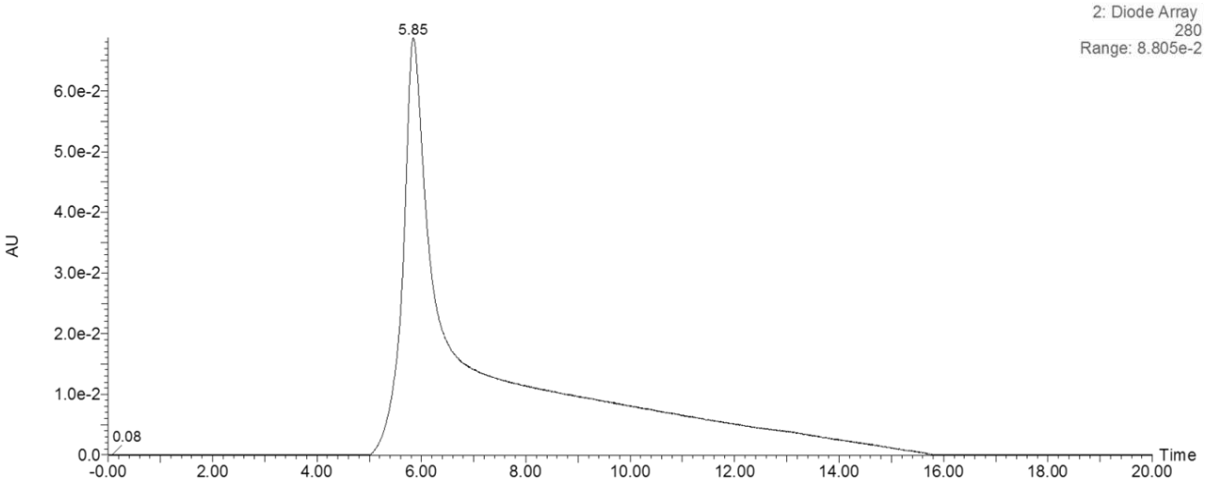
B = Cas9-AFr-1NLS

Column: Sepax SCN-NP5 4.6 × 250 mm 5 µm
Mobile Phase A: 20 mM Hepes + 150 mM NaCl, Mobile Phase B: 20 mM Hepes + 1.0 M NaCl
21% A Hold for 7 min, then 26% A Linear to 27% A over 11 min, linear to 45% from 11–14 min. column wash at 100% A and re-equilibration to 20 min. Retention time = 6.07 min



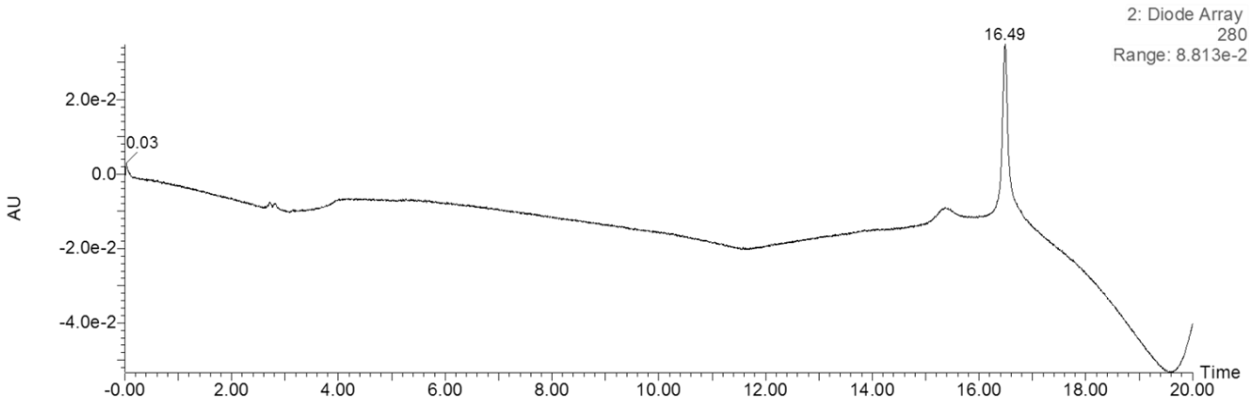
C = Cas9-2lig-AFr-1NLS

Column: Sepax SCN-NP5 4.6 × 250 mm 5 µm
Mobile Phase A: 20 mM Hepes + 150 mM KCl, Mobile Phase B: 20 mM Hepes + 1.0M KCl
15% A Hold for 6 min, linear to 17% A from 6–11 min, linear to 45% from 11–14 min, column wash at 100% A and re-equilibration to 20 min. Retention time = 5.85 min



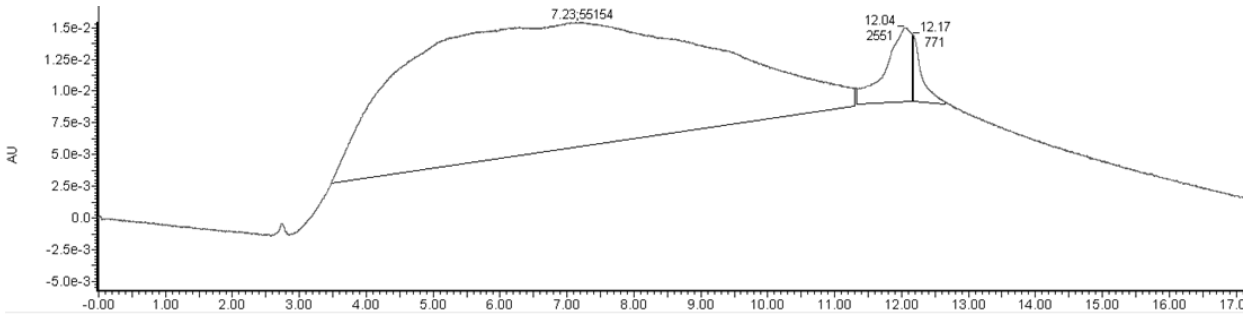
D = Cas9-1NLS

Column: Sepax SCN-NP5 4.6 × 250 mm 5 µm
Mobile Phase A: 20 mM Hepes + 150 mM KCl, Mobile Phase B: 20 mM Hepes + 1.0 M KCl
16% A Hold for 6 min, linear to 19% A from 6–11 min, linear to 45% from 11–14 min, column wash at 100% A and re-equilibration to 20 min. Retention time = 16.49 min



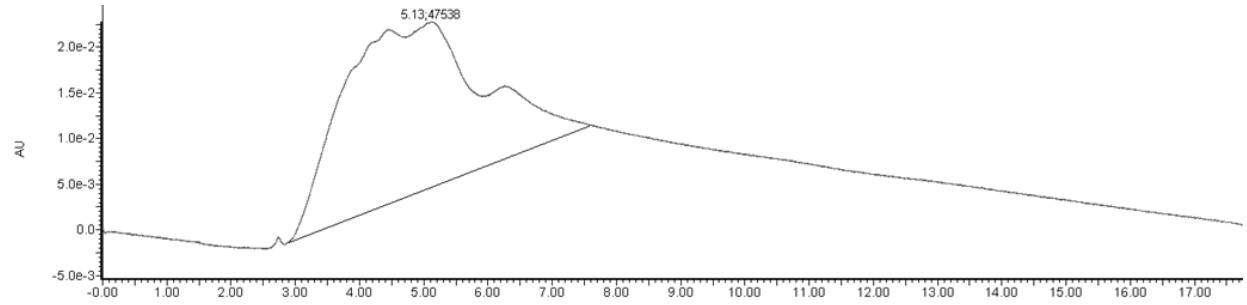
E = AFg-Cas9-2lig-1NLS

Column: Sepax SCN-NP5 4.6 × 250 mm 5 μm
Mobile Phase A: 20 mM Hepes + 150 mM NaCl, Mobile Phase B: 20 mM Hepes + 1.0 M NaCl
21% A Hold for 7 min, then Linear to 26% from 7–11 min, linear to 45% from 11–14 min. Range = 3.5–12.8 min



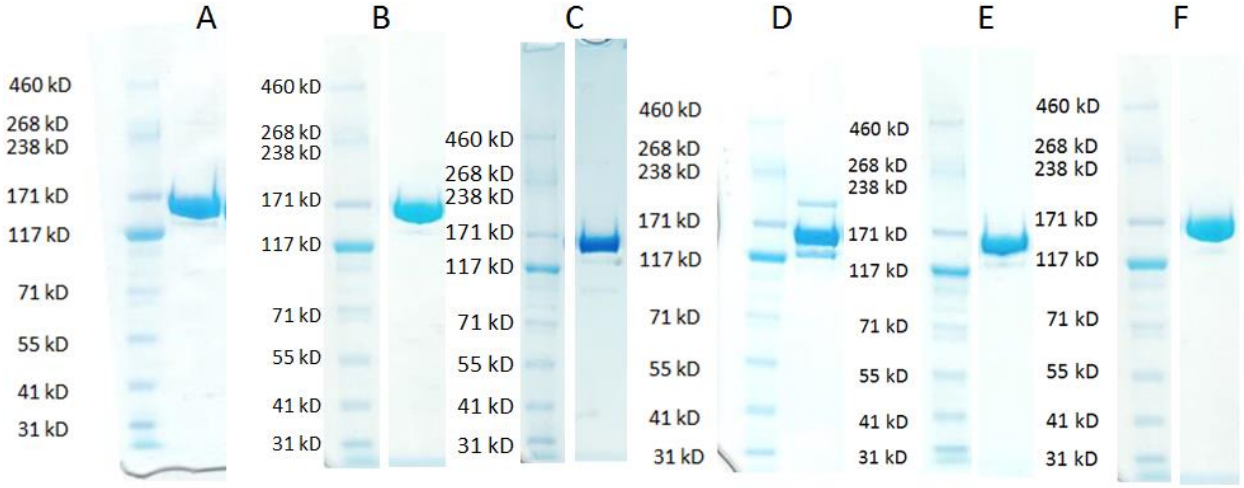
F = AFg-Cas9-2lig-AFr-1NLS

Column: Sepax SCN-NP5 4.6 × 250 mm 5 μm
Mobile Phase A: 20 mM Hepes + 150 mM NaCl, Mobile Phase B: 20 mM Hepes + 1.0 M NaCl
21% A Hold for 7 min, then Linear to 26% from 7–11 min, linear to 45% from 11–14 min. Range = 3.0–7.0 min



SDS-PAGE gels

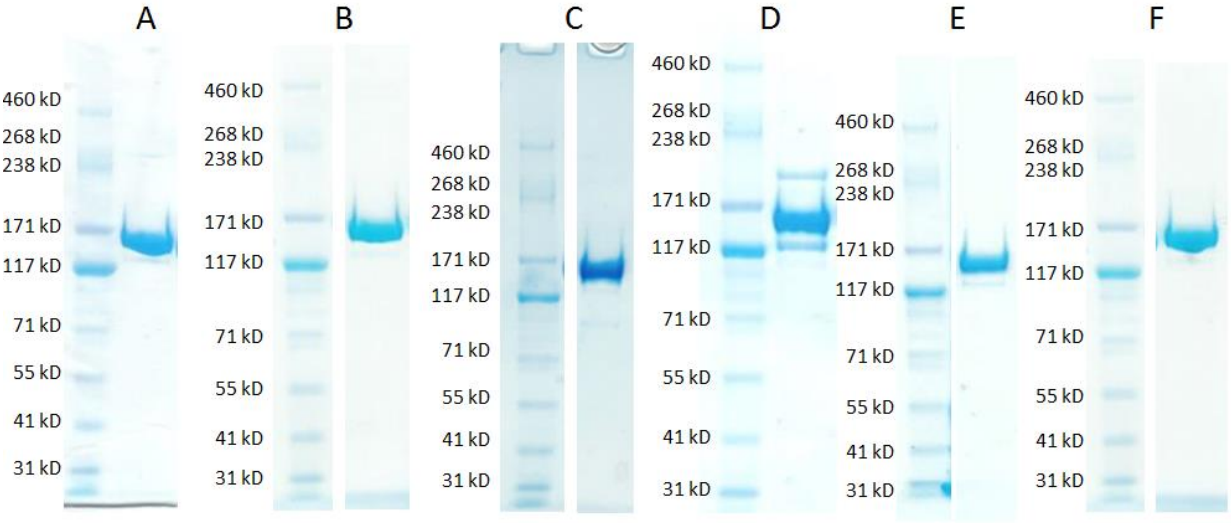
Reduced conditions: Ten microliter samples (3 µM) were added to 5 µL 4× LDS sample buffer (Life Technologies). Reduced samples were prepared by the addition of DTT (final concentration = 50 mM) included in the sample buffer. Ten microliter samples were loaded onto a 3–8% Tris-Acetate gel (Life Technologies) in addition to 10 µL pre-stained HiMark standard (Life Technologies) included on the gels as a molecular weight reference. The gel was run in 1× Tris-Acetate running buffer for 1h at room temperature at 150V, constant voltage. Gel was stained with biological stain (Boston Biological) to visualize protein bands.



A = Cas9-2lig-1NLS B = Cas9-AFr-1NLS C= Cas9-2lig-AFr-1NLS

D = Cas9-1NLS E = AFg-Cas9-2lig-1NLS F = AFg-Cas9-2lig-AFr-1NLS

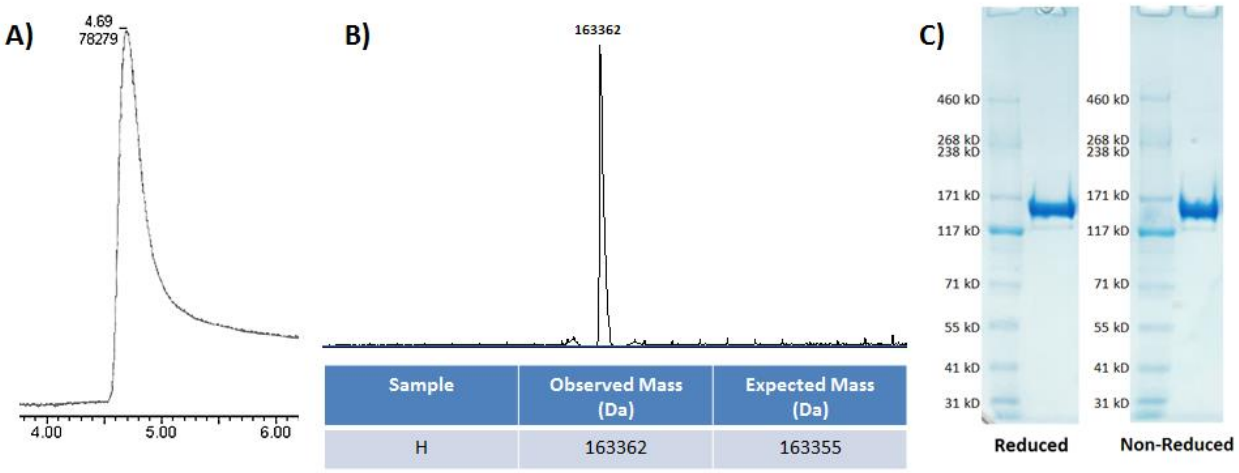
Non-reduced conditions: ThermoFisher Novex Gels; Pre-stained HiMark Standards used (10 µL/well)



A = Cas9-2lig-1NLS B = Cas9-AFr-1NLS C= Cas9-2lig-AFr-1NLS

D = Cas9-1NLS E = AFg-Cas9-2lig-1NLS F = AFg-Cas9-2lig-AFr-1NLS

Combined Analytical Data for Conjugate H



Sample	Observed Mass (Da)	Expected Mass (Da)
H	163362	163355

(A) Analytical SEC was performed on a Waters H-Class Acquity UPLC using a Waters BEH200 SEC 1.8 µm 4.6 × 150 mm ID column. An isocratic gradient was used wherein the mobile phase was sodium chloride GF buffer (20 mM HEPES, 150 mM NaCl, 10% glycerol at pH = 7.4). The flow rate was 0.25 mL/minute at room temperature over 12 minutes. Typical injection size is 10 – 30 µg of protein. Retention time for conjugate H = 4.69 minutes (B) Deconvoluted QTOF Mass spectra (C) SDS-PAGE.

CEX Analytical: Column: Sepax SCN-NP5 4.6×250 mm 5 µm
Mobile Phase A: 20 mM Hepes + 150 mM NaCl, Mobile Phase B: 20mM Hepes + 1.0 M NaCl
20% A Hold for 6 minutes, linear to 25% A from 6–11 minutes, linear to 45% from 11–14 minutes, column wash at 100% A and re-equilibration to 20 minutes. Retention time for conjugate H = 15.20 min.

Trypsin Digestion

Digest Protocol

32 µL of a 0.5 mg/mL solution of protein was diluted to 100 µL. The protein was then precipitated with the addition of 400 µL of ice cold acetone and incubated over night at –20 °C. The sample was centrifuged and the supernatant discarded. The pellet was dried using an Eppendorf SpeedVac for 5 min. The protein pellet was dissolved using 20 µL of 8M urea, 20 mM methylamine, sonicated for 5 min, and 140 µL of 50 mM Tris, 10 mM $CaCl_2$ was added. Trypsin was added to a final concentration of 0.01 mg/mL and incubated overnight at 37 °C.

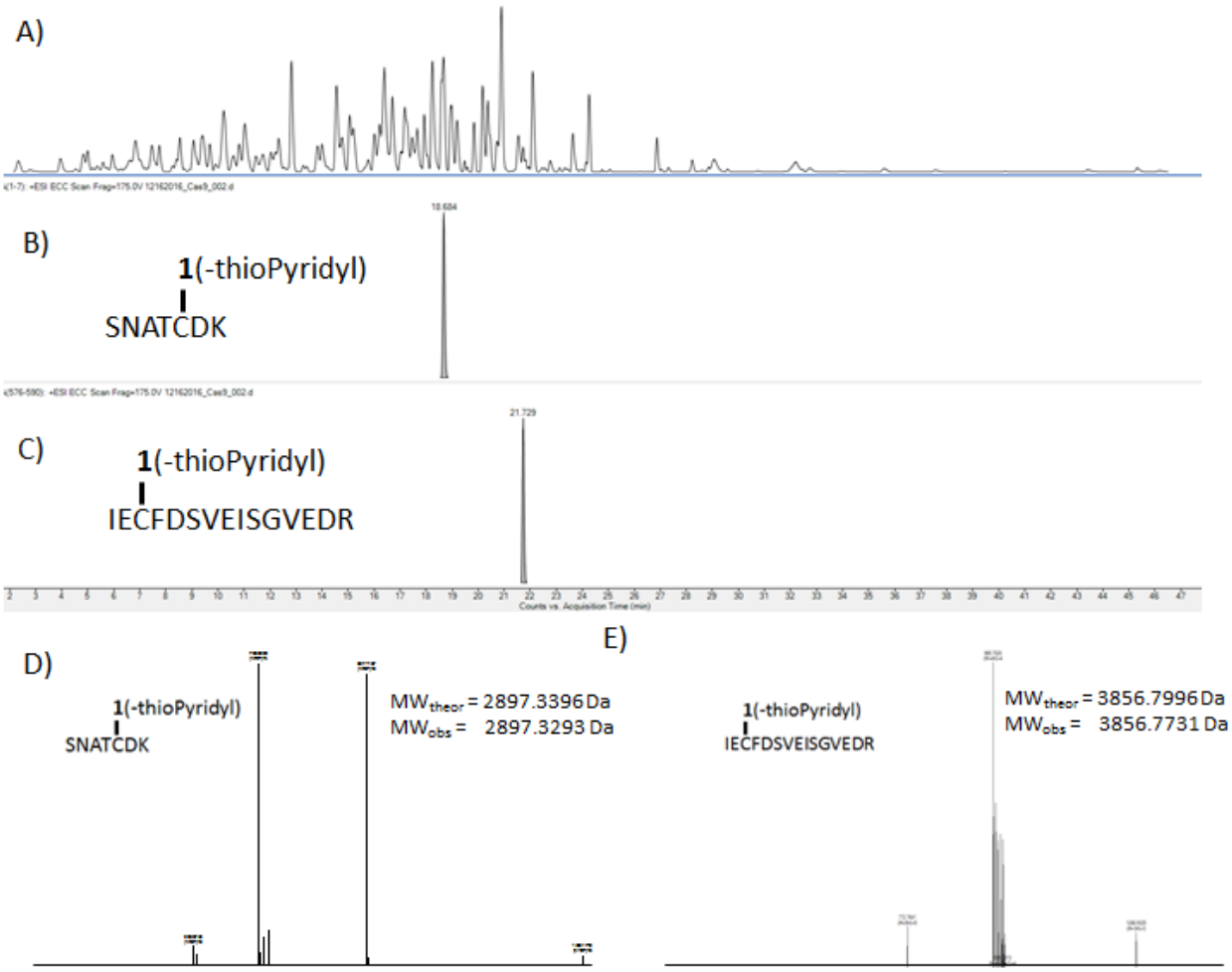
Positive Ion Digest Mass Spectrometry

The sample analysis was carried out on an Agilent 6530 QTof mass spectrometer equipped with a Dual AJS electrospray source operated in positive ion mode using an Auto MS/MS Acquisition method. The mass spectrometer was interfaced with an Agilent 1290 UPLC system complete with a MWD. The MWD was set to collect absorbances of 650 nm, 254 nm, and 210 nm. The Agilent 1290 autosampler injected 10 µL aliquots of the digested samples. The material was separated using a Agilent PLRP-S 100Å 50 × 2.1 mm with 3.0 µm particles column (part no. PL1912-1300). The mobile phases were: A) 0.1% formic acid in water and B) 0.1% formic acid in acetonitrile. Raw mass spectra were viewed using MassHunter (version B.07.00 Service Pack 2, Agilent) and peptide identification was performed using BioConfirm (B.07.00, Agilent).

For the Alexafluor containing constructs the 650 nm trace was used to determine the location of the formed peptide and the mass was used to confirm identification.

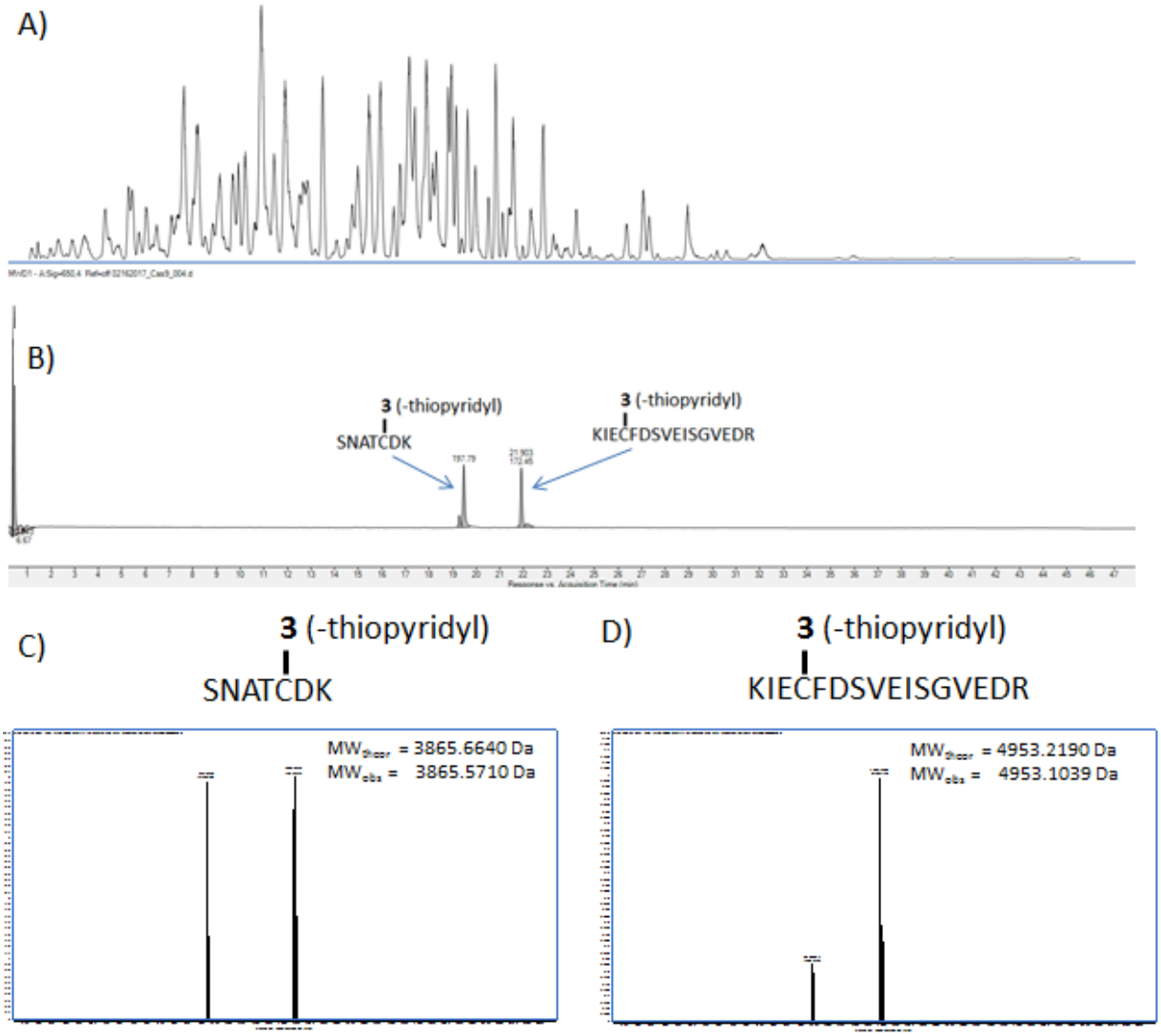
NOTE: all samples used the same analytical gradient for the peptide mapping experiment.

A: Cas9-2lig-1NLS



Trypsin digest data showing evidence for the conjugated peptides derived from Cas9-2lig-1NLS. A) is the Total Compound Chromatogram generated from the BioConfirm software. B) shows the Extracted Compound Chromatogram for the conjugated peptide SNATCDK (sequence 1–7) to compound **1**(- thioPyridyl). C) shows the Extracted Compound Chromatogram for the conjugated peptide IECFDSVEISGVEDR (sequence 576–590) to compound **1** (- thioPyridyl). D) is the Molecular Feature spectra for the conjugated peptide SNATCDK (sequence 1–7) to compound **1** (- thioPyridyl). E) is the Molecular Feature spectra for the conjugated peptide IECFDSVEISGVEDR (sequence 576–590) to compound **1** (- thioPyridyl).

C = Cas9-2lig-AFr-1NLS



Trypsin digest data showing evidence for the conjugated peptides derived from Cas9-2lig-AFr-1NLS. A) is the Total Compound Chromatogram generated from the BioConfirm software. B) shows the UV trace for wavelength= 650 nm. C) is the Molecular Feature spectra for the conjugated peptide SNATCDK (sequence 1–7) to compound **3** (- thioPyridyl). D) is the Molecular Feature spectra for the conjugated peptide KIECFDSVEISGVEDR (sequence 575–590) to compound **3** (- thioPyridyl).

Stability study under T7E1 co-incubation assay conditions

T7E1 assay buffer

700 µL of DMEM 10%FBS (Gibco DMEM low glucose 11885-084) + Gibco (Life Technologies) HI FBS Ref. 16140-063)
500 µL of OptiMEM (Gibco OptiMEM 31985-070)
20 µL RNP buffer
75 µL water
75 µL (74.75 µL water + 0.25 µL DMSO)
Buffer was mixed gently and used in stability study

Stability experiments

Cas9-2lig-1NLS/sgRNA EMX1 RNP was prepared with a final concentration of 12.5 μM, transferred into 2 μL aliquots, flash frozen, and stored at –80 °C until incubation. The Cas9-2lig-1NLS/sgRNA EMX1 RNP was diluted into T7E1 assay buffer to a final concentration of 0.18 μM and incubated a 37 °C. The timing of the experiment was such that all the samples would be finished near the same time to allow for sequential analysis. A 50 μL aliquot for each time point was transferred to a vial and 200 μL of ice cold acetone was added. The samples were vortex and incubated at –20 °C overnight. The samples were centrifuged at 14000 RPM for 30 min. The supernatant was transferred to a new vial and lyophilized to dryness. The samples were reconstituted in 50 μL of 1mM TCEP in water just prior to analysis on the Agilent 6530 QTOF using the same method used for the intact protein analysis. An extracted ion chromatogram was generated using the $(M+3H)^{3+}$ charge state using a +/– 50 ppm symmetrical window. The area of the resulting chromatographic peak was used to determine the concentration of reduced Compound **1** present.

*Standard curve preparation for reduced Compound **1***

An initial 30 μM stock solution of reduced Compound **1** was prepared in 1 mM TCEP. A sequential dilution of the sample was made into T7E1 assay buffer with 10 mM TCEP. A 50 μL aliquot of calibration curve samples were transferred to a vial and 200 μL of ice cold acetone was added. The samples were vortex and incubated at –20 °C overnight. The samples were centrifuged at 14000 RPM for 30 min. The supernatant was transferred to a new vial and lyophilized to dryness. The calibration curve samples were reconstituted in 50 μL of 1mM TCEP in water just prior to analysis on the Agilent 6530 QTof using the same method used for the intact protein analysis. An extracted ion chromatogram was generated using the $(M+3H)^{3+}$ charge state using a +/– 50 ppm symmetrical window. The area of the resulting chromatographic peak was used to determine the concentration of reduced Compound **1** present.

Exp. Table 1:

Time Point	% Reduced **1** Normalized for T=0*
0 h	0.02
1 h	0.60
2 h	2.81
4 h	6.84
8 h	8.16
24 h	36.83

*residual ligand from purification and sample preparation responsible for reduced compound **1** present at T_0. The % reduced compound 1 was then normalized to reflect this starting amount.

QTOF MS Results for 5% FBS/DMEM incubation

RNPs were prepared with a final concentration of 6 μM and diluted with 10% FBS/DMEM solution to simulate T7E1 assay conditions. The samples were incubated at 37 °C and 4 μL was removed for each time point. The subsequent aliquot was diluted (1:4) in water and analyzed using Intact Protein Mass Spectrometry protocol.

Exp. Table 2:

Sample description	1 h %(bis/mono/ unligated)	2 h %(bis/mono/ unligated)	4 h %(bis/mono/ unligated)	8 h %(bis/mono/ unligated)	24 h %(bis/mono/ unligated)
Cas9-2lig-1NLS + sgRNA(PCSK9) RNP	100/0/0	100/0/0	100/0/0	97.4/2.6/0	34.7/65.3/0
Cas9-2lig-AFr-1NLS + sgRNA (PCSK9) RNP	100/0/0	100/0/0	100/0/0	100/0/0	61.2/38.8/0

sgRNA (PCSK9):
GGGCUGAUGAGGCCGCACAUGGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU (4).

Cell-free *in vitro* DNA cleavage to test pH stability, using Cas9 mCherry constructs

Buffers containing 150 mM KCl and 140–185 mM sodium phosphate/citrate (ratio of buffering agent slightly adjusts the concentration) were prepared (according to http://microscopy.berkeley.edu/Resources/instruction/buffers.html) in parallel at the following pH values: 4.0, 4.5, 5.0, 5.5, 6.0, and 6.5. To verify that these buffers could be used to adjust the pH of KCl GF buffer (20 mM HEPES buffer pH 7.5, 150 mM KCl, 10% (v/v) glycerol), 1 μL of each buffer was mixed with 3 μL of KCl GF buffer and the resulting pH was confirmed using pH strips. To prepare RNP, Cas9-mCh at 13.5 μM final concentration (stock at 98 μM) was mixed into EMX1 sgRNA at 16 μM (stock at 16μM), giving a 1:1.2 molar ratio of protein to sgRNA. This mixture was incubated for 10 min at 37 °C. In a strip of PCR tubes, 1μl of phosphate/citrate buffer corresponding to the appropriate pH was added to 3 μL of RNP (in that order) and this was incubated at 37 °C for 1h.

To test cleavage activity following incubation at varying pH conditions, a dsDNA substrate was first generated by amplifying the targeted human EMX1 locus from genomic DNA using PCR (Kapa Biosystems) and quantified on 2% agarose gel stained with SYBR Gold (Invitrogen) by comparison to a standard. 200 ng (~0.5 pmol) of the resulting dsDNA substrate (PCR product of EMX1 locus; primers below) was incubated with 4 pmol of RNP in 20 mM HEPES buffer pH 7.5 with 150 mM KCl and 10% glycerol and incubated at 37 °C for 60 min followed by 95 °C for 5 min, followed by agarose gel run & analysis.

EMX1 Fwd primer: 5′-GCCATCCCCTTCTGTGAATGTTAGAC-3′
EMX1 Rev primer: 5′-GGAGATTGGAGACACGGAGAGCAG-3′

5. EMX1 sgRNA synthesis, purification and analysis

IVT-generated EMX1 sgRNA

Single-guide RNA were synthesized by T7 *in vitro* transcription and purified by PAGE as previously described (5). Briefly, for the EMX1 sgRNA template, the PCR reaction contains 20 nM premix of (5′- TAA TAC GAC TCA CTA TAG GTC ACC TCC AAT GAC TAG GGG TTT AAG AGC TAT GCT GGA AAC AGC ATA GCA AGT TTA AAT AAG G -3′) and (5′- AAA AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT AAA CTT GCT ATG CTG TTT CCA GC -3′) as well as a 1 µM premix of (5′- TAA TAC GAC TCA CTA TAG -3′) and (5′- AAA AAA AGC ACC GAC TCG GTG C -3′), 200 µM dNTP and Phusion Polymerase (NEB, Ipswich, MA) according to manufacturer's protocol. The thermocycler setting consisted of 30 cycles of 95 °C for 10 s, 57 °C for 10 s and 72 °C for 10 s. The PCR product was extracted once with phenol:chloroform:isoamylalcohol and then once with chloroform, before isopropanol precipitation overnight at −20 °C. The DNA pellet was washed three times with 70% ethanol, dried by vacuum and dissolved in DEPC-treated water. An 100-µL T7 in vitro transcription reaction consisted of 30 mM Tris–HCl (pH 8), 20 mM $MgCl_2$, 0.01% Triton X-100, 2 mM spermidine, 10 mM fresh dithiothreitol, 5 mM of each ribonucleotide triphosphate, 100 µg/mL T7 RNA polymerase expressed and purified according to (6) and 1 µM DNA template. The reaction was incubated at 37 °C for 4 h, and 5 units of RNase-free DNaseI (Promega, Madison, WI) was added to digest the DNA template at 37 °C for 1 h. The reaction was quenched with 2× STOP solution (95% deionized formamide, 0.05% bromophenol blue and 20 mM EDTA) at 60 °C for 5 min. The RNA was purified by electrophoresis in 10% polyacrylamide gel containing 6 M urea. The RNA band was excised from the gel, grinded up in a 15-ml tube, and eluted with 5 vol of 300 mM sodium acetate (pH 5) overnight at 4 °C. One equiv. of isopropanol was added to precipitate the RNA at −20 °C. The RNA pellet was collected by centrifugation, washed three times with 70% ethanol, and dried by vacuum. The sgRNA was refolded by heating to 70 °C for 5 min in 20 mM HEPES buffer pH 7.5 with 1 mM $MgCl_2$, 150 mM KCl and 10% glycerol and cooling to room temperature.

EMX1 sgRNA synthesized with ribozymes

Construction of vector pRZG01 and pRZG02

A general-use vector pRZG01 encoding a T7 polymerase promoter followed by a golden gate cloning cassette (7) and an HCV ribozyme (8) was constructed. A pT7CFE1 derivative was a gift from Jamie H.D. Cate (University of California-Berkeley) and was PCR amplified using primers 5′-AGGAGGTGGAGATGCCATGCCGACCCGAGACCACCGGTGCTAGCGGATCCGGTCTCCTATAGTGAGTCGTATTAATTTCACTGGCCGTCGTTTTACAACG-3′ and 5′-GCATGGCATCTCCACCTCCTCGCGGTCCGACCTGGGCATCCGAGGAAACTCGGATGGCTAAGGGAGAGCCGAATTCGATATCTTAATTAAGCTGCAGG-3′ (Integrated DNA technologies) with herculase II DNA polymerase (Agilent Technologies 600675) in 1× reaction buffer, 0.25 mM each dNTP, 0.25 µM each primer, and 1 µL per 50 µL reaction of the supplied Herculase II stock. DMSO was titrated from 0–6% (v/v) in 1% steps. The PCRs were run in a thermal cycler programmed with the following parameters: 95 °C for two min, 20 cycles of touchdown PCR: (95 °C for 20 sec., annealing; 56 °C for cycle 1, then decreased by 0.3 °C for each subsequent cycle; for 20 sec., 4 min at 68 °C) followed by 25 cycles of standard PCR: (95 °C for 20 sec, 52 °C for 20 sec, then 68 °C for 4 min). The reactions were separated on a 1% Agarose/TAE electrophoresis gel pre-stained with ethidium bromide. The reactions with 0 and 1% had the highest yield and were pooled. The major band, which migrated above the 3000 bp-marker was excised and purified using the Qia-Quick (Qiagen

28704) gel extraction kit according to the manufacturer's instructions and quantified using a Nanodrop 8000 spectrophotometer. 0.1 ng of the resulting product was further amplified using primers 5′-CACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAACGAGACCACCGGTGCTAGCG-3′ and 5′-GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC-TTTTTTTGGGTCGGCATGGCATCTCCAC-3′ (Integrated DNA technologies) with Herculase DNA polymerase (Agilent Technologies 600675) under the same conditions except that the DMSO was only titrated from 0–3%. A faint band was observed above the 3000 bp marker. This band was excised and the PCR product was purified using the Qia-Quick (Qiagen 28704) gel extraction kit. The PCR product was circularized using sequence and ligation-independent cloning. For the exonuclease step, 42 ng purified PCR product in 8 µL total volume was combined with 1 µL NEB buffer 2 and 1 µL 0.25 U/µL T4 DNA polymerase (NEB catalog number M0203S) then incubated at ambient temperature for 30 min. The exonuclease activity was terminated by addition of 1 µL 10 mM dCTP. 9 µL of the resulting mixture was combined with 1 µL T4 DNA ligase buffer (NEB) and incubated for 30 min at 37 °C. 5 µL of the reaction was transformed into chemically competent Mach 1 *E. coli*, which were subsequently plated on LB-Agar with 100 µg/mL ampicillin. Single colonies were selected and used to inoculate 5-mL cultures in LB plus 100 µg/mL ampicillin and grown overnight then miniprepped using the Qiagen miniprep kit (catalog number 27106) according to manufacturer's instructions. Individual clones were screened for the correct insert sequence using Sanger sequencing (Qintara Biosciences) with primer 5′-TGTGCTGCAAGGCGATTAAG-3′. A clone with the expected sequence was then transformed into chemically competent *E coli* K12 ER2925 (*dam-/dcm-).* A single colony was used to inoculate 5 mL LB plus 100 µg/mL ampicillin. This culture was grown overnight and the plasmid was miniprepped using the Qiagen plasmid miniprep kit resulting is unmethylated plasmids for golden gate cloning.

Cloning of Dual ribozyme-sgRNA template

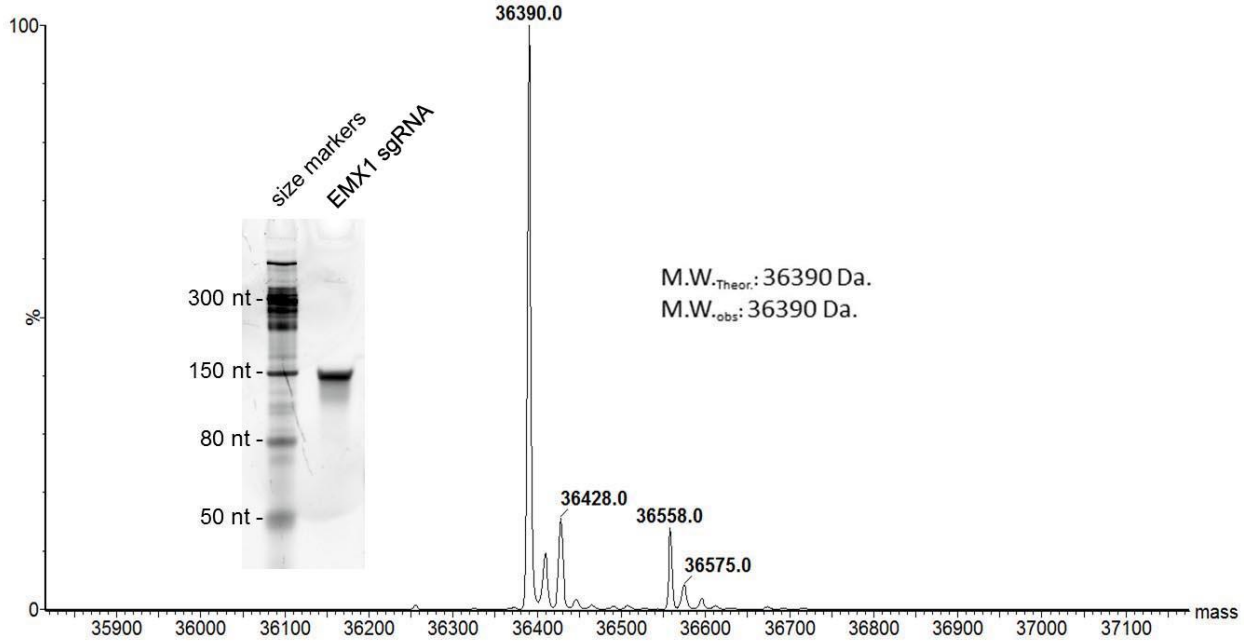
The *in vitro* transcription template pRZG02 was constructed using golden gate assembly (7) of vector pRZG01 and an insert consisting of a 5′-Hammerhead ribozyme followed by the EMX1 spacer. The insert was constructed by overlap extension with synthetic DNA oligonucleotides 5′-AAGCTTGGTCTCCTATAGGGAG-ATTGGAGGTGACCTGATGAGTCCGTGAGGACGAAACGG-3′ and 5′-GAATTCGGTCTCTTAAACCCC-TAGTCATTGGAGGTGACGACGGTACCGGGTACCGTTTCGTCCTCACGGACTCATCAG-3′ (IDT). For the overlap-extension reaction 10 µM of each oligo was combined with 1× Thermopol buffer (NEB), 0.2 mM each dNTP, and 0.1 U/µL Taq DNA polymerase (NEB Catalog M0267S). The resulting mixture was incubated in a thermal cycler at 94 °C for two min then cycled 10 times at 55 °C for 30 sec and 72 °C for 30 sec. For the golden gate assembly, 0.75 µL 10 × T4 Ligase buffer (NEB), 0.375 µL 2 mg/mL BSA (NEB B9000S), 0.5 µL 100 ng/µL unmethylated pRZG01, 0.5 µL BsaI-HF (NEB Catalog number R3535S), and 0.5 µL high concentration (2,000 U/µL) T4 DNA ligase (NEB M0202T) were combined. The resulting mixture was incubated in a thermal cycler for 25 cycles at 3 min 37 °C and 4 min 16 °C. Remaining parent vector was digested by incubation of the mix at 50 °C for 10 min and the enzymes were inactivated by incubation for 5 min at 80 °C. 5 µL of the reaction was transformed into chemically competent Mach 1 *E. coli*, which were subsequently plated on LB-Agar with 100 µg/mL ampicillin. Single colonies were selected and used to inoculate 5 mL cultures in LB plus 100 µg/mL ampicillin and grown overnight then miniprepped using the Qiagen miniprep kit (catalog number 27106) according to manufacturer's instructions. Individual clones were screened for the correct insert sequence using Sanger sequencing (Qintara Biosciences) with primer 5′-TGTGCTGCAAGGCGATTAAG-3′.

In vitro transcription and purification of EMX1 sgRNA

To produce pRGZ02 template at a large scale multiple maxipreps were done according to manufacturer's instructions (Qiagen 12662) and pooled. 1.5 mg of pRGZ02 was linearized in a 30 mL volume by incubation at 37 °C for 2–4 h in the presence of 250 U/mL EcoRV (NEB R0195S) and 1× Buffer 3.1 (NEB). The linearized plasmid was purified using three Zymo clean and concentrator-500 columns (Cat. D4031) according to manufacturer's instructions and eluted in a total volume of 9 mL elution buffer.

8.1 mL of EcoRV-linearized pRGZ02 was combined with 5 mL 15 mM each rNTP (Fisher AC102800100, ICN15076591, AC226250010, ICN10121791), 1.5 mL 10× reaction buffer [500 mM Tris-Cl pH 8.1, 300 mM $MgCl_2$, 0.1% Triton X-100 and 20 mM Spermidine (ICN10047201)], 150 µL 1 M DTT, 105 µL RNAseOUT RNase inhibitor (Invitrogen 10777-019) and 150 µL 10 mg/mL T7 RNA polymerase expressed and purified according to (6) for a total reaction volume of 15 mL. The reaction mixture was incubated overnight at 37 °C. The template was digested by the addition of Turbo DNAse I to 0.04 U/µL and incubation at 37 °C for a further h. The reaction was terminated by the addition 7.5 mL RNA denaturing buffer [93% (v/v) HiDi formamide (Life Technologies 4311320), 0.04 M EDTA pH 8.0 and 0.5 mg/mL bromophenol blue]. The entire reaction was loaded on a 10% Polyacrylamide (29:1 Acrylamide:Bis-acrylamide), 0.5× TBE, 6 M Urea 23 cm long, 30 cm wide and 3 mm thick vertical PAGE gel. The gel was run in 0.5× TBE until the bromophenol blue band reached the bottom of the gel. The RNA was visualized by UV-shadowing and the third-fastest migrating band was excised, crushed, and passively eluted overnight in 0.3 M Sodium acetate at 0.05% (v/v) sodium dodecyl sulfate (SDS) at 4 °C. The RNA was precipitated by adding 2.5 volumes of ethanol and storing at –20 °C at least overnight. The RNA was pelleted by centrifugation at 4000 RCF for 60 min. The pellet was resuspended in 1 mL 70% (v/v) ethanol and transferred to a 1.5 mL Eppendorf tube and repelleted for 20 min at 21,000 RCF. The pellet was washed and repelleted a second time in 70% (v/v) ethanol then dissolved in DEPC-treated water (E&K Scientific EK-65062-500). The purity of the EMX1 sgRNA was confirmed by electrophoresis on a 15% TBE-UREA-PAGE gel (Biorad 4566036) followed by staining with Sybr Gold (Life Technologies S11494). The expected molecular weight was confirmed by mass spec.

LC-MS measurements of the sgRNA samples were conducted using Synapt G2 HDMS (Waters, Milford, MA) instrument equipped with Lockspray system, quadrupole mass analyzer, trap collision cell, and time-of-flight mass analyzer in tandem. Mass spectra were acquired and analyzed using MassLynx 4.1.1 software. Liquid chromatography was carried out using an ACQUITY UPLC system with an Agilent PRLP-S column (1000 Å, 5 µm, 50 × 2.1 mm) at a flow rate of 0.20 ml/min with buffer A consisting of 15 mM triethylamine (TEA) and 400 mM hexafluoroisopropanol (HFIP), and buffer B as 50% methanol in buffer A.



Purity and molecular weight verification of sgRNA EMX1. Insert: TBE-UREA-PAGE analysis of sgRNA EMX1. Main image: Deconvoluted mass spectrum of sgRNA EMX1. Additional higher molecular weight peaks are likely to represent salt adducts.

RNA sequence: Precursor transcript from pRGZ02

5′-3P-
GGGAGAUUGGAGGUGACCUGAUGAGUCCGUGAGGACGAAACGGUACCCGGUACCGUCGUCACC
UCCAAUGACUAGGGguuuaagagcuaugcuggaaacagcauagcaaguuuaaauaaggcuaguccguuaucaacuugaaaaaguggca
ccgagucggugcuuuuuuuggggucggcauggcaucuccaccuccucgcgguccgaccuggggcauccgaggaaacucggauggcuaagggagagc
cgaauucgau-3′OH

EMX1 sgRNA sequence

5′-OH-
GUCACCUCCAAUGACUAGGGguuuaagagcuaugcuggaaacagcauagcaaguuuaaauaaggcuaguccguuaucaacuugaa
aaaguggcaccgagucggugcuuuuuuu-2′:3′-P

Prior to RNP formation, sgRNA EMX1 was refolded. To generate 50 µL of refolded sgRNA at 30 µM concentration, 15 µL 100 µM guide in water was combined with 18.8 µL water, 5 µL 10× refolding buffer [200 mM HEPES-NaOH pH 7.45, 1.5 M NaCl] and 6.25 µL 80% (v/v) glycerol. The resulting mixture was incubated for 5 min at 70 °C then 5 min at ambient temperatures. 5 µL 10 mM $MgCl_2$ was added and the resulting mixture was incubated at 50 °C for 5 min then 5 min at ambient temperatures. This resulted in a sample containing 30 µM refolded guide, 20 mM HEPES-NaOH pH 7.45, 150 mM NaCl, 10% (v/v) glycerol and 1 mM $MgCl_2$. The refolded guide was either used immediately or stored at –80 °C until use.

6. Ribonucleoproteins (RNPs)

Preparation of Ribonucleoproteins (RNPs)

Protein was diluted to the desired concentration (25 µM, 12 µM, or 10 µM) with a final magnesium chloride concentration of 1 mM. This prepared solution was added to a prepared solution of sgRNA at the desired concentration (30 µM, 14.4 µM, or 12 µM) in a 1:1.2 Cas9:RNA molar ratio and the resulting solution was incubated at 37 °C for 10 min. The resulting RNP was cooled to room temperature and was either used immediately or frozen at -80ºC for later use.

Surface Plasmon Resonance Experiments

Experiments were performed on a Biacore^{TM} 3000 instrument (GE Healthcare). Chemically biotinylated ASGPr was captured onto a streptavidin sensor chip to levels ranging from 300 – 500 RU. Compound binding experiments were performed in 20 mM HEPES, pH 7.5, 150 mM NaCl, 20 mM $CaCl_2$, 5 mM $MgCl_2$, and 0.01% P20, at 25ºC.

Experiments were carried out in triplicate for four sample concentrations (10 nM, 3.3 nM, 1.1 nM, and 0.37 nM). Samples were injected at a flow rate of 50 µL/min, a total contact time of 120 sec and dissociation time of 400 sec. 1 min injections (2×) of 200 mM MES, pH 5.3 were performed after each sample injection in order to facilitate ASGPr regeneration.

Binding responses were processed using Scrubber 2 (BioLogic Software Pty Ltd) to zero, x-align, and double reference the data. Rate parameters (k_{on}, k_{off},) and corresponding affinity constant ($K_D = k_{off}/ k_{on}$) were determined by globally fitting the experimental data to a simple 1:1 interaction model using Biaeval (GE Healthcare).

Exp. Table 3: SPR binding

RNP description	Dissociation $t_{1/2}$ (s)	$k_{on} \times 10^6 \pm$ SE ($M^{-1}s^{-1}$)	$k_{off} \times 10^{-4} \pm$ SE (s^{-1})	$K_D \pm$ SE pM
Cas9-2lig-1NLS + EMX1 sgRNA RNP	4714.29 ± 34.64	3.14 ± 0.02	1.47 ± 0.01	46.8 ± 0.5
Cas9-2lig-AFr-1NLS + EMX1 sgRNA RNP	2406.25 ± 11.78	3.30 ± 0.03	2.88 ± 0.01	87.7 ± 0.8
Cas9-2lig-mCh + EMX1 sgRNA RNP	3705.88 ± 34.09	2.06 ± 0.007	1.87 ± 0.02	91.7 ± 0.9
Cas9-1NLS + EMX1 sgRNA RNP	n/a	n/a	n/a	No binding up to 10 nM
Cas9-AFr-1NLS + EMX1 sgRNA RNP	n/a	n/a	n/a	No binding up to 10 nM
Cas9-mCh + EMX1 sgRNA RNP	n/a	n/a	n/a	No binding up to 10 nM
AFg-Cas9-2lig-1NLS + EMX1 sgRNA RNP	14,142.86 ± 136.66	1.80 ± 0.005	0.49 ± 0.05	27.2 ± 0.3
AFg-Cas9-2lig-AFr-1NLS + EMX1 sgRNA RNP	10,612.56 ± 54.93	1.64 ± 0.002	0.65 ± 0.03	39.8 ± 0.2

Experiments were performed on a Biacore™ 3000 instrument (GE Healthcare). Chemically biotinylated ASGPr was captured onto a streptavidin sensor chip to levels ranging from 300 – 600 RU. Compound binding experiments were performed in 10 mM HEPES, pH 7.5, 150 mM NaCl, 20 mM $CaCl_2$, 0.01% P20, and 3% DMSO at 25ºC. Experiments were carried out in triplicate for four sample concentrations (10 nM, 3.3 nM, 1.1 nM, and 0.37 nM). Samples were injected at a flow rate of 50 µL/min, a total contact time of 120 sec and dissociation time of 400 sec. 1 min injections of 200 mM MES, pH 5.3 and 450 µM GalNAc were performed after each sample injection in order to facilitate ASGPr regeneration. Binding responses were processed using Scrubber 2 (BioLogic Software Pty Ltd) to zero, x-align, and double reference the data. Rate parameters (k_{on}, k_{off},) and corresponding affinity constant ($K_D = k_{off}/ k_{on}$) were determined by globally fitting the experimental data to a simple 1:1 interaction model using Biaeval (GE Healthcare).

Compound	Dissociation $t_{1/2}$ (s)	$k_{on} \times 10^6 \pm$ SE ($M^{-1}s^{-1}$)	$k_{off} \times 10^{-4} \pm$ SE (s^{-1})	$K_D \pm$ SE pM
4	587.29 ± 0.59	3.19 ± 0.002	11.8 ± 0.001	369.9 ± 0.4

In vitro Imaging Assays

HEPG2 cells (ATCC HB-8065) and SKHEP cells (HTB-52) were routinely passaged in low glucose DMEM (Thermofisher 1057014 supplemented with 10% FBS (Gibco 1610071) and 1% Penicillin/Streptomycin (Invitrogen 15140-122) in T75 flasks pre-coated with Gelatin (EMD-Millipore ES-006-B) at 37 °C under a 5% CO_2 atmosphere. For imaging experiments, cells were plated into 96 well Cell Carrier black walled microplates (Perkin Elmer) pre-coated with gelatin at a density of 40,000 cells per well. After adhering overnight, cells were treated with Dextran 488 (Thermo Fisher D22910) for 4 h to label endolysosomes and nuclei were labeled with Hoechst 33342 (Thermo Fisher 62249) for 30 min. RNPs were added at 50 pmol in 100 µL culture media and plates were loaded onto an Operetta CLS confocal imager (Perkin Elmer) with controlled environment (37 °C, 5% CO_2). Twelve fields were captured from each well every 15 min over 20 h using a 20× water immersion lens with optimal excitation and emission filters configuration to separately capture images of nuclei (Hoechst), endolysosomes (Dextran 488) and test articles of interest (Alexa Fluor 647 or mCherry). Quantification of labeled RNP internalization into cells and endolysosomal accumulation was performed using Harmony image analysis software (Perkin Elmer). Briefly, nuclei were identified and cell area was defined as a resized nucleus. Within the cellular area endolysosome spots (Dextran 488+ area) were defined using spot detection. Test article related spots (Alexa Fluor 647+ or mCherry+ area) were defined both in endolysosome spot regions, to quantify test article accumulation in endolysosomes, and in cellular area to quantify total cellular accumulation of the test article. Sum intensity values were generated per time point for test article positive spots, and test article positive endolysosome area on a per cell basis. A minimum of 10000 cells/well/time point were quantified for each experimental condition and values were calculated as the sum intensity of spots per cell, mean per well. Data was exported to Prism software (Graphpad) where graphs were prepared and statistical analysis (two-way ANOVA, multiple comparisons) was performed. Images were captured from Harmony software and processed in Metamorph (Molecular Devices) or Zen (Carl Zeiss) software for the final figures.

For the dual-labeled RNPs, cells were imaged in 4 channels to capture nuclei, Dextran 488, AF532 and AF647. Spot detection of each label was performed separately and depicted as the sum spot area per cell, mean per well.

For the ligand competition assay, cells were plated and prepared as above. Cells were pre-treated for 1 h with 1000, 2500 and 5000 pmol of ligand **4** in culture media. Culture media was removed and RNPs were then added at 50 pmol in 100 µL culture media in the presence of ligand at above concentrations. Microscopic imaging, image processing and data analysis and were carried out as above.

For data presented in Figure S2, HEPG2 and SKHEP cells were plated onto collagen coated MatTek glass bottomed wells (MatTek Corp, P35G-1.5-10-C) with the same experimental and culture conditions as above. RNPs were added at 64 pmol in 100 µL media and live cell images were collected on a Zeiss spinning disk confocal microscope (Carl Zeiss, Inc.) with controlled environment (37 °C, 5% CO_2). Images were processed in Zen Blue software (Carl Zeiss, Inc.) with the Hoechst nuclear label portrayed as blue, the Cas9-mCh portrayed as red, and the Dextran-647 endolysosomal label portrayed as green.

Cell culture delivery of Cas9-mCh constructs

Human HEPG2 and SKHEP cells were obtained from ATCC and cultivated in EMEM media (ATCC) supplemented with 10% FBS and penicillin/streptomycin at 37 °C with 5% CO_2. Cells were resuspended by incubation with 0.25% trypsin/EDTA.

For nucleofection assay, cells were harvested from culture plates, washed 1× with PBS and resuspended at 10,000,000 cells/mL in SF buffer (Lonza). After RNP preparation as described previously (5), 10 µL of Cas9 RNP (containing the desired amount of RNP, typically 10–200 pmol) was added to 20 µL of cells in SF buffer (Lonza) and transferred to wells of the 96-well nucleofection plate (Lonza). Cells were electroporated using the 96-well shuttle nucleofector (Lonza) and the HEPG2 settings. Cells were incubated at room temperature for 10 min following electroporation, resuspended in growth media and transferred onto 12-well culture plate (1 ml/well).

Cells were lysed in Quick Extract buffer (Epicenter) after 48 h incubation with RNP, incubated for 10 min at 65 °C followed by 10 min at 95 °C. Genomic DNA concentration was estimated by measuring Abs_{260nm}. The targeted locus was amplified by PCR (Kapa Biosystems) and revealed on agarose gel stained with SYBR Gold (Invitrogen) by comparison to a standard. 150 ng of PCR product was melted then hybridized and subjected to cleavage with T7 Endonuclease I (NEB). The resulting reaction was run on an agarose gel stained with SYBR Gold and the cleavage bands quantified using Image Lab (BioRad).

Cell culture delivery of Cas9-1NLS constructs

Tissue culture conditions

HEPG2 cells (ATCC HB–8065) and SKHEP cells (HTB-52) were routinely passaged in low glucose DMEM (Thermofisher 10567014 supplemented with 10% FBS (Gibco 16140071) and 1% Penicillin/Streptomycin (Invitrogen 15140-122) in T75 flasks pre-coated with Gelatin (EMD-Millipore ES-006-B) at 37 °C under a 5% CO_2 atmosphere.

Nucleofections and coincubations

Nucleofection and coincubation experiments were done in parallel with the same batch of cells. HEPG2 and SKHEP cells were typically grown to be at 80–90% confluence one day prior to the experiment and in log growth phase. Cells were detached using TrypLE Express (Gibco, 12605-010) resuspended in media and counted using a Countess (Thermofisher) automatic cell counter. For nucleofections, the cells were pelleted at 100 RCF for 5 min and the cell pellet was resuspended at $1x10^7$ cells/mL SF solution (Lonza V4XC-2024). 20 µL of the resulting cell suspension was combined with 4 µL of 12.5 µM RNP in a 16-well nucleocuvette strip or 96-well nucleocuvette plate then electroporated in a 96-well shuttle or 4D nucleofector system (Lonza) under the preset HEPG2 settings. The cells were resuspended in pre-warmed media, plated in gelatin-coated 24-well plates and returned to the incubator for 2 days. Genomic DNA was harvested using Epibio Quick Extract buffer (Cat # QE09050) according to manufacturer's directions except that the 65 °C and 95 °C incubations were done for 20 min each and quantified using a nanodrop ND-8000.

For coincubations, cells were diluted in media to 114,000 cells per mL. 700 µL (80,000 cells) were plated in each well of a 24-well gelatin-coated tissue culture plate and returned to the incubator for 30-60 minutes while the peptide-RNP coincubation mixes were prepared. The coincubation mixtures were made by adding 20 µL RNP buffer [20 mM HEPES-NaOH pH 7.45, 150 mM NaCl, 10% (v/v) glycerol and 1 mM $MgCl_2$] and 20 µL 12.5 µM RNP (250 pmol) to 500 µL OptiMEM (Thermofisher 31985062), followed by mixing by gentle pipetting and 5 min incubation at ambient temperatures. 75 µL of water and 75 µL 100 µM ppTG21 salt (7.5 nmol) were then added and the mixture was again incubated at ambient temperature for 5 min. For peptide-free incubations, 150 µL water was added instead. The combined 690 µL contained 0.36 µM RNP and 10.9 µM

ppTG21. The mixture was then added to the plated cells and they were returned to the incubator for two days (44-48 h).Genomic DNA was harvested using Epibio Quick Extract buffer (Cat # QE09050) according to manufacturer's directions except that the 65 °C and 95 °C incubations were done for 20 minutes each. The absorbance at 260 nm of the lysates was quantified using a nanodrop ND-8000.

T7E1 analysis

Genomic DNA samples were diluted to an A_{260} of 0.8 in 10 µL. A PCR product containing the on-target site was amplified using primers 5'- GCC ATC CCC TTC TGT GAA TGT TAG AC-3' and 5′- GGA GAT TGG AGA CAC GGA GAG CAG -3′ (IDT) and the Kappa Hifi HotStart PCR kit (Kappa Biosystems #KK2502) with 1× Kappa GC buffer, 0.3 mM each dNTP, 0.3 µM each primer and 0.015 U/µL Kappa DNA polymerase. The reaction was incubated in a thermal cycler programmed for 5 min at 95 °C followed by 29 cycles of 98 °C for 20 sec., 62 °C for 15 sec. and 72 °C for 30 sec. then a final extension at 72 °C for two min.

The PCR products were quantified by electrophoresis on a 2% egel48 (Thermofisher G800802) alongside a sample of known quantity. The gel was imaged using a Biorad Chemidoc XRS. The bands were quantified using densitometry with Image Studio Version 4.0 (LI-COR).

200 ng of each amplicon was diluted to 9 µL with 1× Kappa GC buffer. 1 µL 0.5 M KCl was added. The amplicon was denatured and rehybridized in a thermal cycler programmed to incubate for 10 min at 95 °C for 10 min followed by 1 min each at 85 °C, 75 °C, 65 °C, 55 °C, 45 °C, 35 °C, and 25 °C with a 2 °C/sec ramp rate. 3 µL water, 1.5 µL 10× NEB Buffer 2 and 0.5 µL 10 U/µL T7E1 (NEB M0302L) were added and the reactions were incubated at 37 °C for 30 min. The reactions were terminated by addition of 3.75 µL Hi-Density TBE sample buffer (5×) LC6678. Half of each sample was electrophoresed on a TBE/acrylamide PAGE gel, stained with Sybr-Gold, and imaged as before. The T7E1 cleaved and uncleaved bands were qualitatively assessed using densitometry with Image Studio Version 4.0.

Next generation sequencing

Targeted deep sequencing library preparation

The concentrations of genomic DNA samples were quantified using the Quant-iT Picogreen dsDNA assay kit (Thermofisher P7589). 2 µL of each sample was combined with 98 µL of the supplied TE buffer and 100 µL of a 200× dilution in TE of the supplied Picogreen dye. The supplied standard was diluted to the recommended concentration and analyzed alongside the samples. The plate was read in a Spectramax M5 with an excitation wavelength of 490 nm and an emission wavelength of 525 nm.

The targeted deep sequencing library preparation incorporated both unique molecular identifier (UMI) tags (9) and heterogeneity tags (10). To incorporate the UMI and heterogeneity tags, the target region was first subjected to two PCR cycles using a mixture of the primers listed in table S5. 30 ng genomic DNA was diluted to 10 µL with water. To the genomic DNA was added 12.5 µL of a master mix consisting of 5.8 µL water, 4.5 µL Phusion HF buffer, 0.5625 µL 10 mM each dNTP, 1.125 µL of the primer cocktail listed in table S5, and 0.5 µL Phusion Hot-start II DNA polymerase (Thermofisher F549S). The samples were denatured at 98 °C for five min the cycled at 98 °C for 10 sec, 61 °C for 2 min. and 72 °C for 15 sec. then cooled to ambient temperature. 1.5 µL Exonuclease I (NEB M0293S) was added to each sample. The reactions were incubated at 37 °C for 1 h. then 5 min at 98 °C to inactivate the exonuclease I.

For the HEPG2 samples, 25 µL of a mixture consisting of 11.9 µL of water, 10 µL 5× Phusion HF buffer, 0.5625 µL 10 mM each dNTP, 0.5 µL Phusion Hot Start II DNA polymerase, and 1 µL each of one forward and one reverse indexing primer from table S6 at 25 µM each. For the SKHEP samples a mixture of 25 µL of a mixture consisting of 16.9 µL of water, 5 µL 5× Phusion HF buffer, 0.5625 µL 10 mM each dNTP, 0.5 µL Phusion Hot Start II DNA polymerase, and 1 µL each of one forward and one reverse indexing primer from table S6 at 25 µM each. The reactions were incubated in a thermal cycler programmed for 30 cycles at 98 °C for 10 sec., 61 °C for 15 sec. and 72 °C for 15 sec. The amplicons were purified using 0.8 volumes (40 µL) AmpPure XP beads (Fisher Scientific NC9959336).

Sequencing

Targeted deep sequencing amplicons were received, quantitated using the Invitrogen QuBit dsDNA BR Assay kit (**Q32850**) per the manufacturer's recommended protocol on the QuBit 3.0 instrument (Q33216) and each library was analyzed for quality on the Agilent 2100 Bioanalyzer using the Agilent DNA 1000 assay kit (5067-1504). The amplicons passing the quality control steps were normalized to 10 nM concentration and equal amounts were pooled and setup for sequencing on the Illumina MiniSeq or NextSeq instrument following the guidelines stated in the Illumina Guide "MiniSeq System Denature and Dilute Libraries Guide". The final pooled and denatured library concentration used for sequencing was 1.8 pM with a spike-in of up to 50% Phix control (FC-110-3001), used as an enhancer for low diversity sequencing runs. Upon sequencing run completion, FASTQ files were generated for further analysis.

Sequence Analysis

The two paired end sequence FASTQ files were merged using PEAR version 2.3.0 (11). A custom python script (see p. 58) was used to extract the UMI tags from the sequences, de-duplicate them and output a new FASTQ containing the read for each UMI tag with the highest average quality score. The indel rates were determined using the Cas-Analyzer tool http://www.rgenome.net/cas-analyzer/#! (12) with the merged, de-duplicated FASTQ file. The input parameters were full reference sequence: GAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAA GAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGAC ATCGATGTCACCTCCAATGACTAGGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGC TGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCACTCCCT; protospacer sequence: GTCACCTCCAATGACTAGGG. For the parameters, the default values were used. These were comparison range: 70, Minimum frequency: 1, WT marker: used, 5. For further reading see: (9) and (13).

Preliminary Peptide:RNP binding evaluation

The preliminary peptide:RNP binding assay utilized the concept of SpeedScreen (Zehender, H., et. al., J. Biomol. Screen, 9, 498; 2004) which utilizes size exclusion chromatography in combination with liquid chromatography/electrospray ionization mass spectrometry in high throughput screening (HTS) to identify binders to targets. The same concept was utilized to determine peptide to RNP ratios semi-quantitatively by determining the molar concentration of each. In control experiments with free peptide in buffer, breakthrough of the free peptide was not detected and therefore any observed peptide in the experiments below is bound to the RNP.

The peptide and RNP analysis was carried out on an Agilent 6530 QTof mass spectrometer equipped with a Dual AJS electrospray source operated in positive ion mode. The mass spectrometer was interfaced with an Agilent 1290 UPLC system. The Agilent 1290 autosampler injected 10 µL aliquots of sample which was diluted to 0.1 mg/mL in MilliQ water just prior to analysis. The material was separated using a Agilent PLRP-S 1000Å 50 × 2.1mm with 5.0 µm particles column (part no. PL1912-1502). The mobile phases were: A) 0.1% formic acid in water and B) 0.1% formic acid in acetonitrile. Raw mass spectra were viewed using MassHunter (version B.07.00 Service Pack 2, Agilent) and mass spectral deconvolution was performed using BioConfirm (B.07.00, Agilent). Due to significant dynamic range differences between the RNP and peptides, a separate analysis was needed for both determinations.

RNPs were prepared in a 1:1.2 ratio (protein:sgRNA targeting EMX1) at a concentration of 10 µM. The RNP solution was diluted to 5.75 µM using RNP buffer. To this solution of RNP was added 30 molar equiv. of a 1 mM (DMSO (1.1%)/water) ppTG21 TFA salt and incubated at room temperature for 1 h. The slight suspension was desalted using 0.5 mL Zeba spin desalting columns (40K MWCO) (eluting with NaCl GF buffer, using the manufacturer's instructions) and the filtrate was utilized in the protein/peptide SpeedScreen experiments.

For the RNP experiment, samples were diluted 1:10 just prior to analysis. Additionally, cytochrome C was added to the sample to a final concentration of 1 µM as an internal standard due to instrument variation. A calibration curve was generated for each RNP from 0–3 µM. The calibration curve was run before and after the sample analysis. The concentration of the RNP was determined using the total ion chromatogram (TIC) peak for the Cas9 protein peak. The area from Cas9 peak was normalized using the TIC peak from cytochrome C. The final concentration was determined via back calculation from the calibration curve. The accuracy for the back calculation of the calibration points ranged from 70–102% for each RNP for a given point.

For the peptide analysis, samples were diluted 1:50 just prior to analysis. Extracted ion chromatograms were generated for the different peptides using the +/– 20 ppm window around the $(M+3H)^{3+}$ charge state. The resulting area for the peptide peak was utilized to generate a calibration curve for quantitation purposes. A calibration curve was generated from 0.0–5.0 µM. The accuracy for the back calculation of the calibration points ranged from 70–130% for a given point. As in the case for the RNP, the calibration curve was run before and after the samples. The final concentration was determined via back calculation from the calibration curve.

Primary Human Hepatocyte Experiments

Primary Human Hepatocyte Culture

Cryopreserved human hepatocytes were obtained commercially from ThermoFisher (Lot#4165). Cells were reconstituted according to the ThermoFisher protocol using Cryopreserved Hepatocyte Recovery Medium (CHRM). The cells were plated on collagen I–coated BD CellCarrier 24 (gene editing endpoints) or 96-well plates (imaging endpoints) at 450,000 or 60,000 cells/well, respectively. Cells were applied to culture dishes in hepatocyte plating medium (Dulbecco's Minimal Essential Medium with 5% FBS, 50 U/mL penicillin, 50 µg/mL streptomycin, 4 µg/mL bovine insulin). All media was purchased from ThermoFisher Scientific, while all media supplements were purchased from VWR International. After 4-6 h under standard incubation conditions (37 °C, 5% CO_2, 100% humidity) hepatocytes had attached to the culture dish surface. The media was then changed to hepatocyte culturing medium (Williams E medium containing 1× ITS+ supplement, 15 mM HEPES, 50 U/mL penicillin, 50 µg/mL streptomycin, 2 mM L-glutamine, 1 µM trichostatin, and 0.1 µM dexamethasone) and the cells were incubated for 20-24 h. On the second day, media was removed and a Matrigel overlay (0.25 mg/mL in hepatocyte culturing media) was applied to the cells according to the ThermoFisher recommended protocol.

Plated Primary Human Hepatocyte Coincubation with RNP

On the third day, the cells underwent a medium change with 700 µL of hepatocyte culturing medium. The coincubation mixtures were made by adding 20 µL RNP buffer [20 mM HEPES-NaOH pH 7.45, 150 mM NaCl, 10% (v/v) glycerol and 1 mM $MgCl_2$] and 20 µL 12.5 µM RNP (250 pmol) to 500 µL OptiMEM (Thermofisher 31985062), followed by mixing by gentle pipetting and 5 min. incubation at ambient temperatures. 75 µL of water and 75 µL 100 µM ppTG21 salt (7.5 nmol) were then added and the mixture was again incubated at ambient temperature for 5 min. For peptide-free incubations, 150 µL water was added instead. The combined 690 µL contained 0.36 µM RNP and 10.9 µM ppTG21. The mixture was then added to the plated hepatocytes and they were returned to the incubator for two days (44-48 h). Genomic DNA was harvested using Epibio Quick Extract buffer (Cat # QE09050) according to manufacturer's directions except that the 65 °C and 95 °C incubations were done for 20 minutes each. The absorbance at 260 nm of the lysates was quantified using a nanodrop ND-8000. The genomic material was harvested and qualitatively evaluated by T7E1 assay (below).

Primary Human Hepatocyte RNP Internalization Imaging

On the third day, the cells underwent a medium change with hepatocyte culturing medium containing 50 µg/mL pHrodo Green Dextran (Thermo Fisher) to label lysosomes. The following day, hepatocyte culture media was exchanged and fresh media containing 1 µg/mL Hoechst dye (ThermoFisher) was added. Thirty minutes later media was removed and 100 µL of hepatocyte culture media containing 50 pmol of RNP pre-complexed with a 30:1 molar ratio of ppTG21 peptide (pre-complex method described in "Nucleofections and Coincubation" section of methods).

Cultured cells were then loaded into an Operetta CLS confocal imager (Perkin Elmer) with controlled environment (37 °C, 5% CO_2). Twelve fields were captured from each well every 15 minutes over 20 h using a 20× water immersion lens with optimal excitation and emission filters configuration to separately capture images of nuclei (Hoechst), lysosomes (pHrodo) and test articles of interest (AlexaFluor 647). Quantification of labeled RNP internalization into cells and lysosomal accumulation was performed using Harmony image analysis software (Perkin Elmer). Briefly, nuclei were identified and cell area was defined as a resized nucleus.

Within the cellular area lysosome spots (pHrodo+ area) were defined using spot detection. Test article related spots (Alexa Fluor 647+) were defined both in lysosome spot regions, to quantify test article accumulation in lysosomes, and in cellular area to quantify total cellular accumulation of the test article. Sum intensity values were generated per time point for test article positive spots, and test article positive lysosome area on a per cell basis. A minimum of 5000 cells/well/time-point were quantified for each experimental condition and values were calculated as the sum intensity of spots per cell, mean per well. Data was exported to Prism software (Graphpad) where graphs were prepared and statistical analysis (two-way ANOVA, multiple comparisons) was performed. Images were captured from Harmony software and processed in Metamorph (Molecular Devices) or Zen (Carl Zeiss) software for the final figures (Figure S17).

T7E1 Assay following plated co-incubation

A T7E1 assay for cleavage at the EMX1 site was performed as described in section 6.4 (Figure S18).

Primary Human Hepatocyte Culture for Suspension Co-incubation

These sections describe a co-incubation experiment that was performed with a co-incubation of primary human hepatocytes and Cas9 RNP that was initiated while cells were in suspension. Pooled human hepatocytes HEP10 were obtained commercially from ThermoFisher (HMCS10). Cells were reconstituted according to the ThermoFisher protocol using Cryopreserved Hepatocyte Recovery Medium (CHRM®). After centrifugation at 100×*g* for 10 minutes, cells were re-suspended in Incubation Medium prepared with Williams Medium E (A1217601) and supplemented with Hepatocyte Maintenance Supplement Pack, Serum-free (CM4000) including HGF at 10 ng/mL (ThermoFisher, PHG0254). Hepatocytes were immediately plated in a 24 well plate at 400,000 cells/well. Cells were maintained for 24 h under standard incubation conditions (37 °C, 5% CO_2, 100% humidity).

Suspension Primary Human Hepatocyte Co-incubation with RNP

One day after thawing, cells were recovered and centrifuged at 100×*g* for 10 minutes to remove dead cells and cell debris. Cells were re-suspended in fresh pre-warmed incubation media supplemented with HFG 10 ng/mL and plated on a 96 well plate at 80,000 cells/well. Fresh RNP was prepared at 6.25 µM concentration in RNP buffer [20 mM HEPES-NaOH pH 7.45, 150 mM NaCl, 10% (v/v) glycerol and 1 mM $MgCl_2$]. First, folding of the sgRNA was performed by heating at 95 °C for 5 minutes and gradual cooling to RT, and then RNP formation was performed with Cas9-2lig-1NLS:sgRNA at a molar ratio (1:1.2) respectively, by incubating the mix at 37 °C for 10 min. Co-incubation mixtures were prepared in OptiMEM (Thermofisher, 31985062) prior to addition to the cells as follows: 20 µL (125 pmol) or 8 µL (50 pmol) of RNP (based on Cas9) were added to 25 µL of OptiMEM and 30 molar equivalents (relative to Cas9) of 100 µM ppTG21 (3750 or 1500 pmol) were added to other 25 µL of OptiMEM separately and incubated for 5 minutes at room temperature. For peptide-free incubations, same volume of 1×PBS was added to the mixture instead. Then RNP and peptide were mixed together and incubated for an additional 10 min at room temperature until the co-incubation mixture was finally added to the plated hepatocytes. After 48 h in culture, cells were washed once with PBS and genomic DNA was isolated by using Epibio Quick Extract buffer (Cat # QE09050) following manufacturer's directions. Genomic DNA concentration was estimated by using the Nanodrop absorbance at 260 nm of the lysates.

T7E1 Assay Following Suspension Co-incubation

A T7E1 assay for cleavage at the EMX1 site was performed as described in section 6.4 (Figure S19).

7. References

1. Liras S, Mascitti V, Thuma B, Doudna J, Rouet R (2017) Tissue-specific genome engineering using CRISPR-Cas9. Patent WO 2017083368 A1.

2. Rittner K et al. (2002) New basic membrane-destabilizing peptides for plasmid-based gene delivery in vitro and in vivo. *Molecular therapy: the Journal of the American Society of Gene Therapy* 5:104-14.

3. Jinek M et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* 337:816-21.

4. Ding Q et al. (2014) Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circulation research* 115:488-92.

5. Lin S, Staahl B, Alla R, Doudna J (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3:e04766.

6. Ellinger T, Ehricht R (1998) Single-step purification of T7 RNA polymerase with a 6-histidine tag. *BioTechniques* 24:718-20.

7. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PloS one* 3:e3647.

8. Avis JM, Conn GL, Walker SC (2012) Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5“ and 3” ends. *Methods in molecular biology (Clifton, N.J.)* 941:83-98.

9. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 108:9530-5.

10. Fadrosh DW et al. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2:6.

11. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England)* 30:614-20.

12. Park J, Lim K, Kim J, Bae S (2017) Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics (Oxford, England)* 33:286-288.

13. Kukita Y et al. (2015) High-fidelity target sequencing of individual molecules identified using barcode sequences: de novo detection and absolute quantitation of mutations in plasma cell-free DNA from cancer patients. *DNA research : an international journal for rapid publication of reports on genes and genomes* 22:269-77.

8. Supporting Tables & Figures

RNP	cells	peptide	method	[RNP] (pmol)	[peptide] (pmol)	% indels (each replicate)										avg. % indels	std. dev.	geom. mean	90% c.i. lower	90% c.i. upper
						r1	*r2*	*r3*	*r4*	*r5*	*r6*	*r7*	*r8*	*r9*	*r10*					
Cas9-2lig-1NLS	HEPG2	ppTG21	co-incubate	250	7,500	4.9	5	4.2	4.1	4.9	3.3	4.7	6.1	4.3	6.2	4.8	0.9	4.7	4.2	5.2
Cas9-1NLS	HEPG2	ppTG21	co-incubate	250	7,500	0.2	0.1	0.3	0.1	0.3	1.7	0.8	0.8			0.5	0.5	0.3	0.2	0.7
Cas9-2lig-1NLS	HEPG2	none	incubate	250	0	0.6	0.1	0.1	0.1	0.1	0.1	0	0.1			0.2	0.2	0.1	0.1	0.2
Cas9-1NLS	HEPG2	none	incubate	250	0	0	0.1	0.4	0.2	0	0.1	0.1				0.1	0.1	0.1	0.1	0.2
Cas9-2lig-1NLS	HEPG2	none	nucleofect	50	N/A	30.3	31.1	22	14	20	20.9	21.5	22.4	20.6	18	22.1	5.2	21.6	18.9	24.6
Cas9-1NLS	HEPG2	none	nucleofect	50	N/A	35.1	40.6	22.8	15.6	27.3	27.3	14.8	19.3			25.4	9.1	24.0	18.8	30.6
untreated control	HEPG2	none	no treatment	0	0	0	0.2	0	0.2	0	0	0	0.1			0.1	0.1	0.1	0.1	0.1
Cas9-2lig-1NLS	SKHEP	ppTG21	co-incubate	250	7,500	0.1	0.2	0.3								0.2	0.1	0.2	0.1	0.5
Cas9-1NLS	SKHEP	ppTG21	co-incubate	250	7,500	0.3	0.2	0.1								0.2	0.1	0.2	0.1	0.5
Cas9-2lig-1NLS	SKHEP	none	incubate	250	0	0.1	0.1	0.2								0.1	0.1	0.1	0.1	0.2
Cas9-1NLS	SKHEP	none	incubate	250	0	0.3	0.2	0.1								0.2	0.1	0.2	0.1	0.5
Cas9-2lig-1NLS	SKHEP	none	nucleofect	50	N/A	9.8	10.8	10.4								10.3	0.5	10.3	9.5	11.2
Cas9-1NLS	SKHEP	none	nucleofect	50	N/A	13.7	15.9	14.5								14.7	1.1	14.7	12.9	16.7
untreated control	SKHEP	none	no treatment	0	0	0.2	0.2	0.1								0.2	0.1	0.2	0.1	0.3

r1-6 was at Groton, r7-10 was at Berkeley

Table S1. Percentage indel rates in HEPG2 and SKHEP cells measured by UMI-library sequencing. RNP: each RNP consists of the listed Cas9 construct in complex with an sgRNA targeting EMX1. Method: RNP was delivered by nucleofection, co-incubation (with 30 molar equivalents of ppTG21), or incubation (no ppTG21). Listed pmol concentrations refer to the amount of RNP and peptide used per well for each replicate. Replicates 1–6 were performed at Pfizer (Groton, CT); replicates 7–10 were performed at U.C. Berkeley. Average % indels is the arithmetic mean of the listed replicates (if n≥3). The arithmetic mean, standard deviation of the arithmetic mean, geometric mean, and upper or lower 90% confidence intervals of the geometric mean were calculated using GraphPad Prism© version 7.02.

Sample	peptide back calc (μM)	peptide dilution factor	peptide conc (μM)	protein back calc (μM)	protein dilution factor	protein conc (μM)	ratio peptide/ protein
Cas9-2lig-AFr-1NLS + ppTG21	3.86	50	193.0	0.263	10	2.63	73.5
Cas9-AFr-1NLS + ppTG21	2.61	50	130.6	0.186	10	1.86	70.3
Cas9-2lig-1NLS + ppTG21	3.01	50	150.6	0.241	10	2.41	62.4
Cas9-1NLS + ppTG21	2.93	50	146.6	0.516	10	5.16	28.4

Table S2. Calculated peptide, protein and ratios following co-incubation (RNP + ppTG21) and SEC-mediated filtration. RNPs were formed using sgRNA targeting EMX1. Peptide recovery was not observed in the absence of RNP.

Sample	protein back calc (μM)	dilution factor	protein conc (μM)	expected protein conc. (μM)	protein recovery (%)
Cas9-2lig-AFr-1NLS + ppTG21	0.263	10	2.63	4.9	54
Cas9-AFr-1NLS + ppTG21	0.186	10	1.86	4.9	38
Cas9-2lig-1NLS + ppTG21	0.241	10	2.41	4.9	49
Cas9-1NLS + ppTG21	0.516	10	5.16	4.9	105
Cas9-2lig-1NLS	0.406	10	4.06	4.9	83
Cas9-1NLS	0.438	10	4.38	4.9	89

Table S3. Protein recovery based on expected concentration following co-incubation (RNP + ppTG21) and SEC-mediated filtration. RNPs were formed using sgRNA targeting EMX1.

Sample	peptide back calc (µM)	dilution factor	peptide conc (µM)	expected peptide conc (µM)	peptide recovery (%)
Cas9-2lig-AFr-1NLS + ppTG21	3.86	50	193.0	147	131
Cas9-AFr-1NLS + ppTG21	2.61	50	130.6	147	89
Cas9-2lig-1NLS + ppTG21	3.01	50	150.6	147	102
Cas9-1NLS + ppTG21	2.93	50	146.6	147	100

Table S4. Peptide recovery based on expected concentration following co-incubation (RNP + ppTG21) and SEC-mediated filtration. Peptide recovery was not observed in the absence of RNP.

Name	Sequence
EMX1v2f_+0	CCTACACGACGCTCTTCCGATCTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+1	CCTACACGACGCTCTTCCGATCTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+2	CCTACACGACGCTCTTCCGATCTTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+3	CCTACACGACGCTCTTCCGATCTGTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+4	CCTACACGACGCTCTTCCGATCTCGTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+5	CCTACACGACGCTCTTCCGATCTCCGTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+6	CCTACACGACGCTCTTCCGATCTACCGTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2f_+7	CCTACACGACGCTCTTCCGATCTAACCGTTDHVBDHVBGAACCGGAGGACAAAGTACAAA
EMX1v2r_+0	TTCAGACGTGTGCTCTTCCGATCTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+1	TTCAGACGTGTGCTCTTCCGATCTTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+2	TTCAGACGTGTGCTCTTCCGATCTTTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+3	TTCAGACGTGTGCTCTTCCGATCTGTTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+4	TTCAGACGTGTGCTCTTCCGATCTCGTTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+5	TTCAGACGTGTGCTCTTCCGATCTCCGTTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+6	TTCAGACGTGTGCTCTTCCGATCTACCGTTDHVBDHVBCAGGGAGTGGCCAGAGTC
EMX1v2r_+7	TTCAGACGTGTGCTCTTCCGATCTAACCGTTDHVBDHVBCAGGGAGTGGCCAGAGTC

Table S5. Primer cocktail for UMI and heterogeneity tag incorporation. Each oligo is present in the stock at 140.6 nM. Light blue text is the heterogeneity spacer and the yellow text is the UMI tag. The ambiguous nucleotide codes are as follows B: C,G, or T; D: A,G, or T; H: A,C, or T; V:A,C, or G. All were synthesized at IDT.

Name	Sequence	Index Sequence
Forward primers for second round		
i501	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	TATAGCCT
i502	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	ATAGAGGC
i503	AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	CCTATCCT
i504	AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	GGCTCTGA
i505	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	AGGCGAAG
i506	AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	TAATCTTA
i507	AATGATACGGCGACCACCGAGATCTACACCAGGACGTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	CAGGACGT
i508	AATGATACGGCGACCACCGAGATCTACACGTACTGACACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	GTACTGAC
Reverse primers for second round		
i701	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	ATTACTCG
i702	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	TCCGGAGA
i703	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	CGCTCATT
i704	CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	GAGATTCC
i705	CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	ATTCAGAA
i706	CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	GAATTCGT
i707	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	CTGAAGCT
i708	CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	TAATGCGC
i709	CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	CGGCTATG
i710	CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	TCCGCGAA
i711	CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	TCTCGCGC
i712	CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	AGCGATAG

Table S6. Primers for the PCR amplification and sample index incorporation for targeted Deep sequencing library preps. *Indicates phosphorothioate linkage; All primers were synthesized by IDT and resuspended to a concentration of 25 μM.

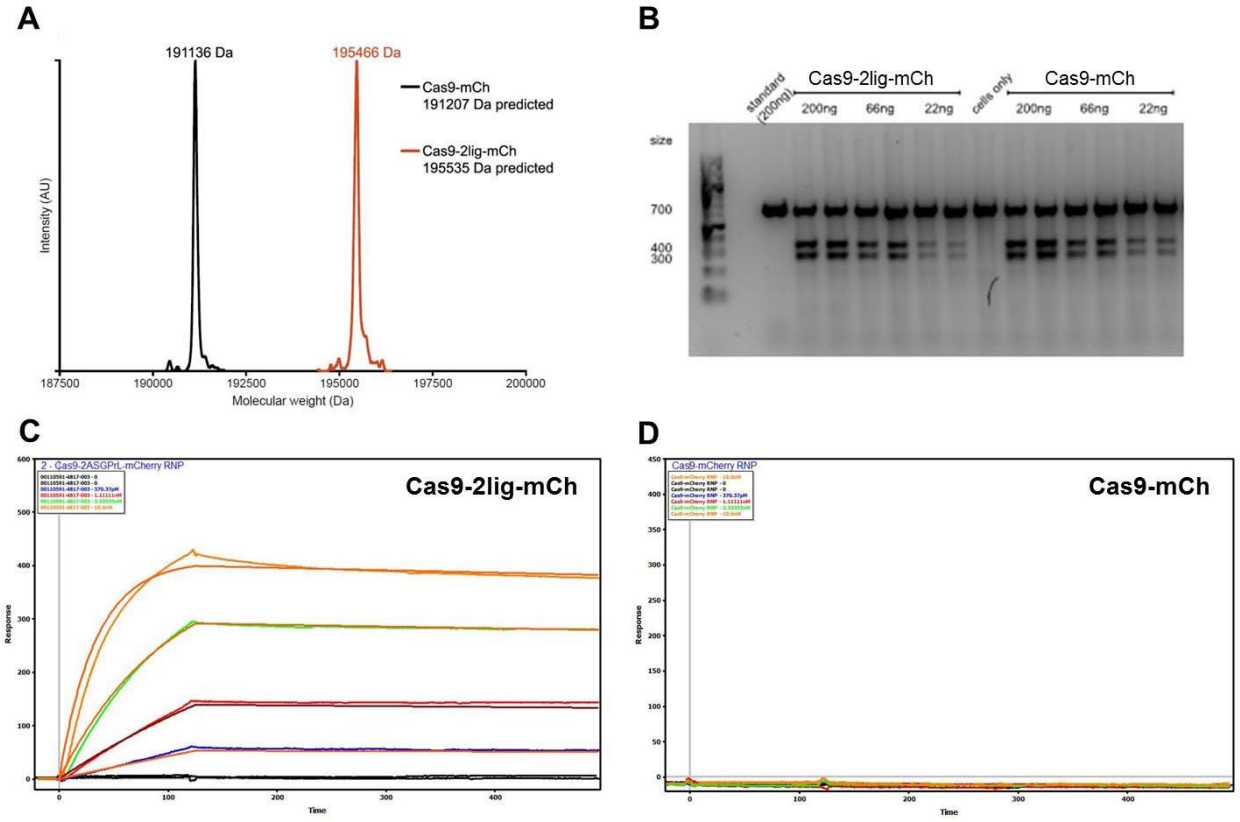


Figure S1. *In vitro* characterization of Cas9-2lig-mCh. (A) Mass spectrometry analyses on Cas9 mCherry unligated (in black) and Cas9-2lig-mCh (in red) shows bis ligation of the ligand onto Cas9. (B) *In vitro* cleavage activity in HEPG2 cells. 200, 66 and 11 ng of Cas9-2lig-mCh or Cas9-mCh RNP were nucleofected in HEPG2 cells and incubated for 48 h at 37 °C. The genomic material was harvested and editing was qualitatively determined by T7E1 assay; associated gel shown. (C) Cas9-2lig-mCh RNP binding kinetics to ASGPr (K_D = 91.7 ± 0.9 pM); time on X axis reported in seconds. (D) Cas9-mCh RNP does not bind to ASGPr; Time on X axis reported in seconds. Corresponding RNPs made from sgRNA targeting EMX1.

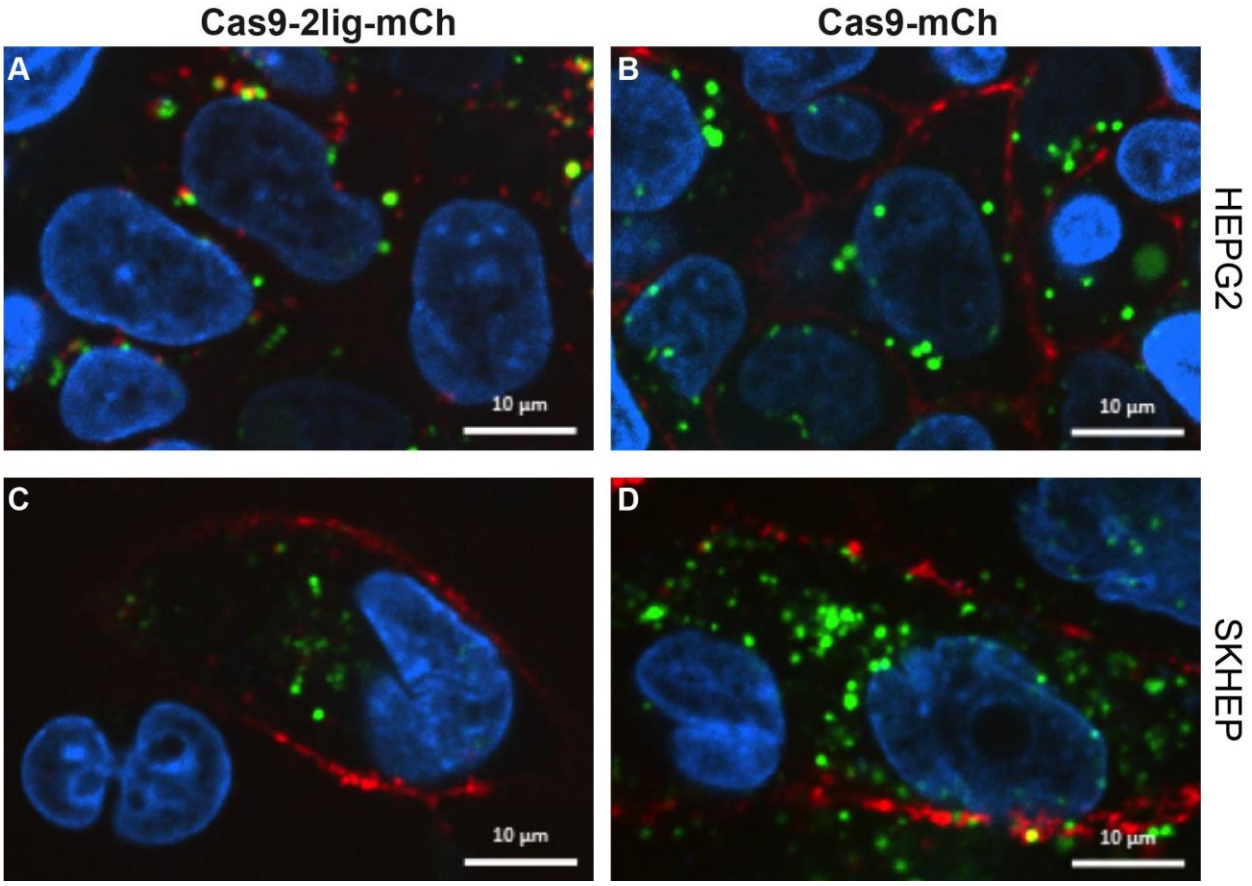


Figure S2. Internalization in HEPG2 cells (ASGPr+, in A, B) or SKHEP cells (ASGPr–, in C, D) of Cas9-2lig-mCh (A, C) and Cas9-mCh (B, D) RNPs observed by live cell imaging at 1 h (corresponding RNPs made from sgRNA targeting EMX1). Zoomed in images of HEPG2 cells show cytoplasmic puncta and cell surface accumulation for the ligated and unligated RNP, respectively. Blue: Hoechst stain of cell nuclei; Red: Intracellular Cas9 visualized via mCherry fluorescence; Green: Endolysosomal compartment stained using dextran488. Note that this preliminary experiment was performed using distinct conditions from the other microscopy in this study; see page S33 of this document.

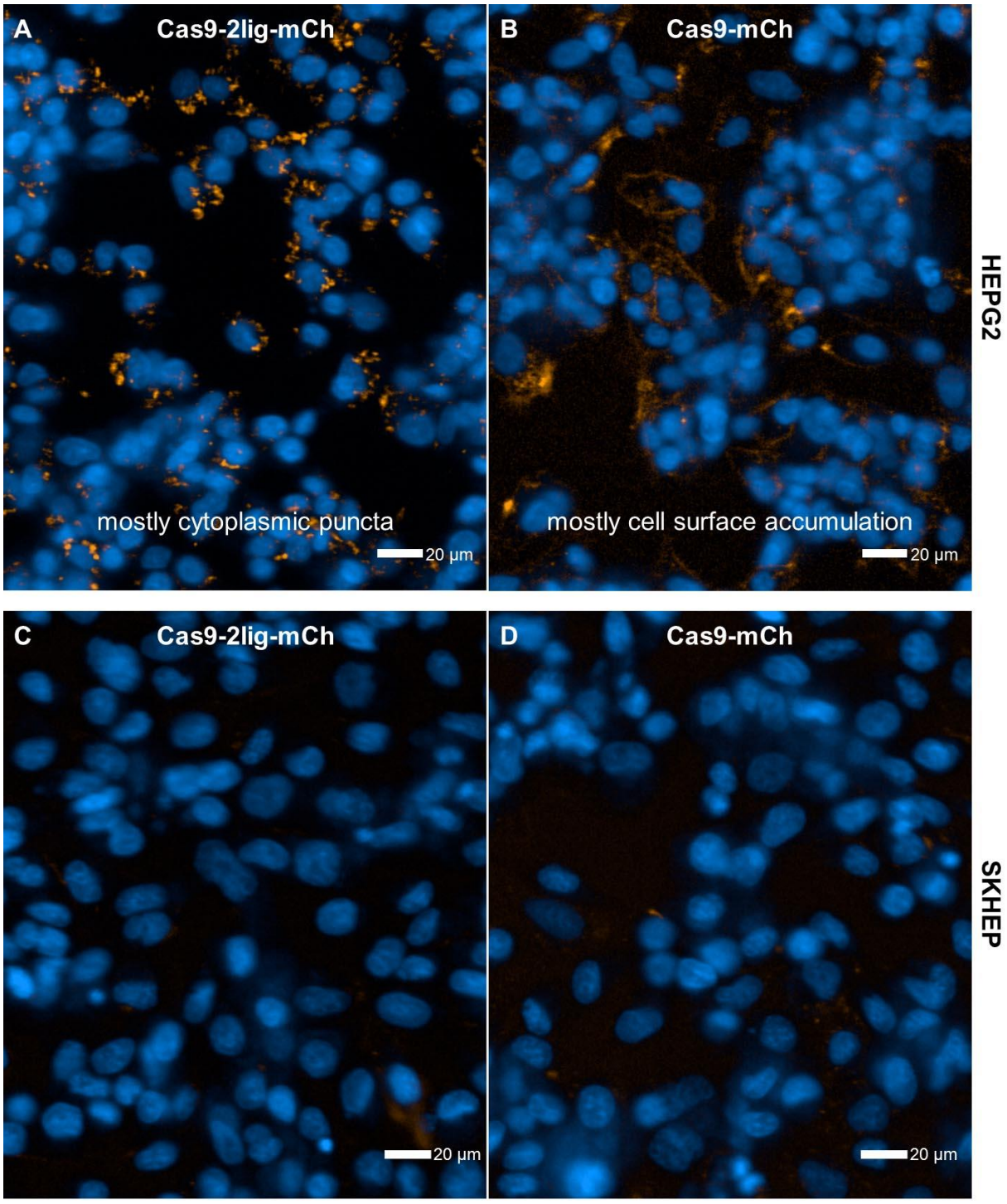


Figure S3. Internalization in HEPG2 cells (ASGPr+, in A, B) or SKHEP cells (ASGPr–, in C, D) of Cas9-2lig-mCh (A, C) and Cas9-mCh (B, D) RNPs observed by live cell imaging at 2.5 h (corresponding RNPs made from sgRNA targeting EMX1). Zoomed in images of HEPG2 cells show cytoplasmic puncta and cell surface accumulation for the ligated and unligated RNP, respectively. Blue: Hoechst stain of cell nuclei; Orange: Intracellular Cas9 visualized via mCherry fluorescence. Contrast of images is normalized so as to allow comparison with constructs in Fig. S7.

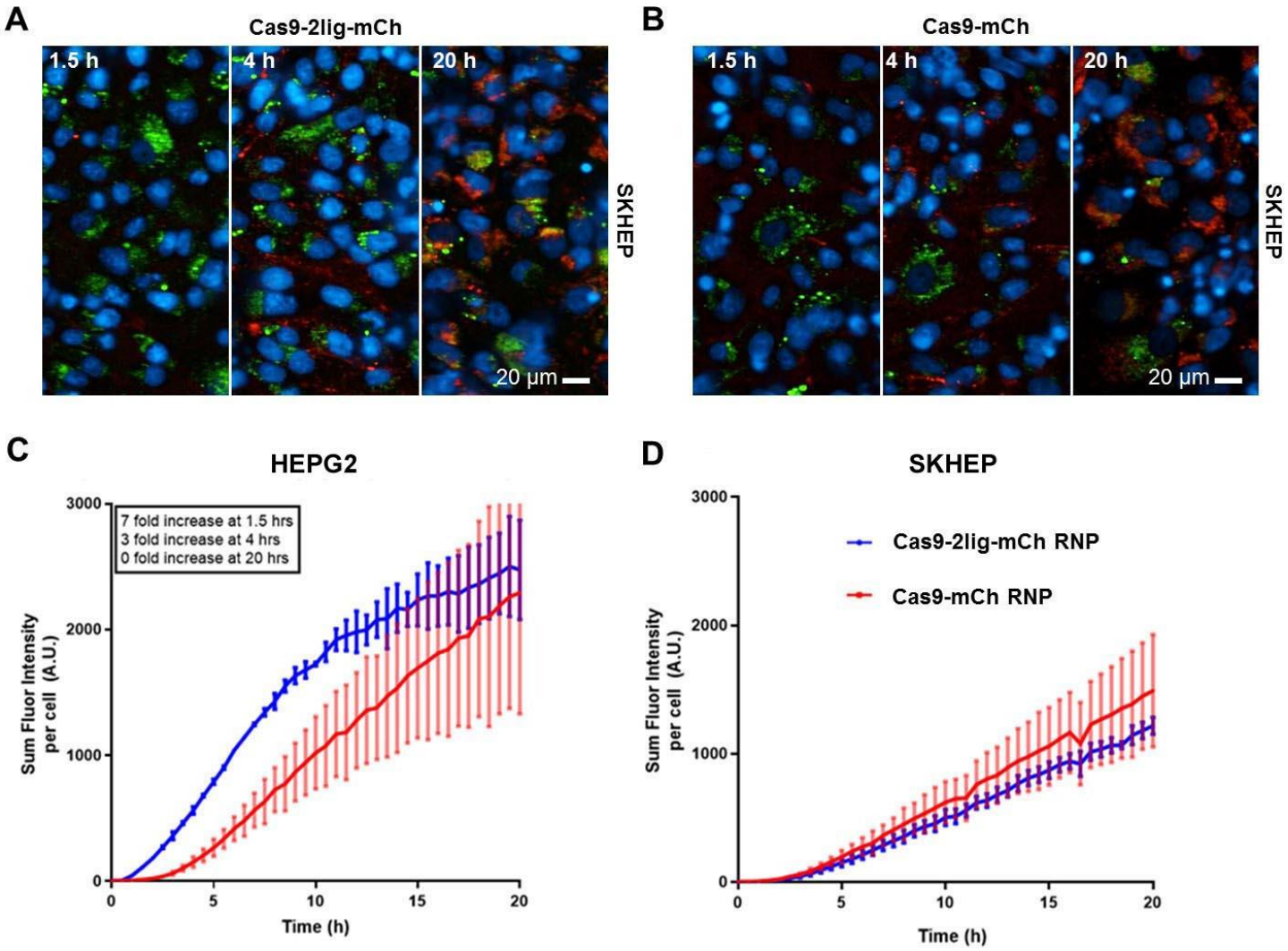


Figure S4. Internalization in SKHEP cells (ASGPr−) of Cas9-2lig-mCh (A) and Cas9-mCh (B) RNPs observed by live cell imaging at 1.5, 4, and 20 h. (C) Quantification of intracellular RNP accumulation in HEPG2 cells over 20 h (same as in Fig. 2C). (D) Quantification of intracellular RNP accumulation in SKHEP cells over 20 h (same as in Fig. 2C). Blue: Hoechst stain of cell nuclei; Green: Endolysosomal compartment stained using dextran488; Red: Intracellular Cas9 visualized via mCherry fluorescence. Fluorescence intensity was quantified using the sum of spots per cell (mean per well). Each data point (C,D) represents 3 technical replicate wells, with a minimum of 10,000 cells quantified per well. For (C) and (D), arithmetic means and standard deviations of the mean were calculated and plotted using GraphPad Prism© version 7.02. Corresponding RNPs made from sgRNA targeting EMX1.

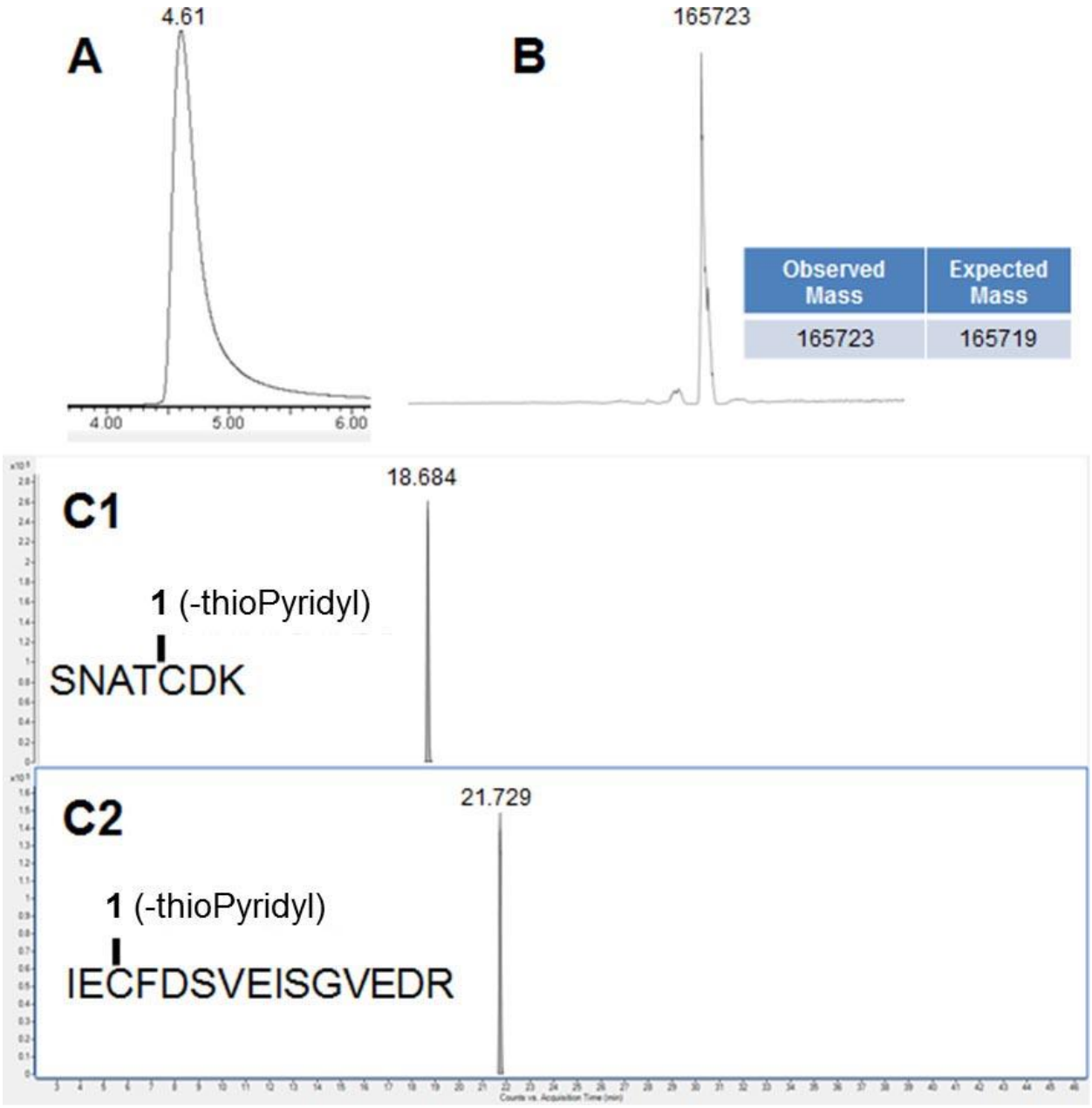


Figure S5. *In vitro* characterization of Cas9-2lig-1NLS. (A) Size exclusion chromatography trace and retention time expressed in minutes (see experimental part for conditions). (B) Deconvoluted QTOF Mass spectra (C) Trypsin digest data showing evidence for the regioselective conjugation of Cas9 with **1** (-2 thioPyridyl); (C1) shows the Extracted Compound Chromatogram for the conjugated peptide SNATCDK (sequence 1-7) to **1** (-2 thioPyridyl); retention time reported in minutes. (C2) shows the Extracted Compound Chromatogram for the conjugated peptide IECFDSVEISGVEDR (sequence 576-590) to **1** (-2 thioPyridyl); retention time reported in minutes.

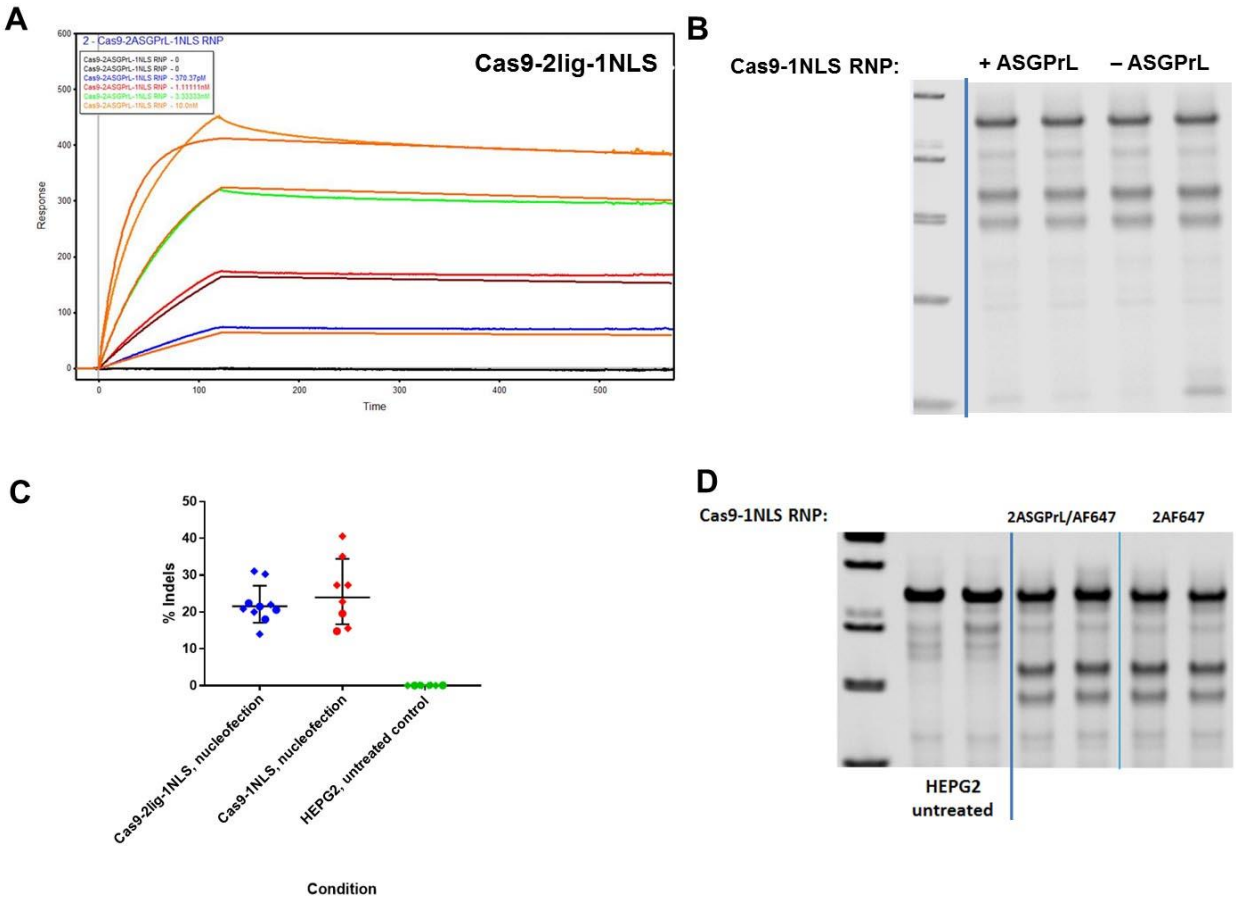


Figure S6. Confirmation of ASGPr binding and functional activity of Cas9-2lig-1NLS (A-C) and Cas9-2lig-AFr-1NLS (D). (A) Cas9-2lig-1NLS RNP binding kinetics to ASGPr (K_D = 46.8 ± 0.5 pM); time on X axis reported in seconds. (B) 50 pmol of Cas9-2lig-1NLS RNP or Cas9-1NLS RNP were nucleofected in HEPG2 cells and incubated for 48 h at 37 °C. The genomic material was harvested and qualitatively evaluated by T7E1 assay; associated gel shown. (C) Percentage indel rates (under nucleofection conditions) derived from deep sequencing. (n = 8-10 replicates; see also Table S1). Blue points represent samples treated with Cas9-2lig-1NLS RNP, red points represent samples treated with Cas9-1NLS RNP and green represents untreated controls. Diamonds represent assays done at Pfizer (Groton, CT) and circles represent assays done at UC-Berkeley. The midpoint bars depict the geometric mean and the error bars depict the geometric standard deviation. The image was generated using Graphpad Prism© version 7.02. (D) 50 pmol of Cas9-2lig-AFr-1NLS RNP or Cas9-AFr-1NLS RNP were nucleofected in HEPG2 cells and incubated for 48 h at 37 °C. The genomic material was harvested and qualitatively evaluated by T7E1 assay; associated gel shown. Corresponding RNPs made from sgRNA targeting EMX1.

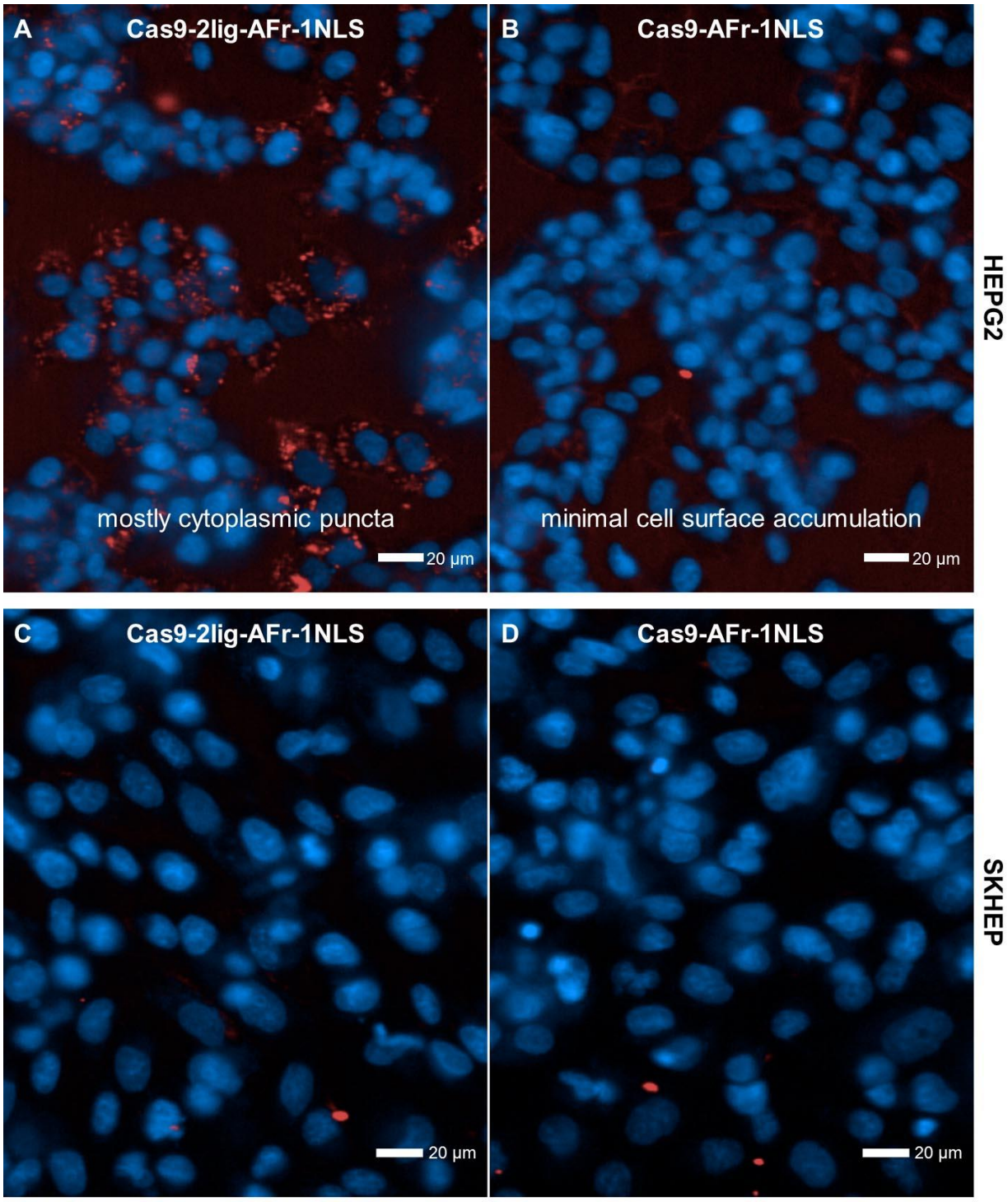


Figure S7. Internalization in HEPG2 cells (ASGPr+, in A, B) or SKHEP cells (ASGPr–, in C, D) of Cas9-2lig-AFr-1NLS (A,C) and Cas9-AFr-1NLS (B,D) RNPs observed by live cell imaging at 2.5 h (corresponding RNPs made from sgRNA targeting EMX1). Zoomed in images of HEPG2 cells show cytoplasmic puncta and minimal cell surface accumulation for the ligated and unligated RNP, respectively. Blue: Hoechst stain of cell nuclei; Red: Intracellular Cas9 visualized via AF647 fluorescence. Contrast of images is normalized so as to allow comparison with constructs in Fig. S3.

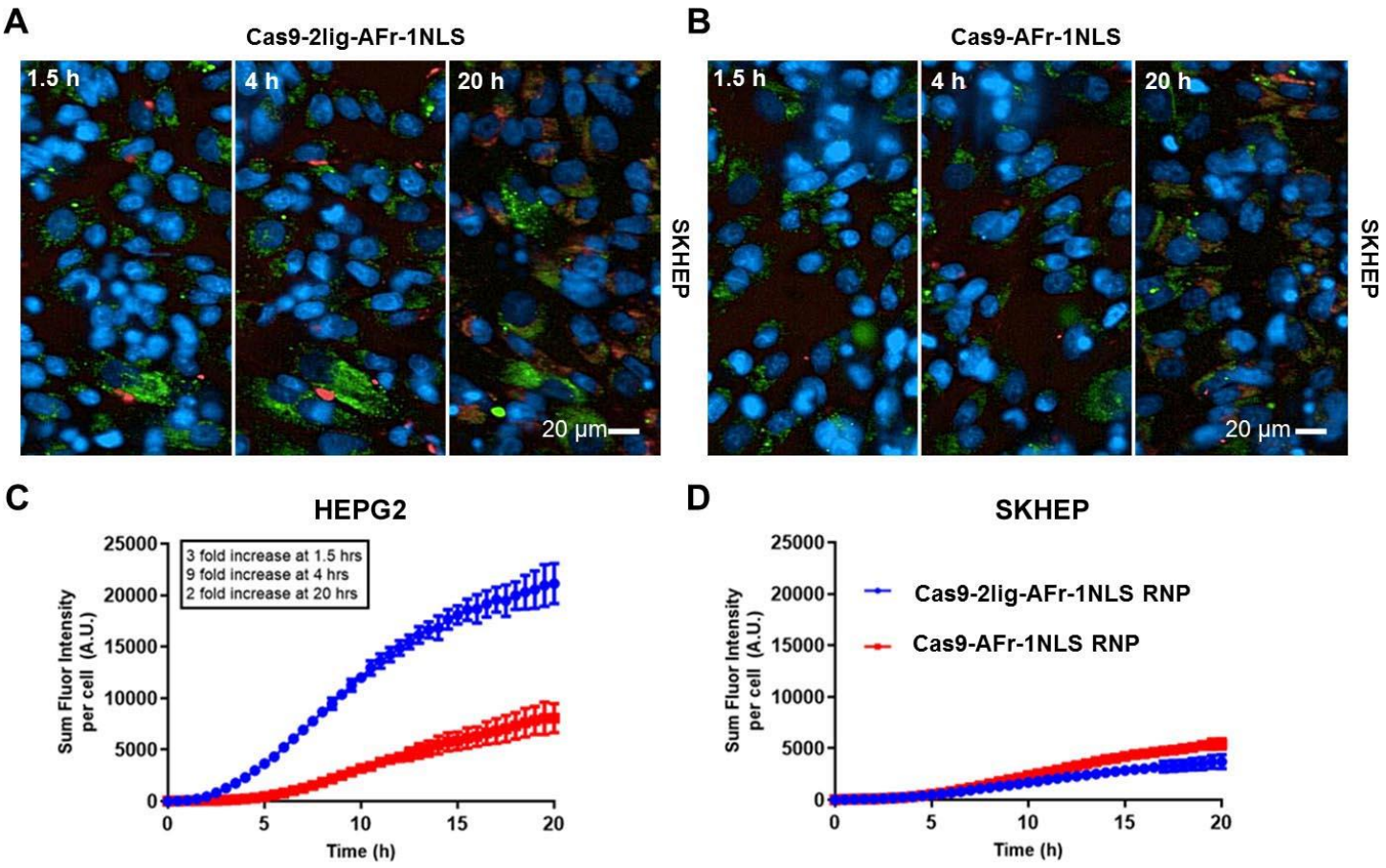


Figure S8. Internalization in SKHEP cells (ASGPr−) of Cas9-2lig-AFr-1NLS (A) and Cas9-AFr-1NLS (B) RNPs observed by live cell imaging at 1.5, 4, and 20 h. (C) Quantification of intracellular RNP accumulation in HEPG2 cells over 20 h (same as Fig. 3C). (D) Quantification of intracellular RNP accumulation in SKHEP cells over 20 h (same as Fig. 3D). Blue: Hoechst stain of cell nuclei; Green: Endolysosomal compartment stained using dextran488; Red: Intracellular Cas9 visualized via AF647 fluorescence. Fluorescence intensity was quantified using the sum of spots per cell (mean per well). Each data point (C,D) represents 3 technical replicate wells, with a minimum of 10,000 cells quantified per well. For (C) and (D), arithmetic means and standard deviations of the mean were calculated and plotted using GraphPad Prism© version 7.02. Corresponding RNPs made from sgRNA targeting EMX1.

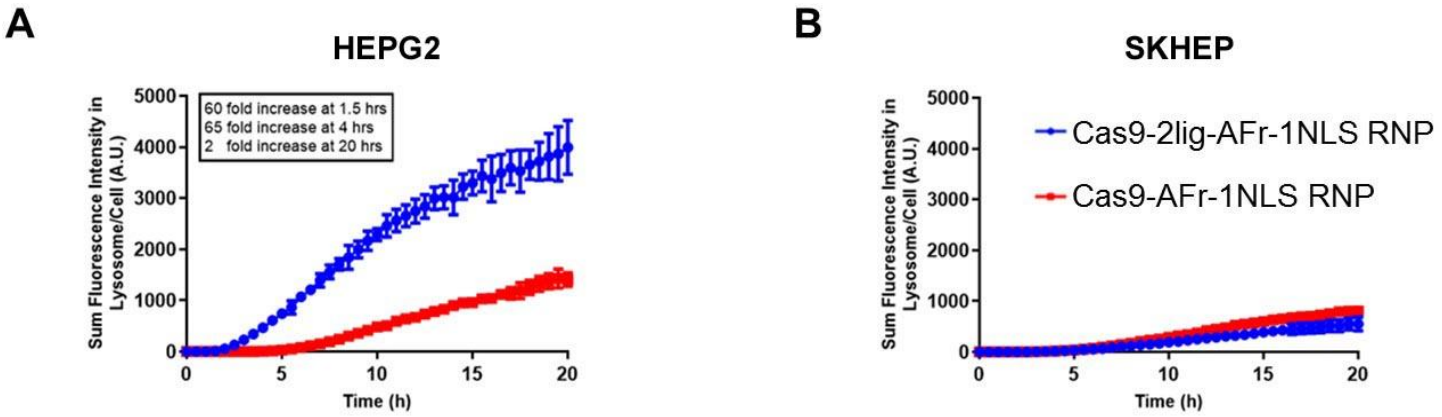
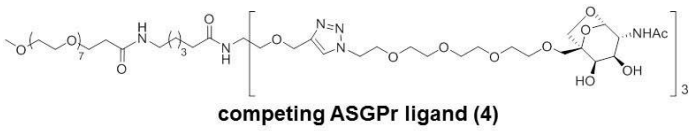


Figure S9. Quantification of endolysosomal accumulation of Cas9-2lig-AFr-1NLS and Cas9-AFr-1NLS RNPs as a function of time; (A) in HEPG2 cells and (B) in SKHEP cells. Fluorescence intensity was quantified using the sum of endolysosomal spots per cell (mean per well). Each data point represents 3 technical replicate wells, with a minimum of 10,000 cells quantified per well. Arithmetic means and standard deviations of the mean were calculated and plotted using GraphPad Prism© version 7.02. Corresponding RNPs made from sgRNA targeting EMX1.



competing ASGPr ligand (4)

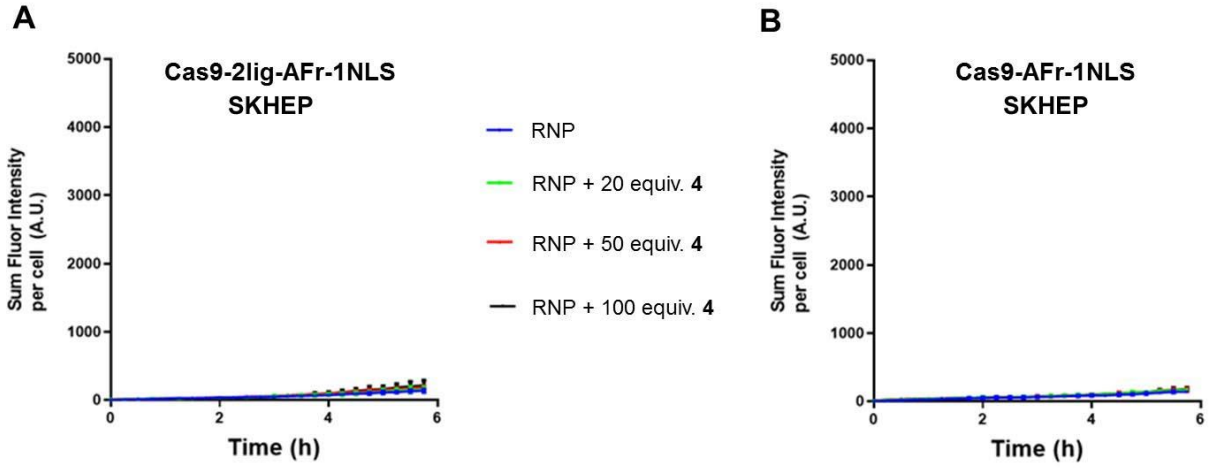


Figure S10. Ligand competition experiment in SKHEP cells (ASGPr−) with Cas9-2lig-AFr-1NLS RNP (A) and Cas9-AFr-1NLS RNP (B). Fluorescence intensity was quantified based on the sum of spots per cell (mean per well). Y axis identical to Fig. 4 for easier comparison. Each data point represents 3 technical replicate wells, with a minimum of 10,000 cells quantified per well. Arithmetic means and standard deviations of the mean were calculated and plotted using GraphPad Prism© version 7.02. Corresponding RNPs made from sgRNA targeting EMX1. "Equiv." represents "molar equivalents".

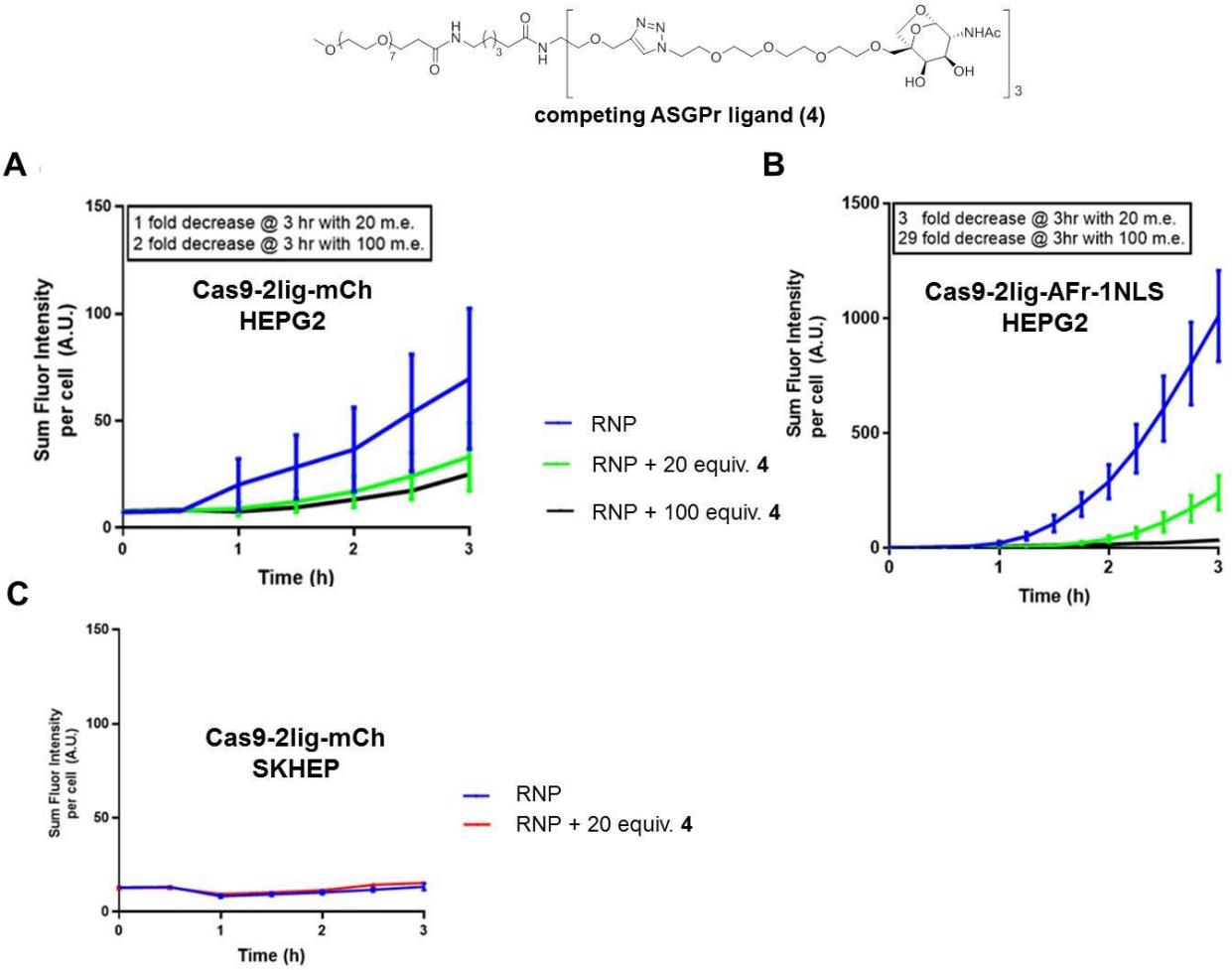


Figure S11. Ligand competition experiment in HEPG2 cells (ASGPr+) with Cas9-2lig-mCh RNP (A) and Cas9-2lig-AFr-1NLS RNP (B). (B) same as in Fig. 4A but plotted to 3 h. (C) Ligand competition experiment in SKHEP cells (ASGPr−) with Cas9-2lig-mCh RNP. Quantified fluorescence intensity represented as the sum of spots per cell (mean per well). "Equiv." and "m.e." represent "molar equivalents". Each data point represents 3 technical replicate wells, with a minimum of 10,000 cells quantified per well. Arithmetic means and standard deviations of the mean were calculated and plotted using GraphPad Prism© version 7.02. Corresponding RNPs made from sgRNA targeting EMX1.

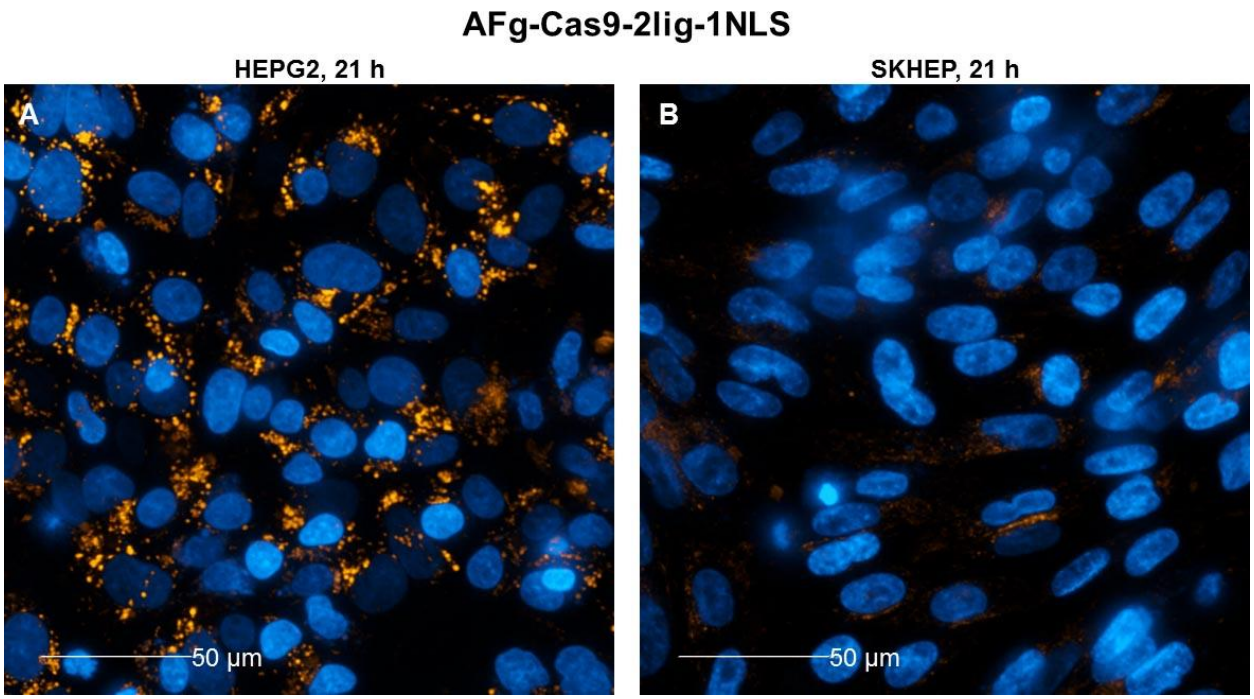


Figure S12. Comparison of Internalization at 21 h of AFg-Cas9-2lig-1NLS RNP in HEPG2 (A) and SKHEP (B) cells. Superior accumulation in endocytotic vesicles was observed for HEPG2 cells at this timepoint. Corresponding RNP made from sgRNA targeting EMX1. Blue: Hoechst stain of cell nuclei; Orange: Intracellular Cas9 visualized via AF532 fluorescence.

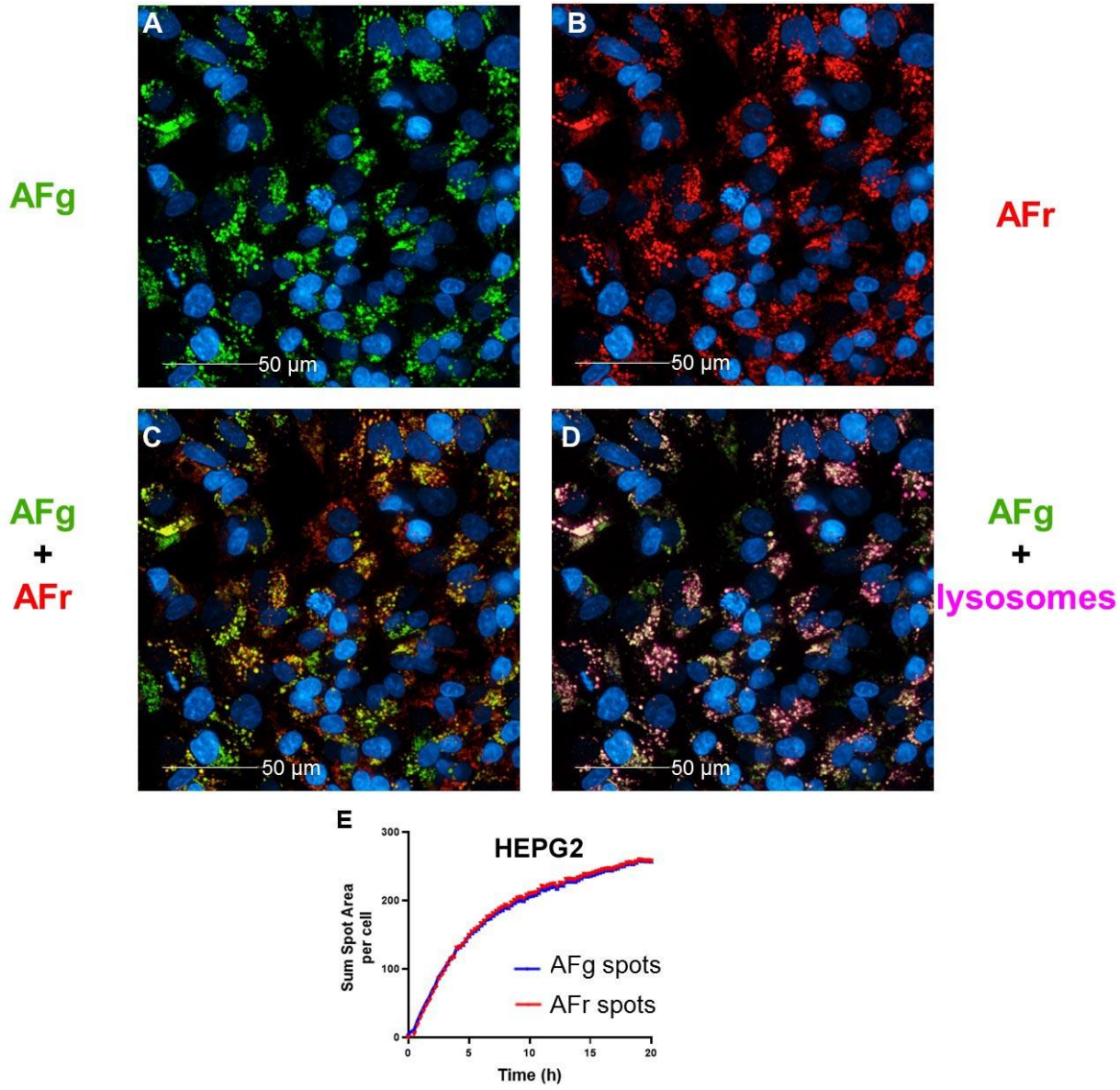
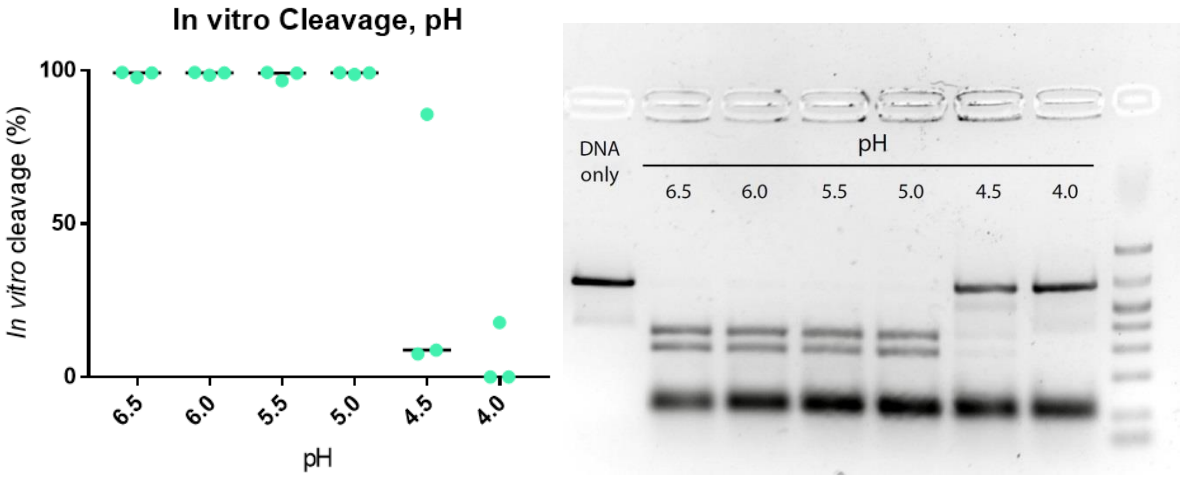


Figure S13. Live cell imaging of AFg-Cas9-2lig-AFr-1NLS RNP in HEPG2 cells (ASGPr+) at 21 h. Corresponding RNP made from sgRNA targeting EMX1. (A) Green AF532 (AFg) channel; (B) Red AF647 (AFr) channel; (C) AFg+AFr channels overlaid; (D) AFg+endolysosome (Dextran 488) channels overlaid. Clear colocalization in endocytotic punctae of linker and protein observed; evidence of endolysosomal accumulation. Blue: Hoechst stain of cell nuclei; Green: Intracellular Cas9 visualized via AFg fluorescence; Red: Intracellular Cas9 visualized via AFr fluorescence; Violet: Endolysosomes visualized via Dextran 488 fluorescence. (E) Sum spot area per cell (mean per well): overlapping of two curves indicates even uptake (of Cas9 and its conjugated ligand, as in (C)) in the cells. Each data point represents 3 technical replicate wells, with a minimum of 10,000 cells quantified per well.



pH	Percent substrate cleaved			Median
	rep 1	rep 2	rep 3	
6.5	97.7	99.2	99.3	99.2
6.0	98.4	99.2	99.4	99.2
5.5	96.6	99.1	99.4	99.1
5.0	98.7	99.2	99.3	99.2
4.5	85.7	8.8	7.6	8.8
4.0	17.8	0.0	0.0	0.0

Figure S14. Cas9 RNP *in vitro* cleavage of dsDNA after incubation at various pH. EMX1-targeting Cas9-mCh RNP was incubated at various pH for 1 h at 37 °C followed by incubation at pH 7.4 with substrate dsDNA for 1 h at 37 °C and analyzed by gel electrophoresis for DNA cleavage. (Left) Percent of DNA substrate cleaved under various pH conditions, black bars represent the median; (right) a representative agarose gel following the cleavage assay.

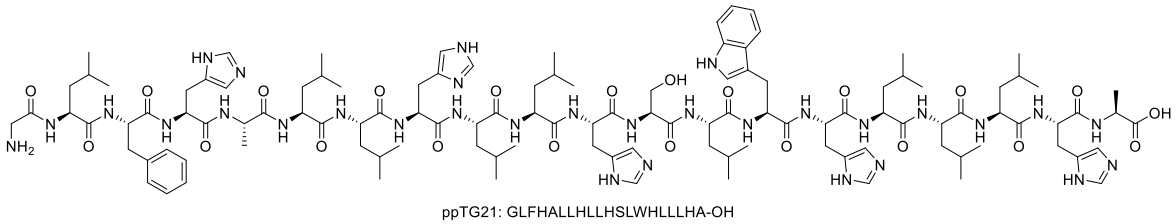


Figure S15. Structure of 20 amino acid peptide ppTG21. For original report see Rittner, K. et al. *Mol. Ther.* **2002**, *5*, 104-114.

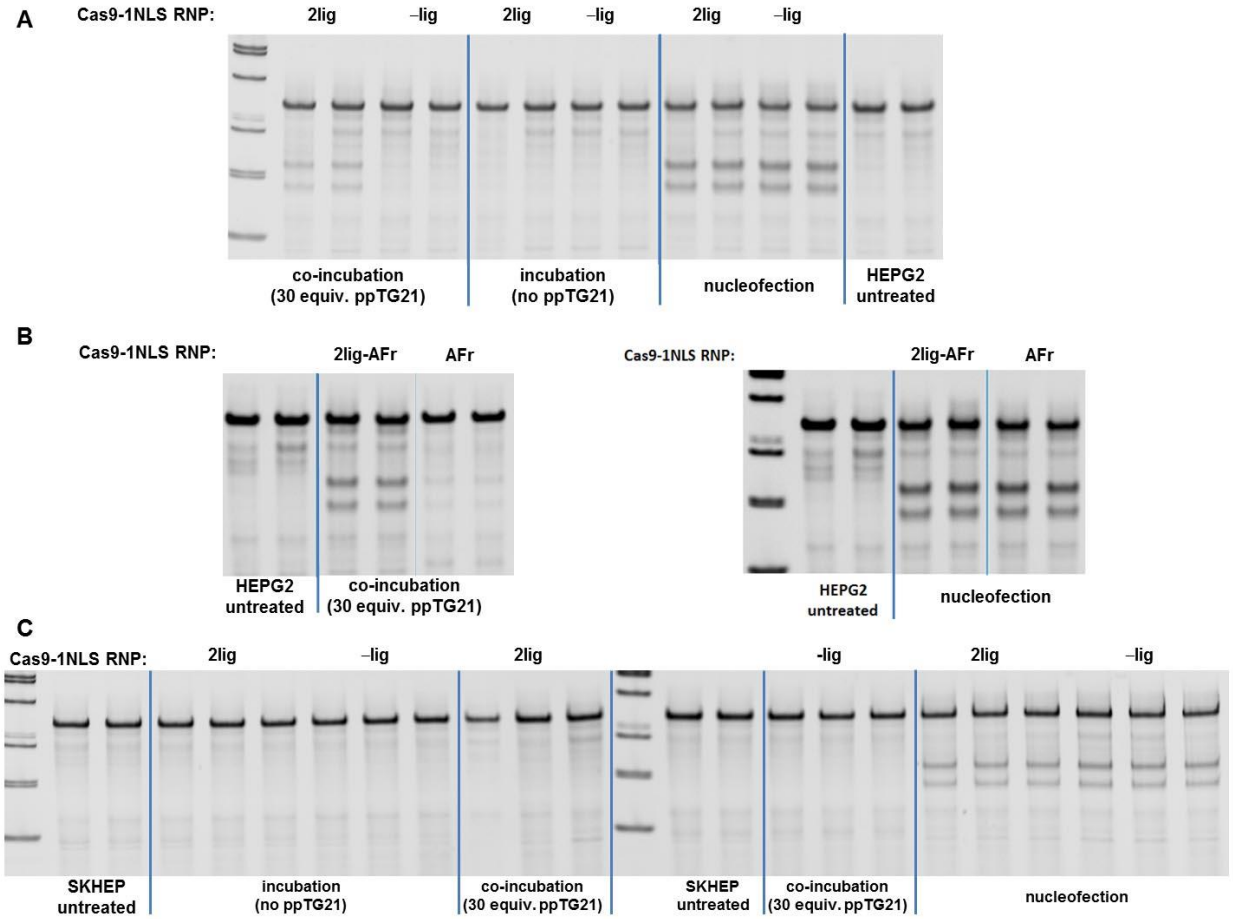


Figure S16. Representative gels showing qualitatively receptor-facilitated gene editing with Cas9-2lig-1NLS RNPs. (A) Cas9-2lig-1NLS and Cas9-1NLS RNP in HEPG2 cells. (B) Cas9-2lig-AFr-1NLS and Cas9-AFr-1NLS RNP in HEPG2 cells (right hand panel same as Fig. S6D). (C) Cas9-2lig-1NLS and Cas9-1NLS RNP in SKHEP cells. HEPG2 or SKHEP cells were treated with indicated RNP under specified conditions and incubated for 48 h at 37 °C. The genomic material was harvested and qualitatively evaluated by T7E1 assay; associated gel shown. See experimental part for more details. Corresponding RNPs made from sgRNA targeting EMX1.

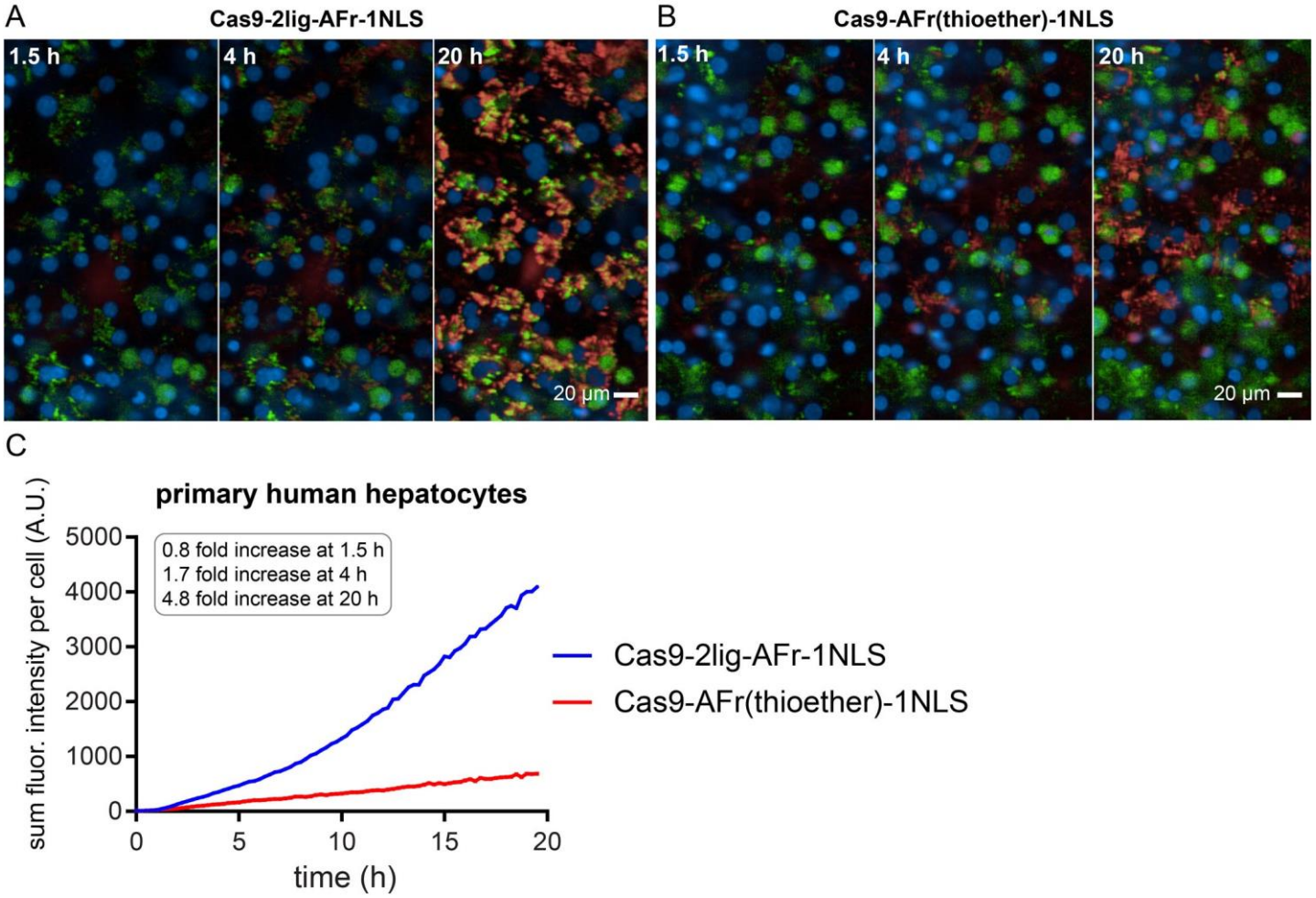


Figure S17. Internalization in primary human hepatocyte cells of Cas9-2lig-AFr-1NLS (A) and Cas9-AFr(thioether)-1NLS (B) RNPs observed by live cell imaging at 1.5, 4, and 20 h; in 20 h images contrast was adjusted down for clarity. Blue: Hoechst stain of cell nuclei; Green: Endolysosomal compartments stained using dextran488; Red: Intracellular Cas9 visualized via AF647 fluorescence. (C) Quantification of intracellular RNP accumulation in endolysosomes of human hepatocyte cells over 20 h. Fluorescence intensity was quantified using the sum of endolysosomal spots per cell (mean per well). Each data point represents a single replicate well, with a minimum of 5,000 cells quantified per well. RNP samples were made using sgRNA targeting EMX1, and were mixed with 30 molar equivalents of ppTG21 before co-incubation with the cells.

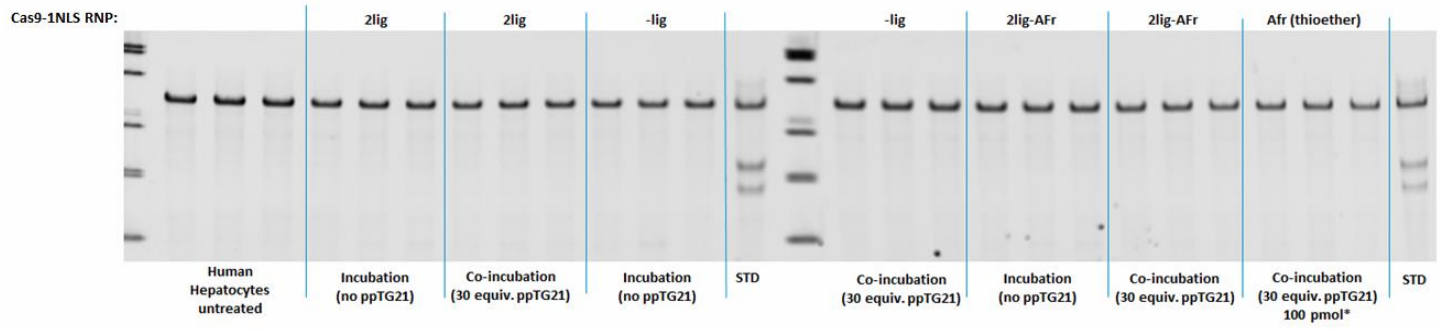


Figure S18. Representative gel assaying qualitatively for gene editing with Cas9-2lig-1NLS, Cas9-1NLS, Cas9-2lig-AFr-1NLS, and Cas9-AFr (thioether)-1NLS RNPs in primary human hepatocytes following co-incubation initiated while cells were plated. RNP samples were made using sgRNA targeting EMX1. Positive control "STD" lanes show editing associated with genetic material harvested from co-incubation experiment with Cas9-2lig-1NLS with 30 equiv. ppTG21 in HEPG2. No editing of the EMX1 locus was observed under experimental conditions. Primary human hepatocyte cells were treated with indicated RNP under specified conditions and incubated for 48 h at 37°C.

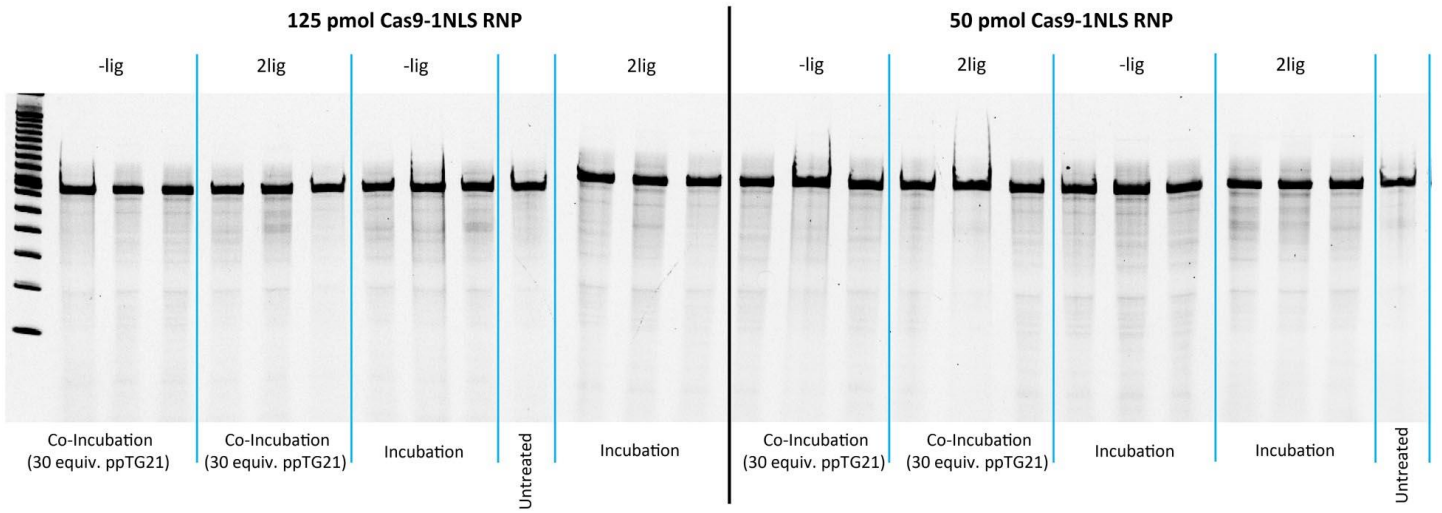


Figure S19. Gels assaying qualitatively for gene editing with Cas9-2lig-1NLS and Cas9-1NLS RNP in primary human hepatocytes following co-incubation initiated while cells were in suspension. RNP samples were made using sgRNA targeting EMX1. No editing of the EMX1 locus was observed under experimental conditions. Successful editing would result in a banding pattern corresponding to the "STD" positive control lanes of Figure S18. Primary human hepatocyte cells were treated with indicated RNP under specified conditions and incubated for 48 h at 37°C.

9. Custom Python script for processing NGS data

```
#!/usr/bin/env python

#This script takes as input a merged paired-end read file containing heterogenity tags
and unique molecular identifiers at the 5' and or 3' ends.
#It writes out a fastq file with only one read per UMI, the one with the highest average
quality score, plus seperate fastq files for the
#sequences and UMI tags.

#The script requires five arguments.
#The first argument is the input file in fastq format. It must be unzipped.
#The second argument is a tag to add to all the output file names.
#The third argument is the length of the UMI tag at the 5' end. If there is also a
heterogenity tag, use the maximun length of the heterogeneity tag plus the length of the
UMI tag.
#The fourth argument is the length at the 3' end. If there is also a heterogenity tag,
use the maximum lenght of the heterogenity tag plus the lenght of the UMI tag.
#The fifth argument is the minimun number of reads a UMI must have to be included in the
output file. Typically we set this value to 1.


#This segment tests whether the right number of input arguments were used.
import sys
if len(sys.argv) != 6:
    print "This script takes five arguments, an input file, an output filename the length
of the UMI at the 5' end and the length of the UMI at the 3' end and the minimun number
of times a UMI must appear."
    sys.exit()

#The following code defines a function to calculate the average quality score from an
illunima quality score string.
def average_qual_score(i):
    qual_sum = 0
    len_seq = 0
    for k in list(i):
        if k == '!':
            qual_sum += 0
        elif k == '"':
            qual_sum += 1
        elif k == '#':
            qual_sum += 2
        elif k == '$':
            qual_sum += 3
        elif k == '%':
            qual_sum += 4
        elif k == '&':
            qual_sum += 5
        elif k == "'":
            qual_sum += 6
        elif k == '(':
            qual_sum += 7
        elif k == ')':
            qual_sum += 8
        elif k =='*':
            qual_sum += 9
        elif k == '+':
            qual_sum += 10
        elif k == ',':
            qual_sum += 11
```

```
elif k == '-':
    qual_sum += 12
elif k == '.':
    qual_sum += 13
elif k == '/':
    qual_sum += 14
elif k == '0':
    qual_sum += 15
elif k == '1':
    qual_sum += 16
elif k == '2':
    qual_sum += 17
elif k == '3':
    qual_sum += 18
elif k == '4':
    qual_sum += 19
elif k == '5':
    qual_sum += 20
elif k == '6':
    qual_sum += 21
elif k == '7':
    qual_sum += 22
elif k == '8':
    qual_sum += 23
elif k == '9':
    qual_sum += 24
elif k == ':':
    qual_sum += 25
elif k == ';':
    qual_sum += 26
elif k == '<':
    qual_sum += 27
elif k == '=':
    qual_sum += 28
elif k == '>':
    qual_sum += 29
elif k == '?':
    qual_sum += 30
elif k == '@':
    qual_sum += 31
elif k == 'A':
    qual_sum += 32
elif k == 'B':
    qual_sum += 33
elif k == 'C':
    qual_sum += 34
elif k == 'D':
    qual_sum += 35
elif k == 'E':
    qual_sum += 36
elif k == 'F':
    qual_sum += 37
elif k == 'G':
    qual_sum += 38
elif k == 'H':
    qual_sum += 39
elif k == 'I':
    qual_sum += 40
else:
    print "Not reading quality score correctly: "+str(k)
    sys.exit()
```

```
        len_seq += 1
    return(float(qual_sum)/float(len_seq))


#This segment reads the input fastq file and splits the sequences and UMI tags into two
fastq files.
f=open(sys.argv[1], 'r')
UMI_list = []
UMI_count = {}
UMI_read_names = {}
UMI_count_list = []
current_read_name = ''
sequences = {}
quality_scores = {}
qual_threads = {}
total_reads = 0
total_UMIs = 0
UMIs_greater_then_cutoff = 0
o=open(str(sys.argv[2])+'UMI.fastq','w')
p=open(str(sys.argv[2])+'read.fastq','w')
q=open(str(sys.argv[2])+'UMI.fasta','w')
line = f.readline()
line_count = 1
while line !='':
    if line_count % 4 == 1:
        o.write(line)
        p.write(line)
        q.write('>'+line)
        current_read_name = line.replace('\n','')
    elif line_count % 4 == 2:
        o.write(line[:int(sys.argv[3])]+line[-1*(int(sys.argv[4])+1):])
        p.write(line[int(sys.argv[3]):-1*(int(sys.argv[4])+1)]+'\n')
        q.write(line[:int(sys.argv[3])]+line[-1*(int(sys.argv[4])+1):])
        UMI = line[:int(sys.argv[3])]+line[-1*(int(sys.argv[4])+1):].replace('\n','')
        sequences[current_read_name] = line[int(sys.argv[3]):-1*(int(sys.argv[4])+1)]
        if UMI not in UMI_list:
            UMI_list.append(UMI)
            UMI_count[UMI] = 1
            total_UMIs += 1
            UMI_read_names[UMI]=[current_read_name]
        else:
            UMI_count[UMI] += 1
            UMI_read_names[UMI].append(current_read_name)
        total_reads += 1
    elif line_count % 4 == 3:
        o.write(line)
        p.write(line)
    else:
        o.write(line[:int(sys.argv[3])]+line[-1*(int(sys.argv[4])+1):])
        p.write(line[int(sys.argv[3]):-1*(int(sys.argv[4])+1)]+'\n')
        quality_scores[current_read_name]=average_qual_score(line[int(sys.argv[3]):-
1*(int(sys.argv[4])+1)])
        qual_threads[current_read_name] = line[int(sys.argv[3]):-1*(int(sys.argv[4])+1)]
    line = f.readline()
    line_count += 1
f.close()
o.close()
p.close()
q.close()

#This segment writes out some statistics about UMI counts and distributions
```

```
r=open(str(sys.argv[2])+'_UMI_count_dist.txt','w')
for i in UMI_list:
    if UMI_count[i] >= int(sys.argv[5]):
        UMIs_greater_then_cutoff += 1
    UMI_count_list.append(int(UMI_count[i]))
UMI_count_list.sort(reverse=True)
r.write('total reads:\t'+str(total_reads)+'\n')
r.write('total UMIs:\t'+str(total_UMIs)+'\n')
r.write('UMIs with '+str(sys.argv[5])+' or more
occurences:\t'+str(UMIs_greater_then_cutoff)+'\n')

UMI_dist = {}
index = []
for i in range(1,UMI_count_list[0]+1):
    UMI_dist[i] = 0
    index.append(i)
for i in UMI_count_list:
    UMI_dist[i] += 1

#this segment writes out statistics about UMI counts and distibutions.

reads_all=0
UMIs_all=0
reads_cutoff=0
UMIs_cutoff=0
for i in index:
    reads_all += i*UMI_dist[i]
    UMIs_all += UMI_dist[i]
    if i >= int(sys.argv[5]):
        reads_cutoff += i*UMI_dist[i]
        UMIs_cutoff += UMI_dist[i]
r.write('average reads per UMI:\t'+str(reads_all/UMIs_all)+'\n')
r.write('average reads per UMI when cutoff of '+str(sys.argv[5])+' is
applied:\t'+str(reads_cutoff/UMIs_cutoff)+'\n')
for i in index:
    r.write(str(i)+'\t'+str(UMI_dist[i])+'\n')
r.close()

#This segment writes out a fastq with just the highest quality read for each UMI group.
s=open(str(sys.argv[2])+'_best_read.fastq','w')
sequences_output = 0
for i in UMI_list:
    if UMI_count[i] >= int(sys.argv[5]):
        best_qual = 0
        best_read = ''
        for k in UMI_read_names[i]:
            if quality_scores[k] > best_qual:
                best_qual = quality_scores[k]
                best_read = k
        s.write(str(best_read)+'\n')
        s.write(str(sequences[best_read])+'\n')
        s.write('+\n')
        s.write(str(qual_threads[best_read])+'\n')
        sequences_output += 1
s.close()

print "Wrote out "+str(sequences_output)+" reads from unique templates."
print "There were "+str(UMIs_greater_then_cutoff)+' UMIs with '+str(sys.argv[5])+'or more
occurances.'
```