Electronic Supplementary Material (ESI) for ChemComm.

Genetic Encoding of A Nonhydrolyzable Phosphotyrosine Analog in Mammalian Cells

I. General material and method

All commercial chemicals are of reagent grade or higher. All solutions were prepared in deionized water that was further treated by Barnstead Nanopure® ultrapure water purification system (Thermo Fisher Scientific). Preparation of LB medium and YPD medium followed reported recipe. Yeast selection media contained yeast nitrogen base without amino acids (Difco™, 6.7 g/L), glucose (10 g/L), and appropriate drop out (DO) supplements (Clonetech). The pH values of all media were adjusted to 7.0. Agar plates were prepared by the addition of Difco agar (15 g/L) to the liquid medium. Antibiotics were added where appropriate to following final concentrations: kanamycin (50 mg/L), ampicillin (100 mg/L). Solutions of antibiotics were filtered through 0.22 μm sterile membrane filters. CMF was synthesized by following previously published procedure with modifications.[1] (Supplementary Information). EGFR and EGF were purchase from ThermoFisher Scientific. Heparin Sepharose 6 Fast Flow resin was purchase from Cytiva.

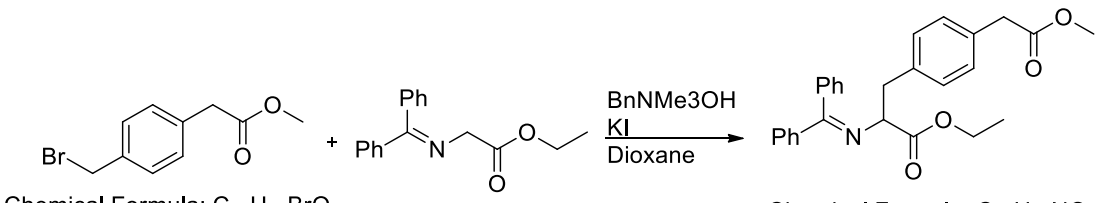
II. Experimental procedures

Chemical synthesis

Chemical Formula: $C_9H_9BrO_2$
Exact Mass: 227.98
1

Chemical Formula: $C_{10}H_{11}BrO_2$
Exact Mass: 241.99
2

Methyl 2-(4-bromomethyl phenyl)-acetate (2). To a solution of 2-(4-bromomethyl phenyl)-acetic acid **1**(10g, 43.7mmol) in anhydrous MeOH(100mL) was added trimethyl silyl chloride (1.1 mL, 8.74mmol) dropwise, and the mixture was stirred at room temperature for 3 h. After the TLC showed the reaction was complete, the solvent was evaporated under vacuum and the residue was purified by flash column chromatography (15/1, Hexane/Ethyl Acetate) to yield compound **2** (10.2g, 96.5%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.37 (d, J = 8.0 Hz, 2H, ArH), 7.28 (d, J = 8.0 Hz, 2H, ArH), 4.50 (s, 2H, $ArCH_2Br$), 3.72 (s, 3H, $COOCH_3$), 3.65 (s, 2H, $ArCH_2COOCH_3$).



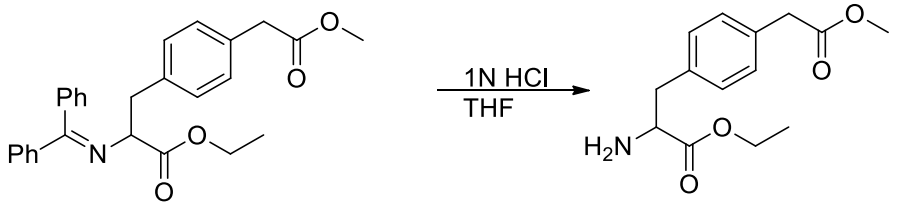
Chemical Formula: $C_{10}H_{11}BrO_2$
Exact Mass: 241.99

2

Chemical Formula: $C_{27}H_{27}NO_4$
Exact Mass: 429.19

3

Ethyl 2-((diphenylmethylene)amino)-3-(4-(2-methoxy-2-oxoethyl)phenyl)propanoate (3). A solution of compound **2** (5g, 20.7 mmol) in dioxane (100 mL) cooled to 5 °C was added to N-(diphenyl methylene)-glycine ethyl ester (6.7g, 25.1mmol) and potassium iodide (343mg, 1.5mmol). Benzyl trimethyl ammonium hydroxide aq. 40% solution (11.5mL, 1.2equivalent) was added. The reaction mixture was stirred at 10 °C for 10 min, diluted with ethyl acetate (100mL) and brine (150mL), then extracted with ethyl acetate (100mL x 2). The combined organic layer was washed with brine (50mL x 2) and evaporated under vacuum to obtain a pale yellow syrup as crude compound **3** (7.2 g).



Chemical Formula: $C_{27}H_{27}NO_4$
Exact Mass: 429.19

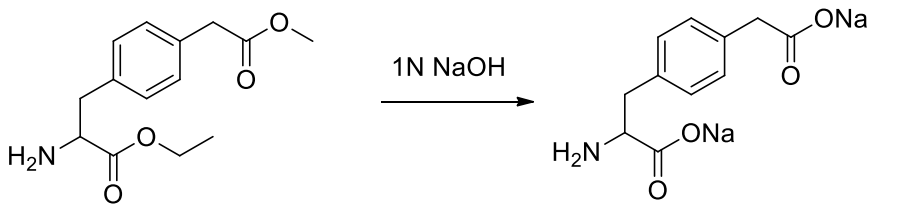
3

Chemical Formula: $C_{14}H_{19}NO_4$
Exact Mass: 265.13

4

Ethyl 2-amino-3-(4-(2-methoxy-2-oxoethyl)phenyl)propanoate (4). Compound **3** (9g, 20.9mmol) was dissolved in THF (70mL) . The solution was cooled in an ice bath, HCl (1N, 70mL) was added and the reaction mixture was stirred at room temperature for 1 h. The mixture was adjusted to pH 7.5 with solid $NaHCO_3$ and extracted with ethyl acetate (100mL x 3). The combined organic layer was dried over Na_2SO_4, evaporated under vacuum and purified by flash column chromatography (90:1 to 60:1 DCM/MeOH) to give compound **4** (3.98g, 72%) as a pale-yellow solid. ^1H NMR(400 MHz, $CDCl_3$) δ 7.23 (d, J = 8.0 Hz, 2H, ArH), 7.17 (d, J = 8.0 Hz, 2H, ArH), 4.16 (q, J = 7.2 Hz, 2H, $COOCH_2CH_3$), 3.71 (dd, J_1 = 5.6 Hz, J_2 = 8.0 Hz,1H, CH_2CHNH_2), 3.70 (s, 3H, $COOCH_3$), 3.62 (s, 2H,

$ArCH_2COOCH_3$), 3.06 (dd, J_1 = 5.6 Hz, J_2 = 13.2 Hz, 1H, CH_2CHNH_2), 2.85 (dd, J_1 = 8.0Hz, J_2 = 13.2Hz, 1H, CH_2CHNH_2), 1.47 (bs, 2H, NH_2), 1.27 (t, J = 7.2 Hz, 3H, $COOCH_2CH_3$).



Chemical Formula: $C_{14}H_{19}NO_4$
Exact Mass: 265.13

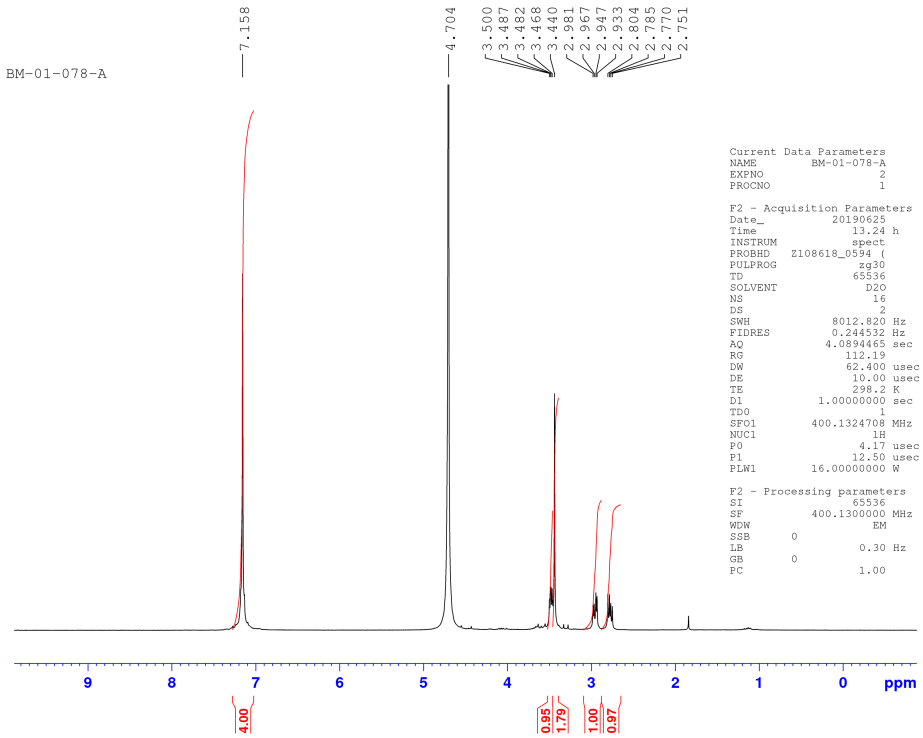
4

Chemical Formula: $C_{11}H_{11}NNa_2O_4$
Exact Mass: 267.05

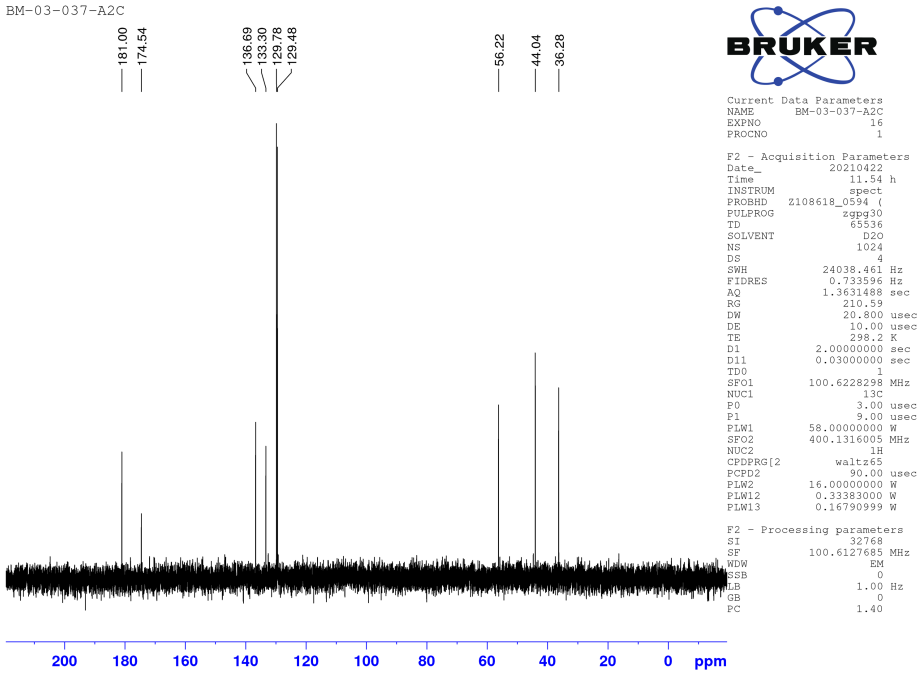
5

***p*-carboxymethyl D,L-phenylalanine sodium salt (5).** To a solution of compound 4 (3.98g, 15 mmol) in THF was slowly added 1 N NaOH (2 equivalent). The reaction mixture was stirred for 12 h, the solvents were evaporated under vacuum. The residue was re-dissolved in water and lyophilized to yield compound 5 (3.5g, 87.4%) as white solid. ^{1}H NMR(400 MHz, D_2O; Supplementary Fig. 14) δ 7.16 (m, 4H, ArH), 3.49 (dd, J_1 = 5.6 Hz, J_2 = 7.6 Hz, 1H, CH_2CHNH_2), 3.44 (s, 2H, CH_2COOCH_3), 2.97 (dd, J_1 = 5.6Hz, J_2 = 13.6 Hz, 1H, CH_2CHNH_2), 2.79 (dd, J_1 = 7.6 Hz, J_2 = 13.6 Hz, 1H, CH_2CHNH_2). ^{13}C NMR (400 MHz; Supplementary Fig. 14) δ 181.00, 174.54 (COO),136.69, 133.30, (ArC), 129.78, 129.48 (ArCH), 56.22 (CH_2CHNH_2), 44.04 ($ArCH_2COOCH_3$), 36.28 (CH_2CHNH_2).

^{1}H NMR



^{13}C NMR



Cell culture and transfection. HEK293T cells and 293T-ΔSTAT1 were maintained in DMEM media supplemented with 10% FBS at 37 °C in a humidified atmosphere of 5% CO_2 (v/v). Transfection was conducted at 70-80% cell confluency using Lipofectamine 2000 or Lipofectamine 3000 (Life Technologies) according to the manufacturer's protocol.

Strains and plasmids.

E. coli GeneHogs® (Thermo Fisher Scientific) was used for routine cloning and plasmid propagation unless otherwise mentioned. *E. coli* NEB® Stable (New England BioLabs) was used for cloning and propagation of unstable plasmids. Plasmid construction was performed using T4 DNA ligase or Gibson Assembly method.[2] *S. cerevisiae* strain MaV203 (MATα; leu2-3,112; trp1-109; his3Δ200; ade2-101; cyh2R; cyh1R; gal4Δ; gal80 Δ; GAL1::lacZ; HIS3UASGAL1::HIS3@LYS2; SPAL10UASGAL1::URA3) (Thermo Fisher Scientific) was used for yeast library construction and selection. Standard protocols were followed for the purification and analysis of plasmid DNA.[3] PCR amplifications were carried out using KOD Hot Start DNA polymerase (Millipore Sigma) by following the manufacturer's protocol. Other molecular cloning reagents were purchased from New England BioLabs. Primer synthesis service was provided by Millipore Sigma, and DNA sequencing service was provided by Eurofins MWG Operon. Primers used in this study are listed in Table S1. Plasmids used and constructed in this study are listed in Table 2S. All plasmids are available from corresponding authors upon reasonable request.

293T-ΔSTAT1 cell. A STAT1-knockout cell line was generated from 293T using CRISPR/Cas9 method. The gRNA sequence[4] (5'-TCATGACCTCCTGTCACAGC-3') was inserted into BbsI-linearized plasmid pSpCas9(BB)-2A-GFP (PX458, Addgene 48138). The resulting plasmid PX458-STAT1 gRNA was transfected into 293T cells. 24 h post transfection, cells were detached and filtered through a 35 μm cell strainer and single cells with GFP expression were sorted into a 96-well plate using BD FACSAria II. Colonies developed from single cells were expanded and maintained.

A pair of primers flanking gRNA binding site were used to amplify PCR products for DNA sequencing in order to identify chromosomal editing events. The selected 293T STAT1-knockout cell line had a 365-bp deletion on both chromosomes, which covered 99 bp of exon and 266 bp of intron. The phenotype of the cell line was confirmed by western blot.

Plasmid pCMFRS. This 7.5-kb pcDNA3.1-derived plasmid contains one copy of the amber suppressor *B. subtilis* tyrosyl-tRNA under the control of a human U6 promoter and a CMFRS-encoding gene behind a CMV promoter. Fragment encoding the CMFRS-1 was amplified from the isolated pEcTyrRS-CMFRS1 and inserted into a 6.2-kb vector that was obtained after treating pcDNA3.1-AzFRS with HindIII and ApaI.

Plasmids pSTAT1-EGFP and derivatives. The 9.2-kb pSTAT1-EGFP plasmid is derived from pcDNA3.1 by replacing the AzFRS-encoding gene in pcDNA3.1-AzFRS with DNA fragment of human STAT1 and EGFP fusion protein. STAT1 and EGFP genes were connected by a short DNA sequence that encodes a GSGSGSAAA linker. To make pSTAT1 Y701D-EGFP, pSTAT1 Y701E-EGFP, or pSTAT1 Y701TAG-EGFP, GAT, GAG, or TAG was introduced at the Y701 position of the STAT1 gene in pSTAT1-EGFP by mutagenic PCR.

Plasmids pSTAT1 and pSTAT1 Y701TAG. The 8.4-kb pSTAT1 and pSTAT1-Y701TAG plasmids were constructed by inserting the PCR products of STAT1 and STAT1 Y701TAG into a 6.2-kb vector that was derived from HindIII and ApaI treatment of pSTAT1-EGFP, respectively.

Plasmids pSTAT1 (132-712) and derivatives. The 8.0-kb pSTAT1 (132-712), pSTAT1 (132-712) Y701D, pSTAT1 (132-712) Y701E, and pSTAT1 (132-712) Y701TAG plasmids are derived from pSTAT1-EGFP. DNA fragments encoding residues 132-712 of STAT1, STAT1 701D, STAT1 701E, and STAT1 701TAG were amplified, purified, and assembled into a 6.2-kb DNA fragment from the digestion of pSTAT1-EGFP by HindIII and XhoI, respectively.

Plasmid pSpCas9(BB)-2A-GFP-STAT1 gRNA. The 9.3-kb pSpCas9(BB)-2A-GFP-STAT1 gRNA plasmid is derived from pSpCas9(BB)-2A-GFP (PX458, Addgene). Primers STAT1 gRNA-F and STAT1 gRNA-R were annealed and phosphorylated following standard protocol.[5] The mixture was ligated into the BbsI-linearized plasmid pSpCas9(BB)-2A-GFP. *E. coli* NEB® Stable was used for efficient transformation and stable propagation.

Plasmids pCMFRS with multiple copies of Bs tRNA. The 8.2-kb plasmid pCMFRS-3xtRNA contains three copies of the *B. subtilis* amber suppressor tyrosyl-tRNA under the control of a human U6 promoter and a human H1 promoter. Promoter H1 was amplified from 293T genome and assembled with Bs tRNA as a H1-Bs tRNA gene cassette by overlapping PCR. The U6-Bs tRNA cassette was amplified from pCMFRS. The two cassettes were inserted into BstZ17I-digested pCMFRS to obtain pCMFRS-3xtRNA. The 8.9-kb pCMFRS-5xtRNA was constructed by ligating the H1-Bs tRNA-U6-Bs tRNA fragment, which was released from pCMFRS-3xtRNA by NheI/XbaI, into the XbaI-treated pCMFRS 3xtRNA. The same cloning strategy for pCMFRS-5xtRNA was applied to the construction of plasmids pCMFRS-7xtRNA and pCMFRS-9xtRNA using pCMFRS-5xtRNA and pCMFRS-7xtRNA as the vector, respectively. *E. coli* NEB® Stable was used as the cloning host.

Library construction and selection.

The EcTyrRS library was constructed previously in MaV203/pGADGAL4 by randomizing four residues (Y37, L71, W129, and D182) using the NNK codon (N = A, C, T, or G, K = T or G).[6-7] A library size of 1×10^7 diversity was obtained and maintained in SD-Leu-Trp media. Positive selection was conducted by culturing cells on the SD-Leu-Trp-Ura plates with 1 mM CMF to identify EcTyrRS variants that can efficiently aminoacylate the amber suppressor tRNA with CMF. In the negative selection, cells were plated on SD-Leu-Trp media plates containing 5-fluoroorotic acid (5-FOA) at 0.1%. EcTyrRS variants that can charge the amber suppressor *Ec*tRNA$_{CUA}$ with any one of the 20 natural amino acids led to GAL4-activated expression of URA3, which converts 5-FOA into a cytotoxic

product and leads to cell death. Single colonies picked from the selection plates were resuspended in 100 μL of SD-Leu-Trp-Ura media. A 2 μL cell resuspension was applied to SD-Leu-Trp-Ura, and SD-Leu-Trp-Ura plates with 1 mM CMF to further verify the phenotypes. Plasmids isolated from potential candidates were propagated in *E. coli* GeneHogs for DNA sequencing.

Fluorescence analysis of yeast culture.

MaV203 transformed with plasmids pyeGFP-N149TAG and pEcTyrRS-mutant was cultured in SD-Leu-Trp media without or with CMF (1 mM or 5 mM) at 30 °C for 12 h. Cells were then collected, washed, and resuspended in an equal volume of PBS (pH 7.4). Cell density was determined by measuring OD_{600nm}. The fluorescence of yeGFP was monitored at Ex = 485 nm and Em = 528 nm. Values of fluorescence intensity were normalized to the OD measurement. Reported data are the average of three measurements with standard deviations. A Biotek Synergy HTX plate reader was used in absorbance, and fluorescence measurements.

Confocal microscopy.

293T cells (1×10^5) seeded in a single well of a 24-well plate were grown for 16-24 h, then transfected with plasmids pCMFRS (0.8 μg) and pEGFP (0.8 μg) in 2 μL of Lipofectamine 2000. Transfected cells were cultured for an additional 24 h in 0.5 mL of media without or with CMF (1 mM or 5 mM). Cells were then washed with 0.5 mL of warm DMEM base medium, fixed with 4% paraformaldehyde (w/v) for 15 min. Following removal of the fixation reagent by washing with 3 x 0.5 mL DPBS, cells were visualized by a Nikon A1R-Ti2 confocal microscope.

Flow cytometry analysis.

For the quantification of amber suppression efficiency by CMFRS, transfected 293T cells from a single well of a 24-well plate were detached with 0.3 mL of 0.05% Trypsin/EDTA, washed with 0.5 mL DPBS, and collected by centrifugation at 300 x *g* for 5 min. Collected cells were fixed in 0.5 mL 4% paraformaldehyde (w/v) for 15 min at room temperature. Following removal of the fixation reagent by

centrifugation, cells were washed twice and resuspended in 0.5 mL DPBS and kept on ice until analysis. Fluorescence of cells were measured using a Beckman Coulter CytoFLEX flow cytometer. A total of 30,000 cells were analyzed for each sample. Data were analyzed using FlowJo. Reported data are the average measurement of three samples with standard deviations.

Mass spectrometry analysis.

Proteins EGFP-40CMFand c-STAT1-701CMF were overexpressed in 293T cells transfected with pCMFRS together with pEGFP or pSTAT1 (132-712) Y701TAG. Cells were lysed by sonication in protein binding buffer (25 mM Tris-HCl, 200 mM NaCl, 10 mM imidazole, 1 x protease inhibitor cocktail, pH 7.5). Following centrifugation at 21,000 x *g* for 30 min at 4 °C, tagged protein was purified using Ni Sepharose 6 Fast Flow resin.

For mass analysis of the full-length EGFP-40CMF protein, the purified fraction was first dialyzed into phosphate buffer (100 mM, pH 7.4) to remove imidazole, then diluted into a buffer with ACN (10%) and formic acid (0.1%). Obtained sample was analyzed on an RSLCnano system (ThermoFisher Scientific) coupled to a Q-Exactive HF mass spectrometer (ThermoFisher Scientific). The protein mass was calculated using the Protein Deconvolution software (ThermoFisher Scientific).

For tandem mass spectrometry analysis, purified protein was further resolved from impurity by SDS-PAGE. Bands corresponding to target proteins were excised, washed, and treated with trypsin overnight at 37 °C. Tryptic peptides were extracted from the gel pieces, dried down, and re-dissolved in 25 μL of an aqueous solution of acetonitrile (2.5%) and formic acid (0.1%). Each digest was run by an RSLCnano system using a 1-h gradient on a 0.075 mm x 250 mm C18 column feeding into a Q-Exactive HF mass spectrometer. All MS/MS samples were analyzed using Mascot. Mascot was set up to search a database that was customized with provided protein sequence. Mascot search has a fragment ion mass tolerance of 0.060 Da and a parent ion tolerance of 10.0 ppm. Deamidation of asparagine and glutamine, oxidation of methionine, and CMF incorporation were specified in Mascot

as variable modifications. Scaffold was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 99.0% probability by the Peptide Prophet algorithm[8] with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm.[9]

Western blot.

Harvested 293T cells were lysed with RIPA buffer (Thermo Scientific) containing Halt™ protease and phosphatase inhibitor cocktail (Thermo Scientific). Following centrifugation, the supernatant was saved as cell extracts. Protein concentrations were then determined using the Quick Start™ Bradford protein assay kit (Bio-Rad Laboratories). Proteins were separated by SDS-PAGE, then transferred onto a nitrocellulose membrane using mini trans-blot cell (Bio-Rad Laboratories). The membrane was blocked with 3% BSA at 4 °C overnight, incubated with primary antibodies at 4 °C for 8 hours, and incubated with secondary antibodies at room temperature for 1 h. Blots were developed using Opti-4CN detection kit (Bio-Rad Laboratories) and imaged using Gel Doc XR+ system (BioRad Laboratories). Primary antibodies used in this study included rabbit anti-phospho-STAT1 (pTyr701) (D4A7) (1:1000, Cell Signaling Technology), mouse anti-STAT1 antibody (C-111) (1:500, Santa Cruz Biotechnology), and mouse anti-histidine tag antibody (clone AD1.1.10) (1:3000, Bio-Rad Laboratories). Secondary antibodies used in this study included HRP conjugated goat anti-mouse IgG (1:2000, Bio-Rad Laboratories) and HRP-linked goat anti-rabbit IgG (1:2000, Cell Signaling Technology).

Electrophoretic mobility shift assay (EMSA).

Proteins c-STAT1, c-STAT1-701D, c-STAT1-701E, c-STAT1-701pTyr, and c-STAT1-701CMF were overexpressed and purified from 293T cells transfected with pSTAT1 (132-712), pSTAT1 (132-712) Y701D, pSTAT1 (132-712) Y701E, or pCMFRS together with pSTAT1 (132-712) Y701TAG. The

same protocol of protein purification for MS analysis was followed. Imidazole was removed from purified protein by dialysis against a storage buffer (25 mM Tris-HCl, 150 mM NaCl, 5% glycerol, 2 mM DTT, pH 7.5). Concentrations of purified c-STAT1 variants were determined by Bradford assay.

EMSA followed a previously reported protocol.[10] PAGE-purified M67-1 and M67-2 primers were dissolved in annealing buffer (5 mM Tris-HCl, 50 mM KCl, 10 mM $MgCl_2$, pH 8.0) to a final concentration of 4 μM each. The mixture was heated to 94 °C, then slowly cooled to room temperature to form DNA duplex. In EMSA assay, 30 nM M67 was mixed with indicated concentrations of c-STAT1 variants in reaction buffer (20 mM HEPES, 4% Ficoll, 40 mM KCl, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 1 mM DTT, 0.2 mg/mL BSA, pH 8.0). The reaction mixture was incubated at room temperature for 30 min, then resolved on a pre-run 4-20% Novex TBE gel for 40 min at 4 °C. The gel was stained with SYBR™ Gold and visualized using Gel Doc™ XR+ imaging system.

Immunofluorescence assay.

The 293T-ΔSTAT1 cell line was used for immunofluorescence assay. Cells were seeded in poly-lysine coated, clear bottom, black wall 96-well plate (Greiner Bio-One). For IFN-γ induction, a final concentration of 100 ng/μL of IFN-γ (Cell Signaling Technology) was added into cell culture 30 min prior to fixation using 4% formaldehyde. Fixed cells were permeabilized with ice-cold 100% methanol at -20 °C for 10 min, incubated with blocking buffer (5% BSA and 0.1% Triton X-100 in DPBS) at room temperature for 1 h, incubated with anti-phospho-STAT1 (pTyr701) antibody (1:50 dilution, 2% BSA and 0.1% Triton X-100 in DPBS) at 4 °C for 12-16 h, then incubated with the Alexa Fluor® 594 conjugated-anti-rabbit IgG (1:500 dilution, 2% BSA and 0.1% Triton™ X-100 in DPBS) at room temperature for 2 h. Buffer of 0.1% Triton X-100 in DPBS (100 μL) was used for washing between each steps and after the incubation with the fluorophore-conjugated secondary antibody. Cells were visualized by a Nikon A1R-Ti2 confocal microscope.

c-STAT1 phosphorylation and purification.

c-STAT1 purified from HEK293T cells was phosphorylated by following a reported protocol with modifications.[11] Recombinant EGFR (5 μg) was activated with EGF (0.15 ng/μL) in 40 μL of kinase buffer (20 mM Tris-HCl, 50 mM KCl, 2 mM DTT, 1x protease/phosphatase inhibitor, pH 8.0) on ice for 10 min. A 100 μL phosphorylation reaction was set up in kinase buffer by adding purified c-STAT1 (20 μg), DTT (2 mM), ATP (5 mM), and $MnCl_2$ (10 mM) to the pre-incubated kinase preparation. Following incubation at 4 °C for 15 h, c-STAT1 was isolated using Ni Sepharose 6 Fast Flow resin, then desalted into HA buffer (20 mM Tris-HCl, 1 mM EDTA, 2 mM DTT, pH 8.0). Phosphorylated c-STAT1 was further purified using heparin Sepharose 6 Fast Flow resin. After eluting non-phosphorylated form at 150 mM KCl, phosphorylated c-STAT1 was eluted at 1 M KCl in HA buffer and collected for EMSA.

qPCR study of IRF1 expression.

The 293T-ΔSTAT1 cell line was used for quantitative PCR (qPCR) measurement of IRF1 expression. After transfection and 24 h of cultivation, cells were either induced or not induced with 100 ng/mL IFN-γ for 6 h, and then harvested for RNA extraction using RNeasy Mini Kit following the manufacture's protocol (Qiagen). RNA (500 ng) was reverse-transcribed into cDNA using iScript™ Reverse Transcription Supermix (Bio-Rad Laboratories). qPCR experiment was performed using SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad Laboratories) in the CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories). The relative expression of human IRF1 was evaluated, while human GAPDH was used as the reference gene. The relative mRNA abundance was determined using the $2^{-\Delta\Delta Ct}$ method.

Table S1. Primer list.

Primers	Sequences (5'->3')	Usage
pcDNA-TyrRS-F	tcccaggtccaactgcacggaagcttgccaccatggcaagcagtaacttgatt	EcTyrRS amplification
pcDNA-TyrRS-R	gaggctgatcagcgggtttaaacgggcccttatttccagcaaatcagacagt	EcTyrRS amplification
STAT1-GCA-F	ccaggtccaactgcacggaagcttgccaccatggcatctcagtggtacgaacttc	STAT1 amplification with GCA
STAT1-R1	agcgctaccgctaccgctacctactgtgttcatcat	STAT1 amplification
STAT1-EGFP-F	ggtagcggtagcgctgctgctgtgagcaagggcgag	EGFP amplification
STAT1-EGFP-R	cgggtttaaacgggcccttagtggtggtggtggtgctcgagcttgtacagctcgtcc	EGFP amplification
STAT1-Y701TAG-F	aactggaTAGatcaagactgagttg	introduce 701TAG
STAT1-Y701TAG-R	tcttgatCTAtccagttcctttagg	introduce 701TAG
STAT1-R	cagcgggtttaaacgggcccttatactgtgttcatcatactg	STAT1 amplification
STAT1 (132-712)-F	ccaggtccaactgcacggaagcttgccaccatggcaagcacagtgatgttagac	STAT1 (132-712) amplification
STAT1 (132-712)-R	tagtggtggtggtggtggtgctcgagaacttcagacacagaaatcaac	STAT1 (132-712) amplification
STAT1 gRNA-F	caccgtcatgacctcctgtcacagc	STAT1 gRNA
STAT1 gRNA-R	aaacgctgtgacaggaggtcatgac	STAT1 gRNA
STAT1 T7-F	gtttagaatttttgtgctaattgc	STAT1 KO sequencing
STAT1 T7-R	ccctggagctttaatgtattactat	STAT1 KO sequencing
M67-1	tccacagtttcccgtaaatgc	M67 duplex
M67-2	gcatttacgggaaactgtgga	M67 duplex
U6 H1-1	gtatcttatcatgtctgtagctagcaaggtcgggcaggaagag	U6-BstRNA$_{CUA}$ amplification
U6 H1-2	cgtcagcgttcgaattaggttcagcacagaaaaatg	U6-BstRNA$_{CUA}$ amplification
U6 H1-3	gctgaacctaattcgaacgctgacgtcatcaaccc	H1 amplification
U6 H1-4	tacccctccagagtggtctcatacagaacttataaga	H1 amplification
U6 H1-5	gtatgagaccactctggaggggtagcgaagtggc	BstRNA$_{CUA}$ amplification
U6 H1-6	ctagctagaggtcgacggtaggtacctctagagtgtcggcgtcccctgagg	BstRNA$_{CUA}$ amplification
IRF1-F	gaggaggtgaaagaccagagca	qPCR
IRF1-R	tagcatctcggctggacttcga	qPCR
GADPH-F	gtctcctctgacttcaacagcg	qPCR
GADPH-R	accaccctgttgctgtagccaa	qPCR

Table S2. Plasmid list.

Plasmids	characteristics	source
pEcTyrRS-lib	AmpR, TRP1, pADH-EcTyrRS, $tRNA_{CUA}$	Ref [6]
pGADGAL4	AmpR, LEU2, pADH1-GAL4 (44TAG+110TAG)	Ref [12]
pEcTyrRS-mutant 1	AmpR, TRP1, pADH-CMFRS-1 (Y37H, L71V, and D182G), $tRNA_{CUA}$	this study
pEcTyrRS-mutant 2	AmpR, TRP1, pADH-CMFRS-2 (Y37H, L71V, W129F, and D182G), $tRNA_{CUA}$	this study
pyeGFP-N149TAG	AmpR, LEU2, pTDH3-yeGFP N149TAG-8xHis	Ref [6]
pEGFP	AmpR, 3x$BstRNA_{CUA}$, CMV-EGFP Y40TAG	Ref [13]
pEGFP 2TAG	AmpR, 3x$BstRNA_{CUA}$, CMV-EGFP 40TAG+150TAG	Ref [6]
pEcTyrRS	AmpR, CMV-EcTyrRS, U6-$BstRNA_{CUA}$	Ref [6]
pCMFRS	AmpR, CMV-CMFRS-1, U6-$BstRNA_{CUA}$	this study
pSTAT1-EGFP	AmpR, CMV-STAT1 GCA-EGFP, U6-$BstRNA_{CUA}$	this study
pSTAT1 Y701TAG-EGFP	AmpR, CMV-STAT1 GCA Y701TAG-EGFP, U6-$BstRNA_{CUA}$	this study
pSTAT1	AmpR, CMV-STAT1, U6-$BstRNA_{CUA}$	this study
pSTAT1 Y701TAG	AmpR, CMV-STAT1 Y701TAG, U6-$BstRNA_{CUA}$	this study
pET30b-STAT1 (132-712)	KanR, T7/lacO-STAT1 (132-712)	Ref [10]
pSTAT1 (132-712)	AmpR, CMV-STAT1 (132-712), U6-$BstRNA_{CUA}$	this study
pSTAT1 (132-712) Y701TAG	AmpR, CMV-STAT1 (132-712) Y701TAG, U6-$BstRNA_{CUA}$	this study
pSpCas9(BB)-2A-GFP (PX458)	AmpR, CMV-SpCas9-2A-GFP	Addgene 48138; Feng Zhang lab
pSpCas9(BB)-2A-GFP-STAT1 gRNA	AmpR, CMV-SpCas9-2A-GFP, U6-sgRNA (STAT1)	this study
pCMFRS-3xtRNA	AmpR, CMV-CMFRS-1, 2xU6-$BstRNA_{CUA}$, 1xH1-$BstRNA_{CUA}$	this study
pCMFRS-5xtRNA	AmpR, CMV-CMFRS-1, 3xU6-$BstRNA_{CUA}$, 2xH1-$BstRNA_{CUA}$	this study
pCMFRS-7xtRNA	AmpR, CMV-CMFRS-1, 4xU6-$BstRNA_{CUA}$, 3xH1-$BstRNA_{CUA}$	this study
pCMFRS-9xtRNA	AmpR, CMV-CMFRS-1, 5xU6-$BstRNA_{CUA}$, 4xH1-$BstRNA_{CUA}$	this study

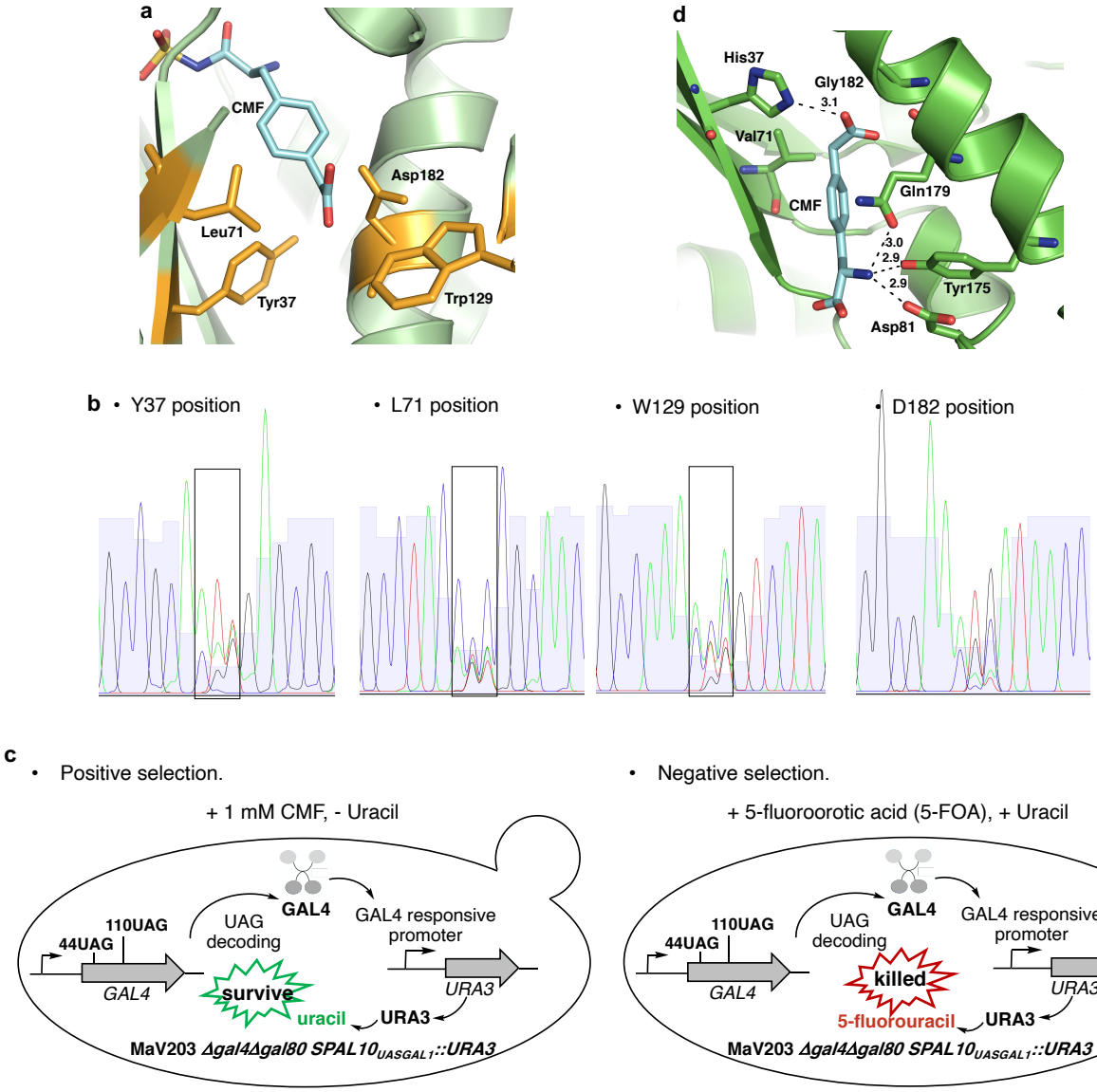


Fig. S1 Identification of CMFRS. **a.** Identification of key residues for the EcTyrRS library. The model was built using the structure of *E. coli* wild-type tyrosyl-tRNA synthetase (PDB ID: 1X8X). Protein ribbon diagrams are colored green. Side chains of the four library residues are shown as sticks in orange. Carbons of CMF ligand are in cyan, oxygens in red, and nitrogen in blue. **b.** DNA sequencing analysis of the EcTyrRS library. **c.** Selection of CMF-specific tRNA synthetase in *S. cerevisiae*. Following the positive selection, yeast cells expressing EcTyrRS variants that recognize CMF or natural amino acids can survive. After the negative selection, yeast cells expressing EcTyrRS variants that recognize natural amino acids will be killed, only ones expressing CMF-specific EcTyrRS can survive. **d.** Model of CMFRS was built from PDB ID: 6WN2 in PyMol. Docking was

conducted in AutoDock Vina. Protein ribbon diagram is colored green. Side chains of the three mutation sites (Y37H, L72V, D182G) and the three residues contribute to binding with the α-amino group are labelled and shown in stick. Carbons of CMF ligand are in cyan, oxygens in red, and nitrogen in blue.

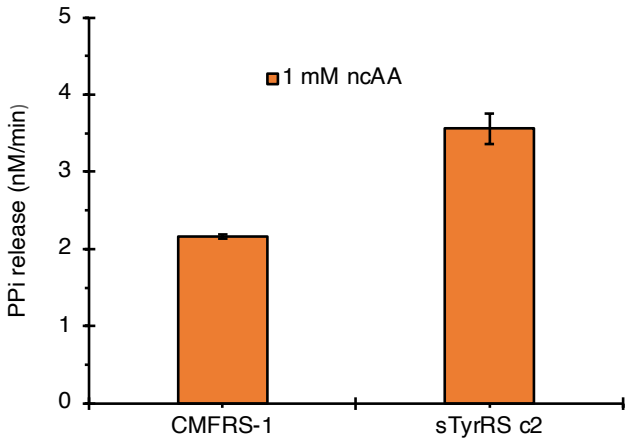


Fig. S2 Initial velocities of pyrophosphate release by CMFRS-1 and sTyrRS. Assays were carried out in the presence of 10 μM recombinant *Ec*tRNA$_{CUA}$. Each data point is the average of triplet measurements with standard deviation.

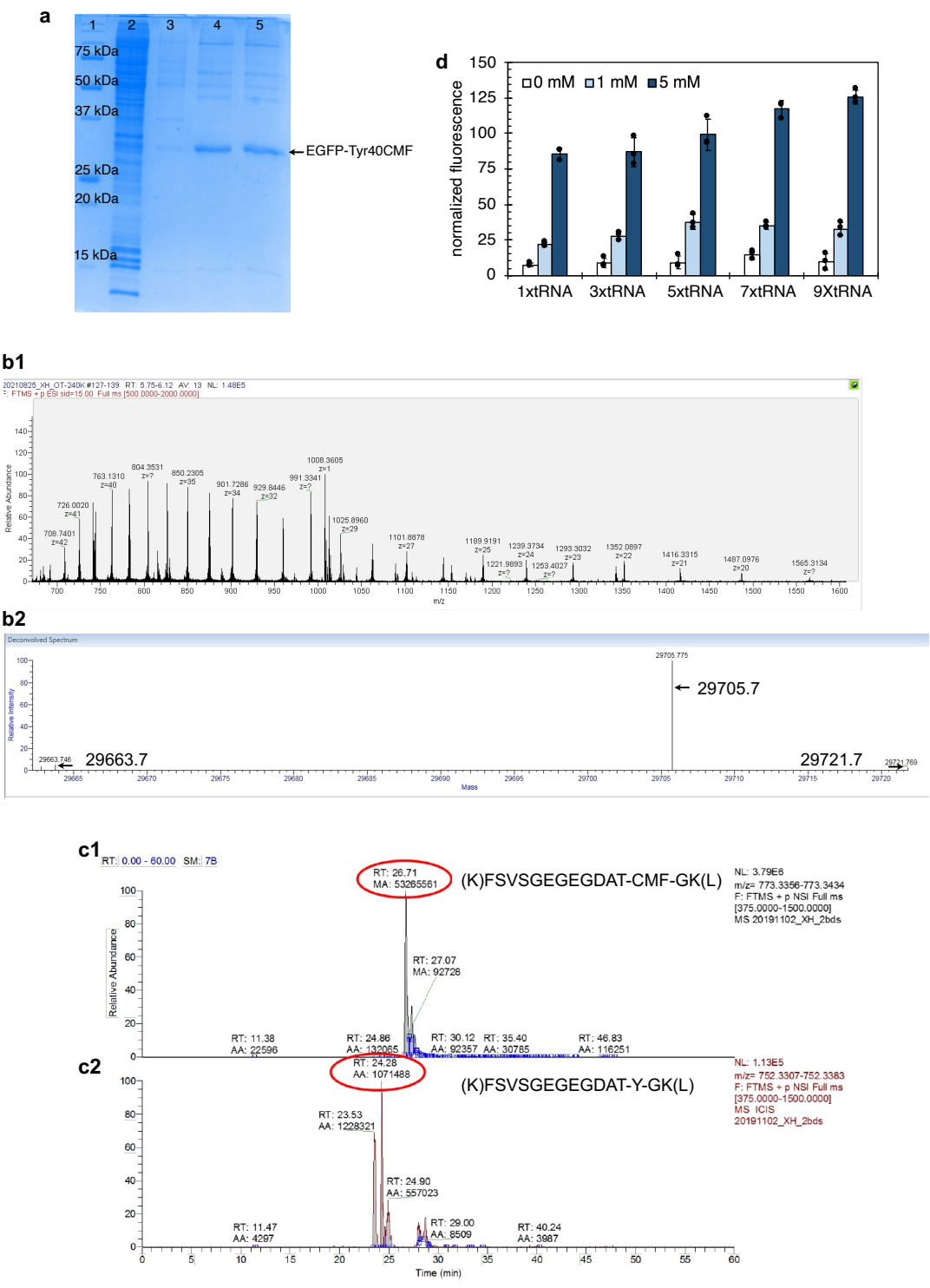


Fig. S3 CMF incorporation into EGFP. **a.** SDS-PAGE of EGFP-Tyr40CMF purified from 293T cells. Lane 1, molecular weight marker; lane 2, total fraction of cell lysate; lane 3, wash fraction; lane 4 and lane 5, elution fraction. The calculated molecular weight of EGFP-Tyr40CMF was 29.7 kDa. **b.**

Mass spectrometry analysis of purified full-length EGFP containing CMF. b1. Charge distribution. b2. Deconvoluted spectrum. Signal at 29705.775 Da corresponds to EGFP protein with CMF following the loss of N-terminal Met and subsequent acetylation at the Val residue. Signal at 29663.7 Da corresponds to EGFP protein following the loss of N-terminal Met and subsequent acetylation at the Val residue. **c.** Tandem mass spectrometry analysis of CMF incorporation into EGFP. c1. The peptide fragment containing CMF. c2. The peptide fragment containing Tyr. The retention time (RT) and peak areas of precursor ions for each peptide (MA and AA) were circled in red. Percentage of CMF incorporation was calculated,

$$\%\ \text{incorporation} = \frac{MA}{MA + AA} = \frac{53265561}{(53265561 + 1071488)} = 98\%$$

d. Suppression of two amber codons in EGFP-40TAG-150TAG using multi-copy tRNAs. 293T cells transfected with plasmids pCMFRS/nxtRNA (n = 1, 3, 5, 7, or 9) and pEGFP 2TAG cultured in the presence of 0, 1 or 5 mM of CMF in 96-well black cell culture plate with clear bottom. Data are plotted as the mean ± s.d. from n = 3 independent experiments.

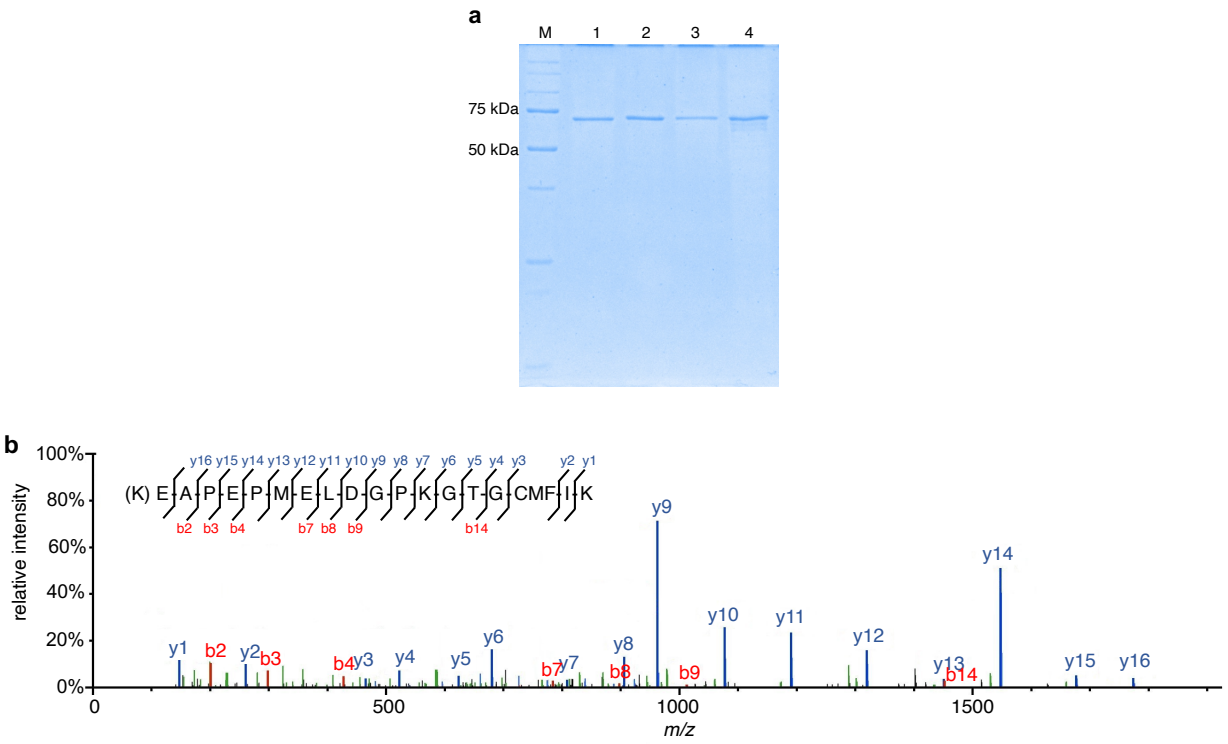
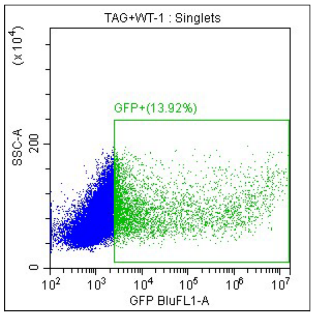
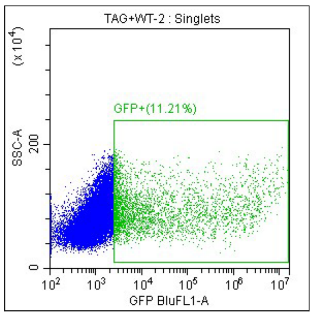
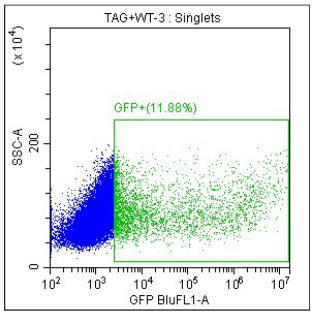


Fig. S4 Analysis of c-STAT1 variants purified from 293T cells. **a** SDS-PAGE analysis of purified c-STAT1 variants. Lanes 1, elute of c-STAT1; lanes 2, elute of c-STAT1-701D; lane 3, elute of c-STAT1-701E; lane 4, elute of c-STAT1-701CMF. The calculated molecular weight of c-STAT1 variants was 68.4 kDa. **b** LC-MS/MS analysis of purified c-STAT1-701CMF.

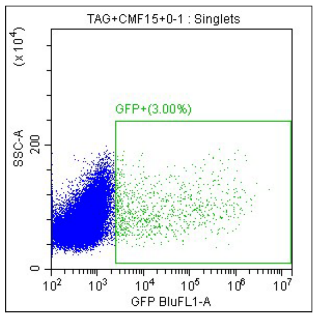
wt EcTyrRS

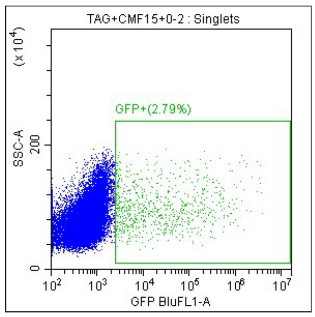


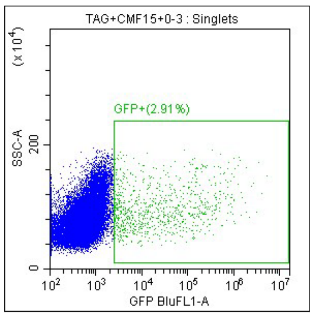




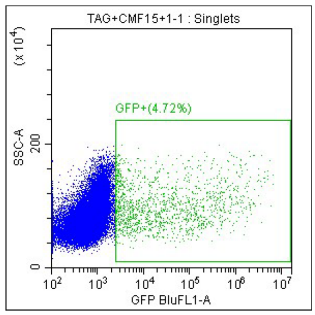
CMFRS, no CMF

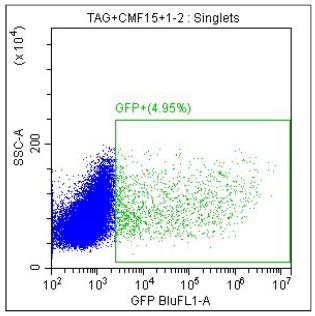


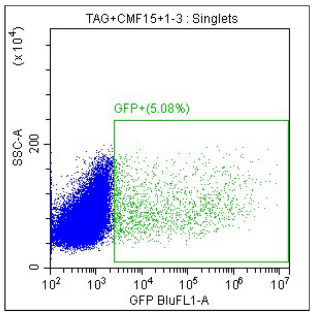




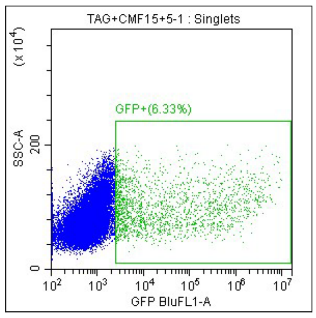
CMFRS, 1 mM CMF

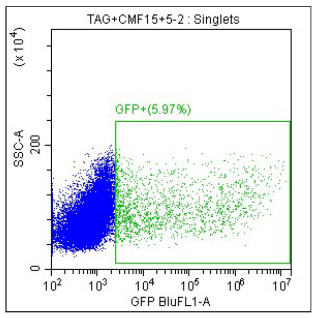


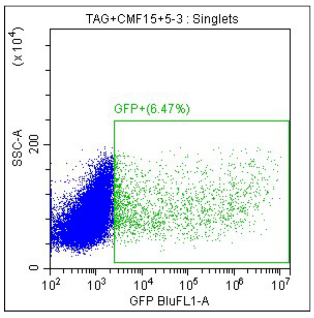




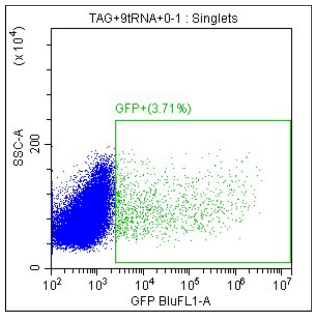
CMFRS, 5 mM CMF

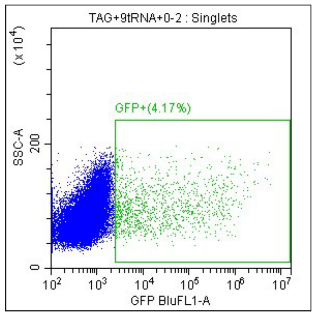


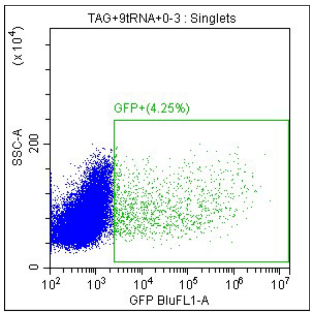




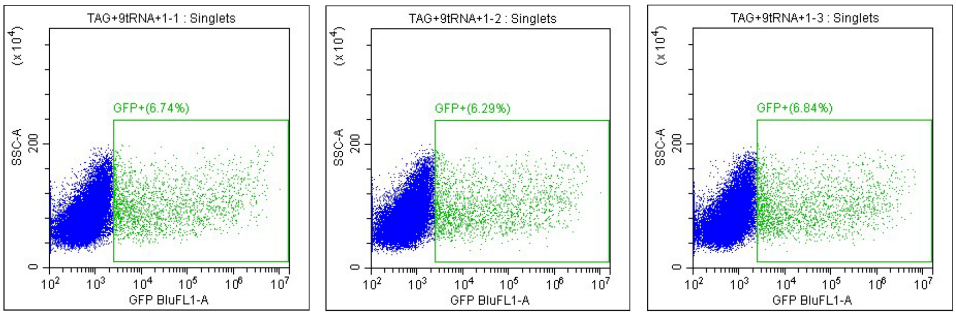
CMFRS 9xtRNA, no CMF



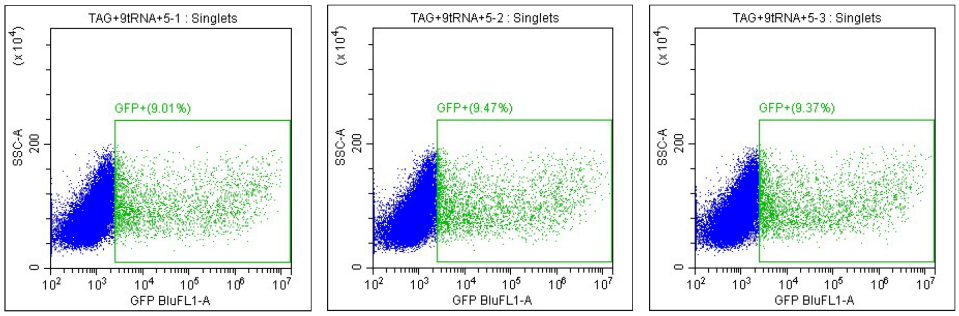




CMFRS 9xtRNA, 1 mM CMF



CMFRS 9xtRNA, 5 mM CMF



gating strategy for the analysis

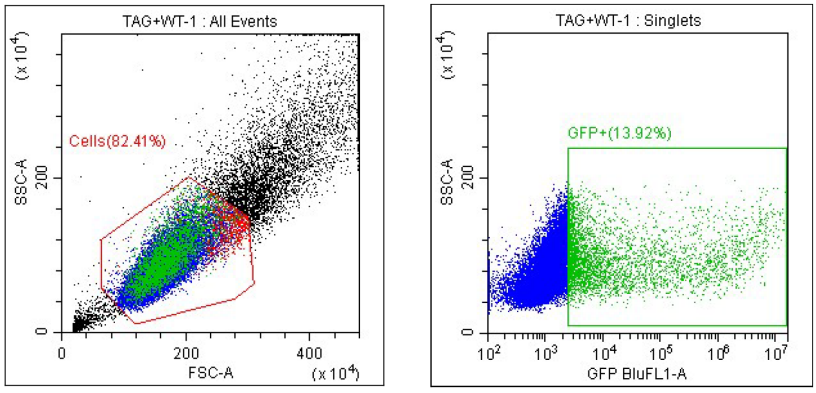


Fig. S5 Flow cytometry data. Flow cytometry analyses of 293T cells expressing the evolved CMFRS with an EGFP mutant that contains an amber mutation at position Tyr40

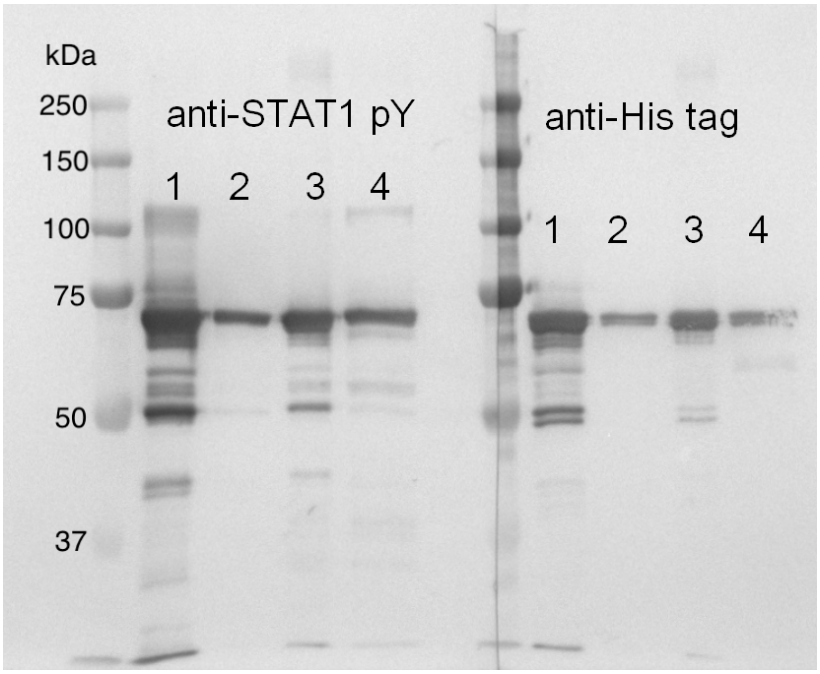


Fig. S6 STAT-1 phosphorylation. The protein samples were detected by phospho-STAT1 (pTyr701) antibody and anti-His tag antibody. Sample 1, phosphorylation reaction; sample 2, elute from Ni resin; sample 3, desalted prior to heparin resin; sample 4, elute from heparin resin.

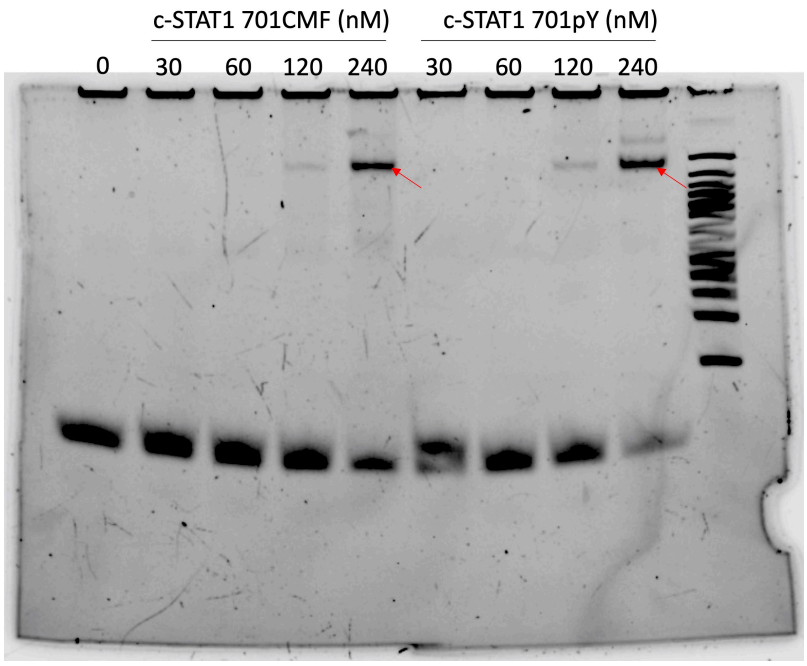


Fig. S7 EMSA assay. The DNA concentration was fixed at 30 nM. The protein concentrations varied from 0 to 240 nM. Free and bound (DNA-protein complex) DNAs were visualized by SYBR® Gold EMSA nucleic acid gel stain.

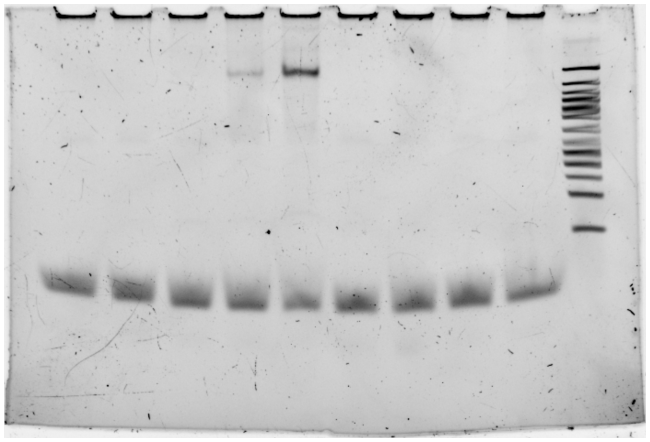


Fig. S8 Original gel image of Figure 2b.

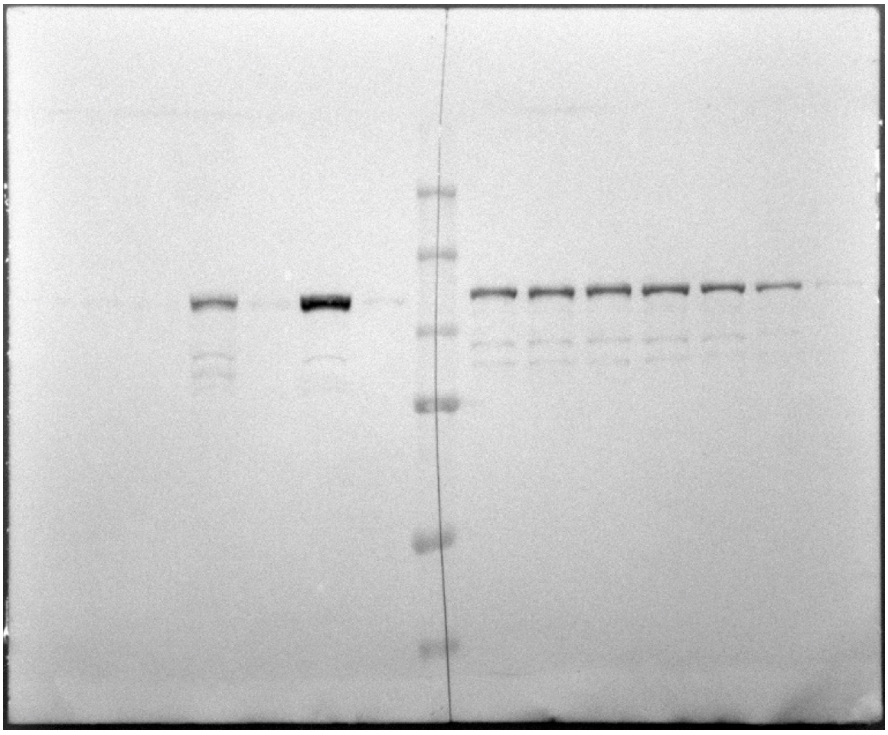
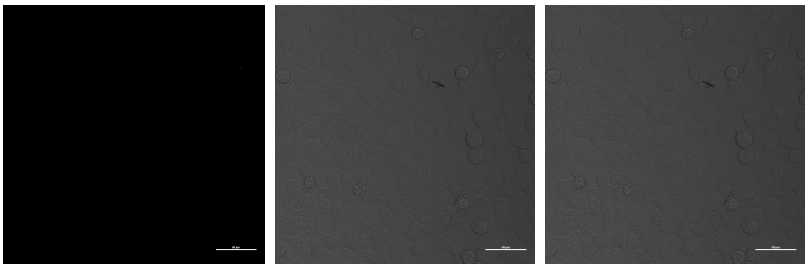
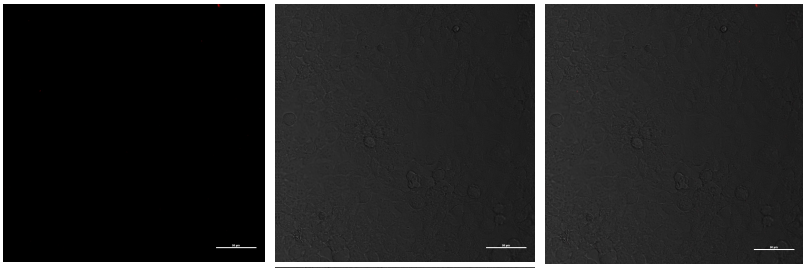


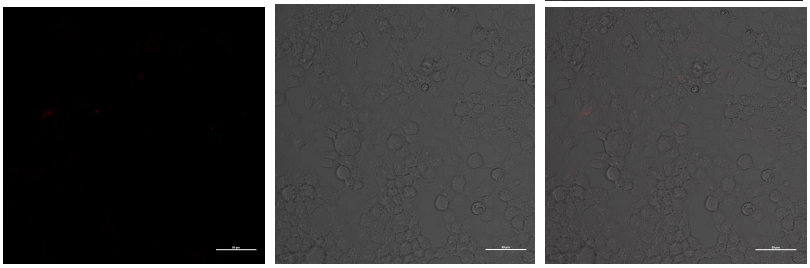
Fig. S9 Original gel image of Figure 3a.



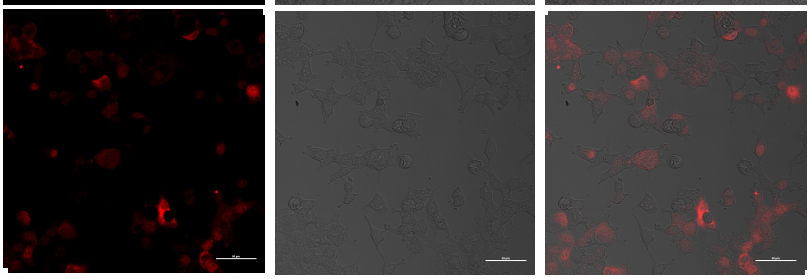
293T-ΔSTAT1 cells
no IFN-γ



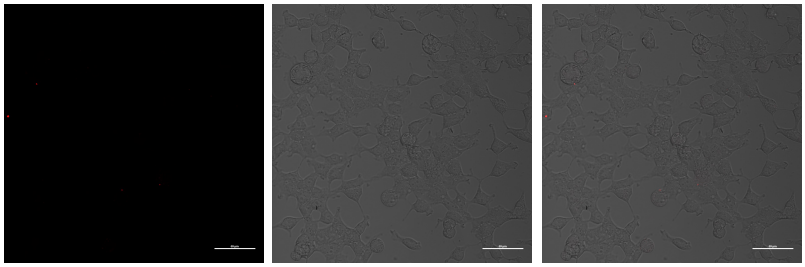
293T-ΔSTAT1 cells
with 100 ng/μL IFN-γ



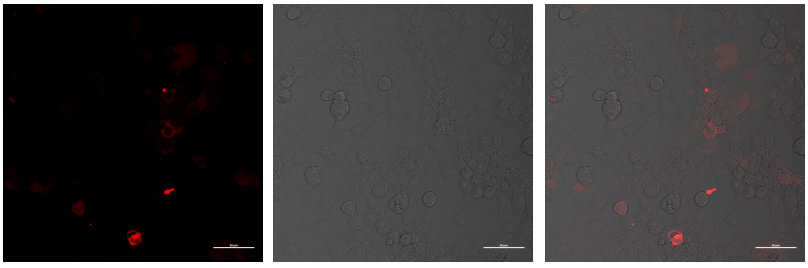
293T-ΔSTAT1 cells expressing
wild-type STAT1
no IFN-γ



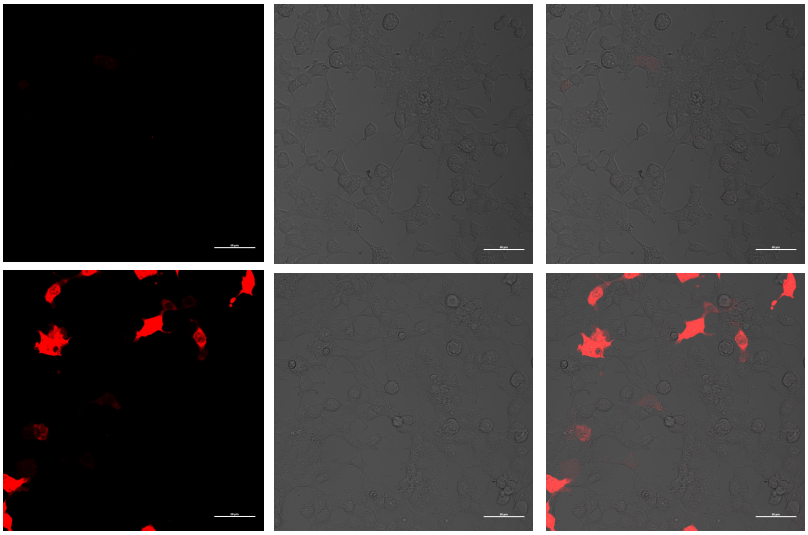
293T-ΔSTAT1 cells expressing
wild-type STAT1
with 100 ng/μL IFN-γ



293T-ΔSTAT1 cells expressing
STAT1-701TAG and wild-type EcTyrRS
no IFN-γ



293T-ΔSTAT1 cells expressing
STAT1-701TAG and wild-type EcTyrRS
with 100 ng/μL IFN-γ



293T-ΔSTAT1 cells expressing STAT1-701TAG and CMFRS no CMF

293T-ΔSTAT1 cells expressing STAT1-701TAG and CMFRS 5 mM CMF

Fig. S10 Images of 293T-ΔSTAT1 cells expressing STAT1 variants. The left panel shows fluorescence images, the middle panel shows brightfield images, and the right panel shows composite images of bright-field and fluorescence images. Scale bars, 50 μm.

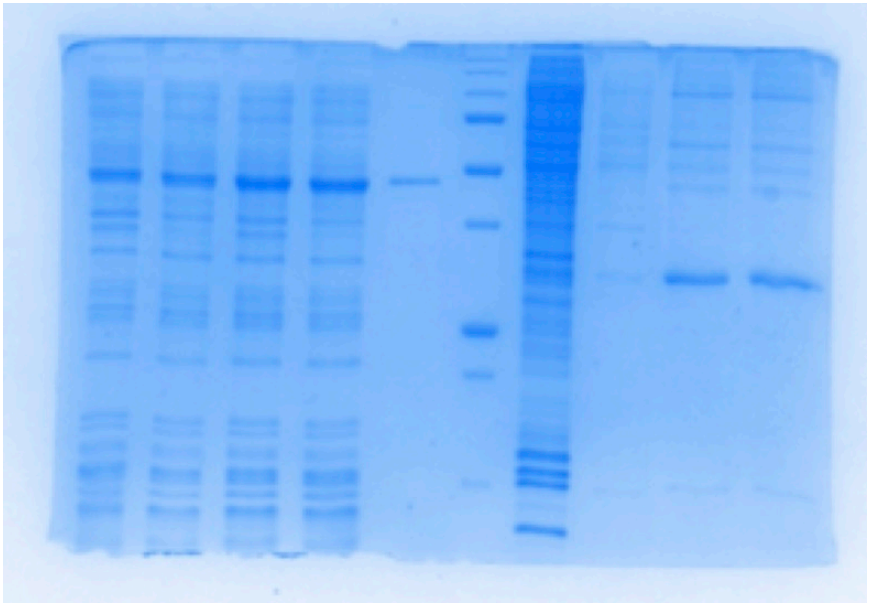


Fig. S11 Original SDS-PAGE of Figure S2a.

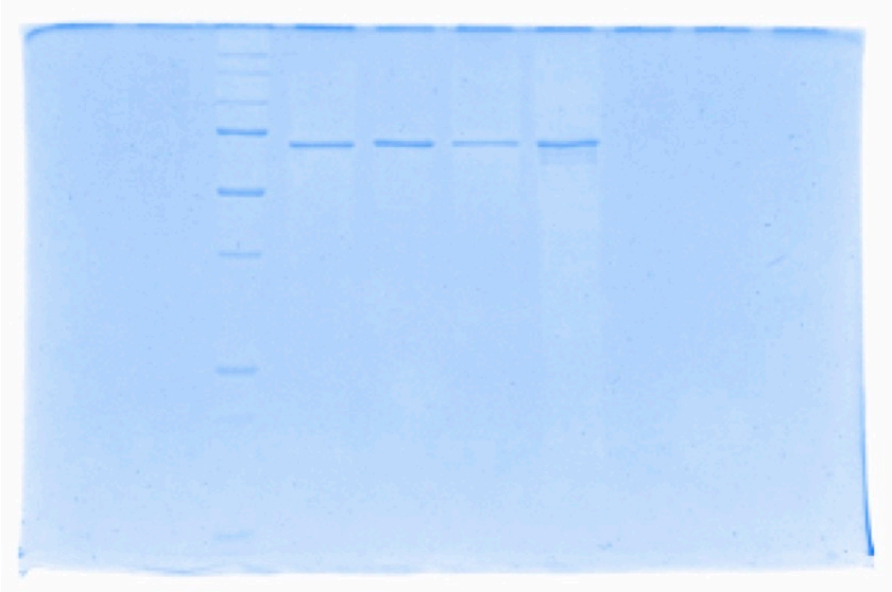


Fig. S12 Original SDS-PAGE of Figure S3a.

References

1. J. Xie, L. Supekova and P. G. Schultz, *ACS Chem. Biol.*, 2007, **2**, 474-478.
2. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat. Methods*, 2009, **6**, 343-345.
3. J. F. Sambrook, D. W. Russell and Editors, *Molecular cloning: A laboratory manual, third edition*, Cold Spring Harbor Laboratory Press, 2000.
4. S. Yamauchi, K. Takeuchi, K. Chihara, C. Honjoh, Y. Kato, H. Yoshiki, H. Hotta and K. Sada, *Scientific reports*, 2016, **6**, 1-11.
5. F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang, *Nat. Protoc.*, 2013, **8**, 2281-2308.
6. X. He, Y. Chen, D. G. Beltran, M. Kelly, B. Ma, J. Lawrie, F. Wang, E. Dodds, L. Zhang, J. Guo and W. Niu, *Nature Communications*, 2020, **11**, 4820.
7. L. Benatuil, J. M. Perez, J. Belk and C.-M. Hsieh, *Protein Engineering, Design and Selection*, 2010, **23**, 155-159.
8. A. Keller, A. I. Nesvizhskii, E. Kolker and R. Aebersold, *Anal. Chem.*, 2002, **74**, 5383-5392.
9. A. I. Nesvizhskii, A. Keller, E. Kolker and R. Aebersold, *Anal. Chem.*, 2003, **75**, 4646-4658.
10. T. Ju, W. Niu, R. Cerny, J. Bollman, A. Roy and J. Guo, *Mol. BioSyst.*, 2013, **9**, 1829-1832.
11. U. Vinkemeier, S. L. Cohen, I. Moarefi, B. T. Chait, J. Kuriyan and J. E. Darnell, Jr., *EMBO J.*, 1996, **15**, 5616-5626.
12. J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang and P. G. Schultz, *Science*, 2003, **301**, 964-967.
13. Z. Yuan, N. Wang, G. Kang, W. Niu, Q. Li and J. Guo, *ACS Synth. Biol.*, 2017, **6**, 721-731.