Science
AAAS

Supplementary Materials for

Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery

Michael Segel[1,2,3,4,5], Blake Lash[1,2,3,4,5], Jingwei Song[1,2,3,4,5], Alim Ladha[1,2,3,4,5], Catherine C. Liu[1,2,3,4,5,6], Xin Jin[2,3,4,7], Sergei L. Mekhedov[8], Rhiannon K. Macrae[1,2,3,4,5], Eugene V. Koonin[8], and Feng Zhang[1,2,3,4,5,*]

Correspondence to: zhang@broadinstitute.org

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Data S1

Affiliations: (1) Howard Hughes Medical Institute, Cambridge, MA 02139, USA; (2) Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; (3) McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; (4) Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; (5) Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; (6) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; (7) Society of Fellows, Harvard University, Cambridge, MA 02138 USA. (8) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

* Correspondence should be addressed to zhang@broadinstitute.org (F.Z.).

Materials and Methods

Cell culture

Human embryonic kidney cells (HEK-293FT, ThermoFisher R70007) were maintained at 37°C with 5% CO_2 in DMEM+GlutaMAX (Gibco) supplemented with 10% fetal bovine serum (VWR Seradigm 1500-500) and 1% penicillin/streptomycin (Gibco). Neuro-2A (N2a, ATCC #CRL-131) cells were maintained similarly to HEK293FT cells except in the absence of penicillin/streptomycin. HEK and N2a cells at 80-90% confluence were transfected with Lipofectamine 3000 (ThermoFisher, L3000001) according to manufacturer's guidelines. Mouse primary cortical neurons (ThermoFisher, A15585) were cultured according to manufacturer's guidelines.

Plasmids

A list of relevant plasmids can be found in Supplementary Table S4 with sequences found in Data S1. All coding sequences of mouse orthologs were amplified from a mouse embryonic stem cell cDNA library with PCR using gene-specific primers designed from the mm10 annotation. The cDNA library was prepared from a P30 mouse brain using the Protoscript II First Strand cDNA synthesis kit (NEB, E6300S). RNA from mouse brains were purified using the Directzol RNA Miniprep kit (Zymo, R2061). Plasmids were cloned using PCR amplification with Phusion Flash 2x Master Mix (Thermo Fisher Scientific F548S) and assembled with Gibson Assembly 2x Master Mix (NEB, E2611L), or blunt-end ligated with KLD Enzyme Mix (NEB, M0554S).

Western Blot

Whole-cell lysate of cultured HEK293FT and N2a cells were harvested using M-PER Mammalian Protein Extraction Reagent (ThermoFisher, 78501) with Halt Protease Inhibitor Cocktail (ThermoFisher, 78429). Harvested mouse brain tissue was homogenized using a glass dounce homogenizer in M-PER containing Protease Inhibitor Cocktail. Harvested mouse plasma was prepared from blood collected by cardiac puncture, and spun with 5 µL of 0.5 M EDTA (ThermoFisher, AM9260G). Plasma particles were precipitated using the Total Exosome Isolation Kit, according to the manufacturer's guidelines (ThermoFisher, 4484450). Protein concentrations were analyzed with the Pierce 660nm Protein Assay Reagent (ThermoFisher, 22660) by Nanodrop. Protein samples were loaded onto 4-12% Bis-Tris gels (ThermoFisher, NW04125) and electrophoresis was performed according to manufacturer's guidelines. Proteins were transferred to PVDF membranes using the iBlot Transfer Stack (ThermoFisher, IB401031). Blots were blocked for 1 hour in Intercept (TBS) Blocking Buffer (LI-COR, 92760001), incubated with primary antibodies in Intercept (TBS) Blocking Buffer overnight at 4°C, washed 3 times in 1X TBST, incubated with secondary antibodies in Intercept (TBS) Blocking Buffer for 1 hour at room temperature, washed 3 times in 1X TBST, then imaged with BioRad ChemiDoc XRS+ . Quantification was performed using ImageJ western blot analysis toolkit.

RT-qPCR

For all RT-qPCR experiments, total RNA was extracted with TRI Reagent (Zymo Research, R2050-1-200) and purified using the Direct-zol RNA Microprep Kit (Zymo Research, R2061).

Total RNA was reverse transcribed using the Agilent cDNA qPCR Synthesis Kit (Agilent, 600559) according to manufacturer's guidelines. Gene-specific primers used for RT-qPCR are shown in Supplementary Table S3. PCR was performed using Fast SYBR Green Master Mix (ThermoFisher, 4385610) on a Roche Lightcycler 480 according to manufacturer's guidelines.

Purification of VLPs from cell culture media

Unless otherwise noted, cell culture media containing extracellular VLPs were harvested, clarified by low centrifugation at 2000 *g* for 10 mins, filtered through a 0.45 μm Durapore PVDF filter (EMD Millipore, #SE1M003M00), and ultracentrifuged at 120,000xg *g* for 2 hours at 4°C in a Beckman Coulter Optima XPN-80 ultracentrifuge. The supernatant was decanted, and the remaining pellet was resuspended up to 100 uL 1X PBS. To remove DNA and RNA not protected by a vesicle, purified VLPs were treated with 2.5 μL Micrococcal Nuclease (NEB, M0247S) in 1X Micrococcal Nuclease buffer and 100 ug/mL BSA for 2 hours at 37°C. Micrococcal nuclease was then inactivated by the addition of EDTA to a final concentration of 10 mM. RNA was extracted with Tri reagent using the Direct-zol RNA Microprep Kit.

Mice

Wild-type female C57BL/6J-Elite mice were obtained from Charles River Laboratories. Female SpCas9 mice were obtained from a breeding colony maintained by our group (*42*). All housing, breeding, and procedures were performed according to protocols approved by the Institutional Animal Care and Use Committees (IACUC) of the Broad Institute of MIT and Harvard.
All animal line generation was performed at Harvard Genome Modification Facility (GMF). Briefly, SpCas9 RNPs and ssDNA donors with the HA tag and 130 bp of homology were injected into wild-type C57BL/6J embryos. Neonates were genotyped for correct insertions using NGS on DNA isolated from ear clippings (Lucigen; QE09050)

Homology-based mining of retroviral proteins

In order to obtain a comprehensive census of Gag homologs encoded in human and mouse genomes, we ran HHpred search with a multiple alignment of mammalian orthologs of the human Arc protein employed as the query (https://toolkit.tuebingen.mpg.de/tools/hhpred (PMID: 33315308). In addition to the highly significant hits in the human proteome, this search retrieved the homologous domains from the Pfam database: Gag-p24, Gag-p30, gag-gag2, SCAN, PNMA, DUF1759, and DUF4219. Although all these domains are homologous, they belong to two different Pfam clans, namely, clan GAG-polyprotein (CL0523) and clan Viral_Gag (CL0148). HHpred searches were repeated using multiple alignments of each of these Pfam domains as queries. These seed alignments were also used to build HMM profiles (PMID: 22039361). These profiles and the profile for mammalian Arc orthologs were used to search the human proteome (Annotation Release 109.20210226) and the mouse proteome (Annotation Release 109), from NCBI ftp site
(https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/405/GCF_000001405.39_GRCh38.p13/)

(https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/10090/109/GCF_000001635.27_GRCm39/.) In order to confirm the significant hits produced by these search as Gag homologs, they were used as queries for reciprocal searches of the Pfam (PMID: 30357350), SMART (PMID: 33104802), and CDD (PMID: 31777944) databases. The confirmed Gag homologs encoded in the human and mouse genomes with apparent orthologous relationships among them are listed in Supplementary Table S1. Additional confirmed Gag homologs encoded in the mouse genomes are listed in Supplementary Table S2. Importantly, retroviruses integrated into mammalian genomes in multiple waves because of which one to one orthologous relationships can be demonstrated only for a subset of retrovirus-derived genes.

Protein production, purification, and size-exclusion chromatography

Gag proteins cloned from mouse cDNA were subcloned into a TwinStrep-SUMO expression backbone and transformed into Rosetta™ 2(DE3) (Millipore Sigma, 71400). Overnight cultures were inoculated with single colonies and grown. The following day 1 mL of overnight culture was used to inoculate 50 mL of Terrific Broth (TB) medium. Cultures were grown at 37°C until O.D. 0.6-0.7 and induced with 0.3 mM IPTG at 25C for 16-18 hours. Cultures were spun down and resuspended in 50 mM Tris-HCl, pH 8.0 (Invitrogen), 500 mM NaCl (Millipore Sigma), 1 mM DTT (Millipore Sigma) with the addition of cOmplete™ Protease Inhibitor Cocktail (Roche). Bacteria were lysed with 25 kpsi of pressure on an LM-20 Microfluidizer (Microfluidics International Corporation). Insoluble components were spun at 20,000 *g* for 30 mins at 4°C and the supernatant was decanted and 0.45 um filtered. Protein was isolated by flowing over Strep Tactin Superflow Plus resin (Qiagen, 30004) followed by 5 C.V. washing with 50 mM Tris-HCl, pH 8.0, 500 mM NaCl. Protein was eluted with 50 mM Tris-HCl, 500 mM NaCl, 2.5 mM desthiobiotin. The resulting proteins were cleaved overnight with homemade SUMO protease to remove tags and size excluded into 50 mM Tris-HCl, 150 mM NaCl on an ÄKTA Pure system using a Superdex 200 Increase 10/300 GL column (Cytiva).

Negative stain TEM and cryo TEM

Negative stain TEM and cryo TEMmicroscopy experiments were performed at MIT's Koch Nanotechnology Materials Core Facility. In sample preparation for cryo-electron microscopy, 3 uL of sample and buffer containing solution was dropped on a lacey copper grid coated with a continuous carbon film and blotted to remove excess sample without damaging the carbon layer by Gatan Cryo Plunge III. Grid was mounted on a Gatan 626 single tilt cryo-holder equipped in the TEM column. The specimen and holder tip were cooled down by liquid-nitrogen, which keeps maintaining during transfer into the microscope and subsequent imaging. Imaging on a JEOL 2100 FEG microscope was done using a minimum dose method that was essential to avoid sample damage under the electron beam. The microscope was operated at 200 kV and with a

magnification in the range of 10,000~60,000 for assessing particle size and distribution. All images were recorded on a Gatan 2kx2k UltraScan CCD camera.

In sample preparation for negative stain-electron microscopy, 7 uL of sample and buffer containing solution was dropped on a 200 meshes copper grid coated with a continuous carbon film and waited for 60 seconds and removed excess solution by touching the grid with a kimwipes and then 10uL of negative staining solution, phosphotungstic acid, 1% aqueous solution was dropped on the TEM grid and immediately removed it by kimwipes and 10 uL of the stain is then applied to the grid and after 40 seconds, the excess stain is removed by touching the edge with kimwipes. Finally, dried the grid at room temperature. After that, the grid was mounted on a JEOL single tilt holder equipped in the TEM column. The specimens were cooled down by liquid-nitrogen and imaging on an JEOL 2100 FEG microscope was done using minimum dose methods that were essential to avoid sample damage under the electron beam. The microscope was operated at 200 kV and with a magnification in the range of 10,000~60,000 for assessing particle size and distribution. All images were recorded on a Gatan 2kx2k UltraScan CCD camera.

CRISPRa of CA-containing-like genes in N2a cells

Guides were selected from the mouse SAM V1 guide library (*19*) for each of the screened CA-containing genes (Supplementary Table S3). Guides were annealed and cloned via golden gate into the UniSAM addgene plasmid (Addgene, 99866) as covered in previous publications from our lab (*19*).

Plasmids were transfected into N2a cells at 80% confluency in 1x T225 flasks per replicate per condition. Cell culture media was replaced with fresh media 5 hours post-transfection. VLPs were harvested by filtration and ultracentrifugation 48 hours after transfection as indicated above, treated with micrococcal nuclease, and lysed in trizol, as covered in the previous section.

Co-IP Mass spectrometry

N2a cells cultured in 1x T225 flasks per replicate were transiently with lipofectamine 3000 with N-terminal and C-terminally HA tagged-*Peg10* overexpression plasmids. Cells were lysed and PEG10 protein was immunoprecipitated using the Pierce™ HA-Tag Magnetic IP/Co-IP Kit (Thermofisher; 88838). Briefly, cells were lysed in the Pierce kit IP lysis buffer and rotated at

room temperature for 30 mins with 25 µL of Pierce HA magnetic beads. Protein was eluted from the beads using the kit's acidic elution buffer and neutralized.

In-Gel Protein Digestion

Gel bands were cut out, reduced with 5 mM DTT and alkylated with 10 mM iodoacetamide and digested with either trypsin at 37°C essentially as described previously (*43*).

Immunoprecipitation Protein Digestion

Beads were digested according to the S-Trap Micro Spin Column Digestion Protocol (www.protifi.com).

LC-MS/MS Analysis for Gel Bands

The dried peptide mix was reconstituted in a solution of 2% formic acid (FA) for MS analysis. Peptides were loaded with the autosampler directly onto a 50cm EASY-Spray C18 column (ES803a, Thermo Scientific). Peptides were eluted from the column using a Dionex Ultimate 3000 Nano LC system with a 14.8 minute gradient from 1% buffer B to 23 % buffer B (100% acetonitrile, 0.1% formic acid), followed by a 0.2 minute gradient to 80% B, and held constant for 0.5 min. Finally, the gradient was changed from 80% buffer B to 99% buffer A (100% water, 0.1% formic acid) over 0.5 minute, and then held constant at 99 % buffer A for 14 more mins. The application of a 2.2 kV distal voltage electrosprayed the eluting peptides directly into the Thermo Exploris480 mass spectrometer equipped with an EASY-Spray source (Thermo Scientific). Mass spectrometer-scanning functions and HPLC gradients were controlled by the Xcalibur data system (Thermo Scientific). MS1 scans parameters were 60,000 resolution, AGC at 300%, IT at 25ms. MS2 scan parameters were 30,000 resolution, isolation width at 1.2, HCD collision energy at 28%, AGC target at 100% and max IT at 55ms. 15MS/MS scans were taken for each MS1 scan. Expected peptides for Retrotransposon-derived protein PEG10 GN=Peg10 were put into an inclusion list.

LC-MS/MS Analysis for IPs

The dried peptide mix was reconstituted in a solution of 2% formic acid (FA) for MS analysis. Peptides were loaded with the autosampler directly onto a 50cm EASY-Spray C18 column (ES803a, Thermo Scientific). Peptides were eluted from the column using a Dionex Ultimate 3000 Nano LC system with a 3 minute gradient from 1% buffer B to 5% buffer B (100 % acetonitrile, 0.1% formic acid), followed by a 36.8 minute gradient to 25%, and a 10.2 minute gradient to 35% B, followed by a 0.5 minute gradient to 80% B, and held constant for 4.5 mins. Finally, the gradient was changed from 80% buffer B to 99% buffer A (100% water, 0.1% formic acid) over 0.1 minutes, and then held constant at 99% buffer A for 19.9 more mins. The application of a 2.2 kV distal voltage electrosprayed the eluting peptides directly into the Thermo Exploris480 mass spectrometer equipped with an EASY-Spray source (Thermo Scientific). Mass spectrometer-scanning functions and HPLC gradients were controlled by the Xcalibur data

system (Thermo Scientific). MS1 scans parameters were 120,000 resolution, AGC at 300%, IT at 50ms. MS2 scan parameters were 30,000 resolution, isolation width at 1.2, HCD collision energy at 28%, AGC target at 300% and IT set to Auto. Cycle time for MS2 was 3 sec for each MS1 scan.

Database Search

Tandem mass spectra were searched with Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version IseNode in Proteome Discoverer 2.5.0.400). Sequest was set up to search a mouse uniprot database (database version March 21, 2020; 55699 entries containing common contaminants) assuming the digestion enzyme trypsin. Sequest was searched with a fragment ion mass tolerance of 0.02 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine was specified in Sequest as a fixed modification. Oxidation of methionine was specified in Sequest as a variable modification. Scaffold (version Scaffold_4.11.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Percolator posterior error probability calculation (*44*). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (*45*). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Cloning of PEG10 variants

Structural domains within PEG10 were identified via homology detection of the full Peg10 peptide sequence against known domains via HHPred with the HHsuite v3.0 release at default settings (database: PDB_mmCIF70) (*46*). The protease domain in PEG10 was made inactive by mutating the catalytic aspartic acid to alanine (D491A) using oligos listed in Supp. Table S3. In-frame deletions of the NC and the RT domains were made using oligos listed in Supp. Table S3. *Peg10* codon optimization was performed using the IDT codon optimization tool for human and gene blocks were ordered from IDT.

AAV PHP.eB production and delivery

AAV CRISPR guides were designed using Benchling online CRISPR gRNA design tool and validated for efficiency in mouse embryos, and cloned with golden gate method into PX552 (Addgene, 60958). PHP.eB AAV vectors were generated as previously described (*47*). In brief, HEK293 cells were grown in T225 flasks plastic dishes and transfected with the pHelper plasmid, the PHP.eB capsid plasmid, and the Px552 transgene plasmid. Five days later, cells were lysed, and virus was isolated using Optiprep density-gradient medium (Sigma; D1556) and ultra-centrifuged at 350,000 *g*. The viral layer was isolated and concentrated using Amicon Ultra-15 centrifugal filter units (Sigma; Z648043-24EA). AAV titer was determined using SYBR

Green qPCR. Two vectors with two separate *Peg10* targeting gRNAs were produced in parallel and pooled in equal titer. For in vivo administration of the virus, P1 mice were anesthetized and 5e11 viral genomes for each virus were retro-orbitally injected into in-house generated SpCas9 mice. Mice are publicly available (Jackson labs, 026179).

FACS of neuronal nuclei

Prefrontal cortex of P25 mice was dissected and flash frozen in liquid nitrogen. Nuclei were prepared from fresh frozen brains using the Nuclei EZ Prep kit (Sigma, NUC101), as previously described (*48*). Briefly, nuclei were dounced exactly 20 times in 1 mL of Nuclei lysis buffer on ice immediately after removal from dry ice, and washed twice in Nuclei lysis buffer. Following prep, nuclei were counterstained with DAPI and resuspended in PBS with 0.5% BSA (ThermoFisher, 15260037). For each condition, 20,000 neuronal nuclei were sorted based on DAPI and GFP on a Sony MA900 Cell Sorter directly into Tri reagent. RNA was extracted using the Direct-zol RNA Microprep Kit and mRNA sequencing library was prepared as listed below.

mRNA-sequencing of nuclei, whole cell RNA and EVs

N2a cells, purified EVs, sorted neuronal nuclei, and primary mouse cortical neurons were lysed with TRI Reagent (Zymo Research, R2050-1-200), and total RNA was extracted using Direct-zol RNA Microprep Kit (Zymo Research, R2061) and treated with DNase I (Zymo Research, E1010). Whole cell mRNA was enriched using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490), and the multiplexed RNA sequencing library was prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7775). Libraries were sequenced on the Illumina NextSeq 550 with 50 cycles for read 1 (forward), 25 cycles for read 2 (reverse) for a read coverage of approximately 15-20 million reads per sample.

Raw reads were trimmed using Trimmomatic (*49*) and quality control was performed using fastqc (*50*). Pseudoalignments and differential gene expression analysis was performed using Kallisto and Sleuth (*51*, *52*). Full read alignments were generated using STAR (*53*) and converted to indexed BAM files with SAMtools to generate read tracks.

Indel sequencing of in vitro edited cells and sorted neuronal nuclei and N2a cells

In vitro 96-well plates of tissue culture cells were lysed with 20 µL of QuickExtract (Lucigen, QE09050). The target region was amplified from genomic DNA by PCR with primers described in Supp. Table S3 and barcoded with indexed Illumina P5 and P7 primers for NGS. Libraries were purified with Ampure XP (Beckman Coulter), quantified with Qubit hsDNA reagent, and sequenced on an Illumina MiSeq system (300bp read 1). Indels were quantified from the

resulting library using CRIS.py, a python based tool used to detect indels from NGS libraries (*54*).

In vivo and in vitro eCLIP

Female prefrontal cortex of P30-P32 HA-tagged mice generated and bred in house were dissected and flash frozen in liquid nitrogen. To prepare the CLIP library, 300 mg of brain was pulverized with a pestle on ice and immediately UV irradiated twice at 400 mJ/cm^2 (*55*). The tissue was then processed using an adapted eCLIP protocol (*56*). Briefly, tissue was resuspended in the recommended CLIP lysis buffer and sonicated for 5 mins at low intensity on a Bioruptor Pico sonicator (Diagenode, B01080010). For in vitro eCLIP experiments, ~2 million N2A cells were transiently transfected in a 6-well plate in the same manner as listed above. Three days following transfection, media was changed to PBS (ThermoFisher, 10010023) and UV irradiated once at 400 mJ/cm^2, resuspended in the CLIP lysis buffer and scraped from the plate.

Unless otherwise stated, all methods and components for the CLIP preparation were exactly as listed previously (*56*). Briefly, lysed samples were nuclease treated as recommended and immunoprecipitated using Pierce Anti-HA magnetic beads (ThermoFisher, 88836). As per the protocol, immunoprecipitated samples were washed and overnight ligated with the recommended CLIP 3' RNA adaptor. Simultaneously, a subsample was blotted for HA to determine the efficacy of the precipitation and expected band size. Following overnight ligation, samples were loaded onto a NuPAGE gel (ThermoFisher, NP0321PK2) and transferred onto a nitrocellulose membrane (ThermoFisher, IB23001). Bands were excised with a scalpel on a sterile bioassay plate, transferred to an eppendorf tube, and sequentially digested in Proteinase K with urea, and then Acid-Phenol:Chloroform (ThermoFisher, AM9722) on an orbital thermoshaker at 37ºC. RNA was extracted and prepared for sequencing as recommended from the CLIP protocol. The samples were run on the NextSeq 550 with 50 cycles for read 1 (forward), 25 cycles for read 2 (reverse). Sequencing reads were aligned with STAR, quantified using htseq-count, and differential expression analysis was performed with DESeq2.

Cre-reporter cell line generation

A lentiviral loxP-GFP Cre reporter cassette was generated by subcloning the loxP-GFP cassette from RV-Cag-Dio-GFP (Addgene #87662) into a lentiviral transfer plasmid encoding a blasticidin resistance gene. To produce virus HEK293FT cells were seeded at 1e7 cells per 15-cm dish. 16 hours later cells were co-transfected with 5 ug psPAX2 (Addgene #12260), 4.7 ug pMD2.G (Addgene #12259), and 7.7 ug of the Cre reporter plasmid using PEI HCl MAX (Polysciences 24765-1), media was changed 4 hours post transfection. 48 hours later viral

supernatant was harvested, spun at 2000 *g* for 10 mins to remove cell debris, 0.45 um filtered, aliquoted, and frozen at -80°C for later use.

N2a reporter cell lines were created by seeding cells at 50% confluency in 6-well plates at day 0. 1 mL of clarified virus supernatant was added to the cells along with 8 ug/mL polybrene (TR1003G). Media was changed one day later, and cells were selected for two weeks starting on day 3 with 10 ug/mL Blasticidin-HCl (Thermo Fisher Scientific A1113903). Successful reporter line generation was confirmed by transfection of a Cre encoding plasmid.

Lentivirus and VLP production transfer.

Both lentivirus and Peg10 VLPs packaging Cre-encoding mRNA were produced and purified in an identical manner. HEK293FT cells were seeded at a density of 4e6 cells per 10-cm dish. The next day, cells were transfected with 2 ug each of i) [vector RNA], ii) a Peg10 overexpression plasmid or psPAX2 (Addgene #12260), and iii) pMD2.G (Addgene #12259). Cells were washed with PBS and fresh medium 4 hours post-transfection. 48 hours later, the culture supernatant was harvested, and particles were purified in the same manner as the purification of EVs from the CRISPRa experiment described above.

For all N2a transfer experiments, Cre-reporter N2a cells were plated onto 96-well plates at 5e4 cells per well one hour before transfer. 20 ul of purified VLPs or lentivirus derived from a 10-cm dish were applied to N2a cells in triplicate. Cells were incubated for 72 hours before flow cytometry analysis.

For tail-tip fibroblast transfer experiments, cells were plated onto gelatin coated 96-well plates at 3e4 cells per well 24 hours before transfer. 20 ul of purified VLPs or lentivirus derived from a 10-cm dish were applied to cells in triplicate the next day. For endogenous fusogen pseudotypes, the transduction enhancer vectofusin-1 (Miltenyi Biotec) was added per manufacturer's instructions. Cells were incubated for 72 hours before analysis.

Flow cytometry

N2a cells in 96-well plates were washed once with 1X PBS, and dissociated with TrypLE (ThermoFisher, 12604013). Cells were resuspended with 2% FBS in PBS, centrifuged at 500 *g* for 5 min, and washed once in PBS. Cells were viability stained for 30 mins with Zombie NIR Fixable Viability Dye (Biolegend, 423105) in PBS at 4°C. Cells were washed twice in 2% FBS in PBS and resuspended in 2% FBS in PBS + 2 mM EDTA for analysis on a Beckman Coulter

Cytoflex S flow cytometer. Analysis was performed using FlowJo v10.7 (BD Biosciences). Representative gating schemes are shown in Figure S8.

Cell imaging

N2a, Ai9 tail-tip fibroblasts, or HEK293FT cells were plated on glass bottom plates (Cellvis, P96-1.5H-N). Following VLP transfer experiments, cells were fixed with 4% PFA in 1X PBS for 15 mins at room temperature, washed 3 times with 1X PBS for 5 mins each, permeabilized with 0.1% Tween in PBS for 5 mins, and counterstained with DAPI (ThermoFisher, R37606). Confocal images were obtained using a Leica Stellaris 5 confocal microscope and an Opera Phenix high content imaging system. TTFs were stained with anti-H2A (Abcam, ab18255) and detected with anti-rabbit Alexa fluor 488 (ThermoFisher, A-21206). For H2B-mCherry transfer experiments, cells were stained with anti-mCherry (Abcam, ab167453) and detected with anti-rabbit Alexa fluor 488.

Immunogold labeling and EM

To reduce noise in immunogold labeling, VLPs were purified more extensively than in previous experiments. Supernatant from transfected cells was harvested, spun at 2000 *g* for 10 mins to remove cell debris, 0.45 um filtered, and spun at 150,000 *g* for 2 hours at 4°C through a 20% sucrose cushion. The pellet was resuspended in PBS and further purified over an 8-30% iodixanol gradient (Opti-prep, Sigma-Aldrich) with 2% steps spun at 250,000 *g* for 1.5 hours at 4°C. The gradient was fractionated in 0.5 mL fractions, desalted with centrifugal filters, and western blotted against HA to determine which fractions contained PEG10 (fig. S9A). Immunogold labeling protocol was derived from previous work (*10*), first, PEG10 positive fractions were fixed overnight in 2% PFA at 4°C. Samples were adhered to carbon coated nickel TEM grids (300 mesh, Ted Pella) for 20 mins at 4°C, washed in PBS, quenched with 50 mM glycine, permeabilized with 0.1% saponin for 20 mins, and blocked for 30 mins in 5% BSA-c (EMS-Aurion). Samples were stained in blocking buffer with a rabbit anti-PEG10 (Proteintech) for 1 hour at room temperature, washed, and stained with 6nm goat anti-rabbit gold antibody conjugates (EMS-Aurion). Samples were washed in PBS, stabilized in 1% glutaraldehyde, washed again in water, and negative stained with 2% uranyl acetate (EMS). Grids were visualized on a FEI Tecnai G2 Twin Spirit TEM equipped with a Gatan CCD camera at 120 kV at a magnification between 40,000-60,000X.

Cas9 and sgRNA guide cell line generation

A guide against mouse *Kras* was cloned using golden gate assembly under control of a U6 promoter into a custom lentiviral vector. Virus was produced using this genome as previously described and supernatant was added to N2a cells at 70% confluency. After 2 days, cells were split and selected for 7 days in Zeocin (Thermo Fisher). Cells expressed H2B-mCherry upon transduction, as verified visually. Cells were used for transfer experiments as described previously. For SpCas9 cell lines, N2a cells were transduced similarly, but with a genome

encoding an EF1a driven SpCas9 and blasticidin resistance. Cells were periodically maintained under selection and frozen at low passage.

Tail-tip fibroblast isolation

Tails were removed from culled female homozygous Ai9 mice with scissors and doused in 70% ethanol. The outer skin was removed with tweezers and scissors and the tail was minced in warm DMEM. Liberase TH (Roche) was added to a concentration of 200 ug/mL and the samples were incubated for 30 mins at 4°C. After washing, the tissue was placed on a 2% gelatin coated tissue culture plate in warm DMEM and fibroblasts were allowed to migrate out for 5 days before a media change and tissue removal. Media was changed every 2 days thereafter until cells reached confluence. Cells were split, frozen, and used for experiments at low passage to prevent senescence.

Supplementary Text

Supplementary Text 1

The full length MmPEG10 polyprotein encodes four conserved retroelement domains: CA, NC, PRO, and a catalytically inactive RT (*34*, *35*). Like some retroelements, the transcript is translated in two protein isoforms, namely, reading frame 1 (RF1, *gag*), which encodes only CA and NC, and reading frame 2 (RF2, *pol*), which is generated via a -1 ribosomal frameshift (*36*) and encodes all four domains (Fig. 1A). The MmPEG10 polyprotein is then processed in cis by the protease into individual proteins (Fig. 2E) as shown previously (*21*) and confirmed by western blotting. The band sizes on the western blot correspond to the sizes of the predicted functional domains of MmPEG10 , as confirmed by examination of the peptide signatures from these bands by mass spectroscopy (fig. S5B-D).

To characterize protein interactions that might be relevant for the functions of MmPEG10, we performed Co-IP mass spectroscopy on N and C-terminally HA-tagged *MmPeg10* that was transiently transfected into N2a cells (fig. S5, E and F) and found that MmPEG10 is associated with RNA processing proteins, such as MOV10 and GEMIN2 (fig. S5G).

Supplementary Text 2

We injected AAV-PHP.eB encoding a neuron specific expression of nuclear membrane GFP protein (GFP-KASH) and sgRNAs against *MmPeg10* into P1 SpCas9 mice (fig. S7A). After 24 days, we harvested and performed FACS for GFP+ nuclei (fig. S7B) from the frontal cortex of the *MmPeg10* knockout mice and non-targeting control mice. Next generation sequencing for indels in the *MmPeg10* locus confirmed robust gene knockout with 55-70% indels (fig. S7C). In cortical neurons, many genes were significantly downregulated and a few genes were upregulated in *MmPeg10* knockout nuclei compared to the control sorted nuclei (fig. S7D). Pathway analysis of these genes shows significant enrichment of downregulated genes responsible for cytoplasmic RNA processing, as well as pathways involved in neuron migration, homeostasis, and synaptic endocytosis (fig. S7E).

Supplementary Text 3

Immunogold TEM analysis of VLPs derived from HEK293FT cells demonstrates MmPEG10 VLPs are membrane bound and confirms their size to be around 100 nm (Fig. 3B). MmPEG10 particles fractionate alongside the exosomal marker TSG101 and contain visible capsid-like structures (fig. S9, A and B). To ensure that MmPEG10 VLPs transferred mRNA rather than protein, we blotted the VLPs for HA as well as Cre recombinase, calnexin, and CD81 (fig. S9C). The VLP fraction contained HA tagged MmPEG10 and CD81, but not Cre recombinase or calnexin, a marker of cell contamination. By contrast, Calnexin, Cre, and MmPEG10 were readily detectable in the whole cell lysate

Supplementary Text 4

While both UTRs of the *MmPeg10* mRNA are bound by the MmPEG10 protein, eCLIP data indicate that there is also binding inside the *MmPeg10* coding sequence (Fig. 2I). We hypothesized that this affinity led to *MmPeg10* mRNA packaging, thereby reducing the efficiency of cargoRNA transfer. We generated six mouse and five human *Peg10/PEG10* overexpression constructs, each with an overlapping 500-700 bp window of recoded codons to eliminate self mRNA binding (fig. S11A). For mouse, the most effective was rc4, referred to hereafter as *rMmPeg10*, and for human rc3, referred to hereafter as *rHsPeg10*. For VLPs produced with both rHs and rMmPeg10 there was an increase in the ratio of secreted cargo(Cre) relative to *MmPeg10*/*HsPEG10* 'self' RNA by 10-20 fold in the MNaAse treated VLP fraction without a corresponding change in vesicle secretion (fig. S11, B and C).

Supplementary Text 5

Publicly available tissue-wide sequencing databases show that *MmPeg10* is highly expressed in the placenta (fig. S13A) (*13*), which also expresses the fusogenic syncytin proteins MmSYNA and MmSYNB; these can be used to pseudotype lentivirus for efficient transgene delivery (*25*). Re-analyzing previous MmPEG10 eCLIP data in mouse trophoblast stem cells shows a direct interaction between MmPEG10 and the mouse syncytin transcript *MmSyna* (fig. S13B). Previously published single-cell sequencing databases of human syncytiotrophoblasts confirms that analogous endogenous fusogens (*ERVW-1 and ERVFRD-1*) are expressed in the same cells as *HsPEG10* (fig. S13C) (*37*).

Supplementary Text 6

To understand the effects of MmPEG10 VLPs on target cells, which may indicate a potential biological function for *MmPeg10*-mediated intercellular mRNA transfer, we treated primary mouse cortical neurons with MmSYNA pseudotyped MmPEG10 VLPs containing Mm.cargo(Peg10). After 72 hours we performed mRNA sequencing on the recipient neurons and identified a large number of key developmental genes that are differentially expressed (fig. S14A). For example, we detected upregulation of *MmShank1,* which we showed to be bound by PEG10 protein in our eCLIP data (fig. S6, F and G) and downregulated upon knockout of *MmPeg10* in the developing mouse brain (fig. S7,D and F). Pathway analysis suggests these upregulated genes are involved in neurodevelopment more broadly (fig. S14B).

Supplementary Text 7

We also investigated whether our SEND VLPs delivering alternative cargos like Cre would have substantial effects on target cells compared to delivering the native *MmPeg10* mRNA. This is particularly important since PEG10 has been implicated in cancer and can regulate cell growth and apoptosis (*38–41*). To test this, we produced VSVg pseudotyped VLPs containing either Mm.cargo(Peg10) or Mm.cargo(Cre) and treated loxP-GFP-N2a cells with equal amounts of the

VLP (as assayed by copy number of cargoRNA). 72 hours later we performed mRNA sequencing on these cells and naïve cells and found that cells treated with Mm.cargo(Peg10) VLPs had over 500 differentially expressed genes compared to naïve cells (fig. S15A), while those treated with Mm.cargo(Cre) had 42 (fig. S15, B and C). Cells treated with Mm.cargo(Peg10) VLPs upregulated genes involved in metabolic pathways (fig. S15D) and downregulated genes involved in translation, protein catabolism, and some biosynthetic pathways (fig. S15E). Cells treated with Mm.cargo(Cre) differentially expressed genes involved in vesicle exocytosis (fig. S15F). This suggests transferring a reprogrammed cargo does not have the same impact on recipient cells as transferring Peg10, although further studies are needed to explore the effect on recipient cells in vivo.

Fig. S1.

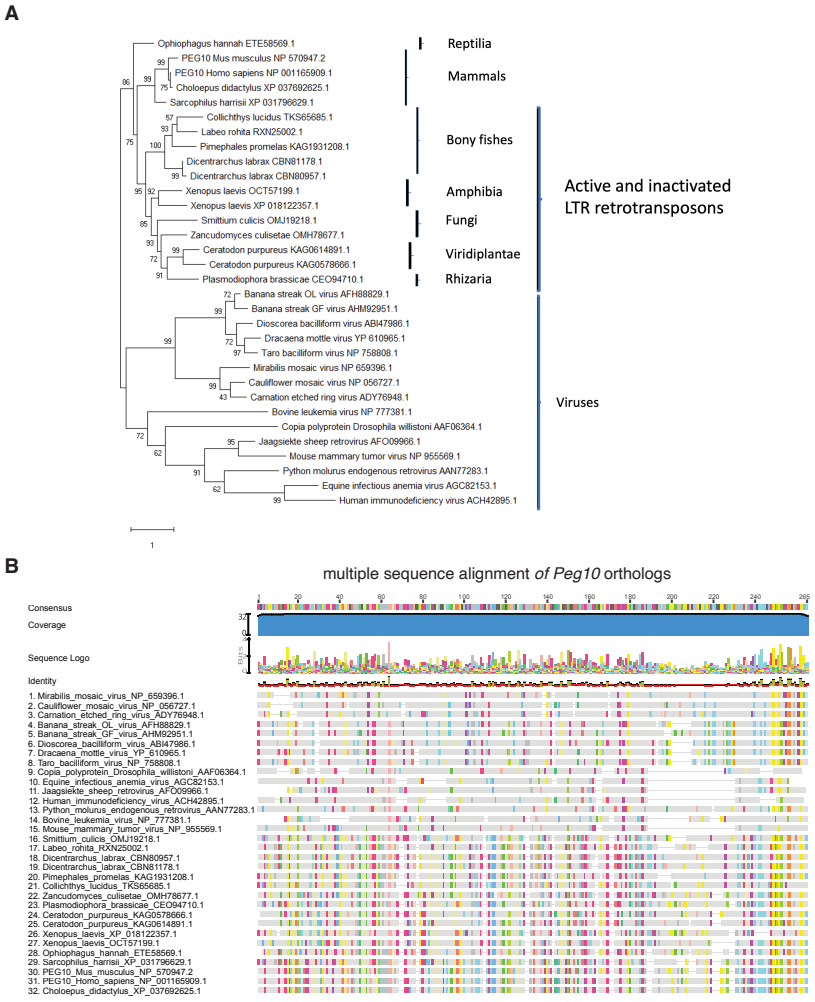

Fig. S1 Evolutionary provenance of PEG10.

A. Phylogenetic tree of the CA-domain of PEG10 and its homologs encoded by LTR retrotransposons and retroviruses. PEG10 homologs were identified using BLASTP against the non-redundant protein sequence database at the NCBI Sequences were selected to cover the diversity of eukaryotic lineages. Multiple alignment was made using Muscle PMID: 15034147 and trimmed manually to 265 positions (See Supplementary Data). The tree was constructed using FastTree 2 with default parameters (WAG evolutionary model, gamma-distributed site rates) PMID: 20224823. The sequences are denoted by their species of origin and their identifiers in the NCBI protein database. The numbers at forks indicate bootstrap support (percentage points) that was calculated by FastTree.

B. Multiple sequence alignment used to generate (A).

Fig. S2

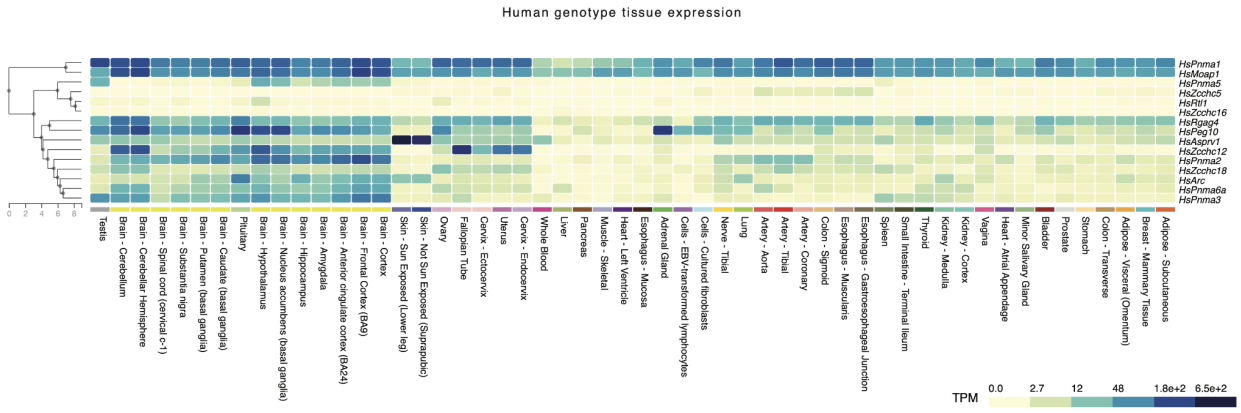

Fig. S2 Adult tissue expression of CA-containing proteins. Heatplot from the Broad Institute GTEx portal (gtexportal.org) of tissue-specific expression of CA-containing genes in human tissues.

Fig. S3

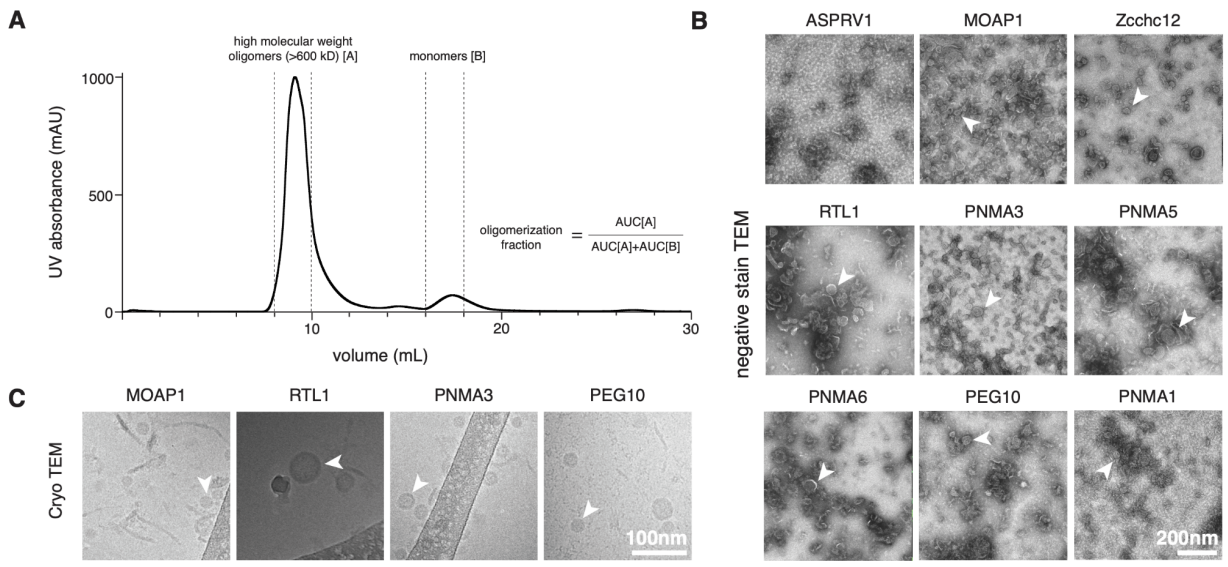

Fig. S3. Bacterially produced mouse orthologs of CA-containing proteins form oligomers/capsids as assayed by size exclusion and EM.

A. Representative chromatogram from size exclusion of MmPEG10 on a Superdex 200 Increase 10/300 GL. The first annotated peak represents high molecular weight oligomers in the void fraction of the column, while the second represents monomers. The fraction oligomerized was calculated by taking fractions of the area under the curve.
B. Widefield negative stain transmission electron micrographs (TEM) of bacterially produced CA-containing proteins. Arrows indicate capsids. Scale bar, 200 nm.
C. Additional electron micrographs from cryogenic transmission electron microscopy (cryo TEM) of bacterially produced CA-containing proteins. Arrows indicate capsids. Scale bar, 100 nm.

Fig. S4

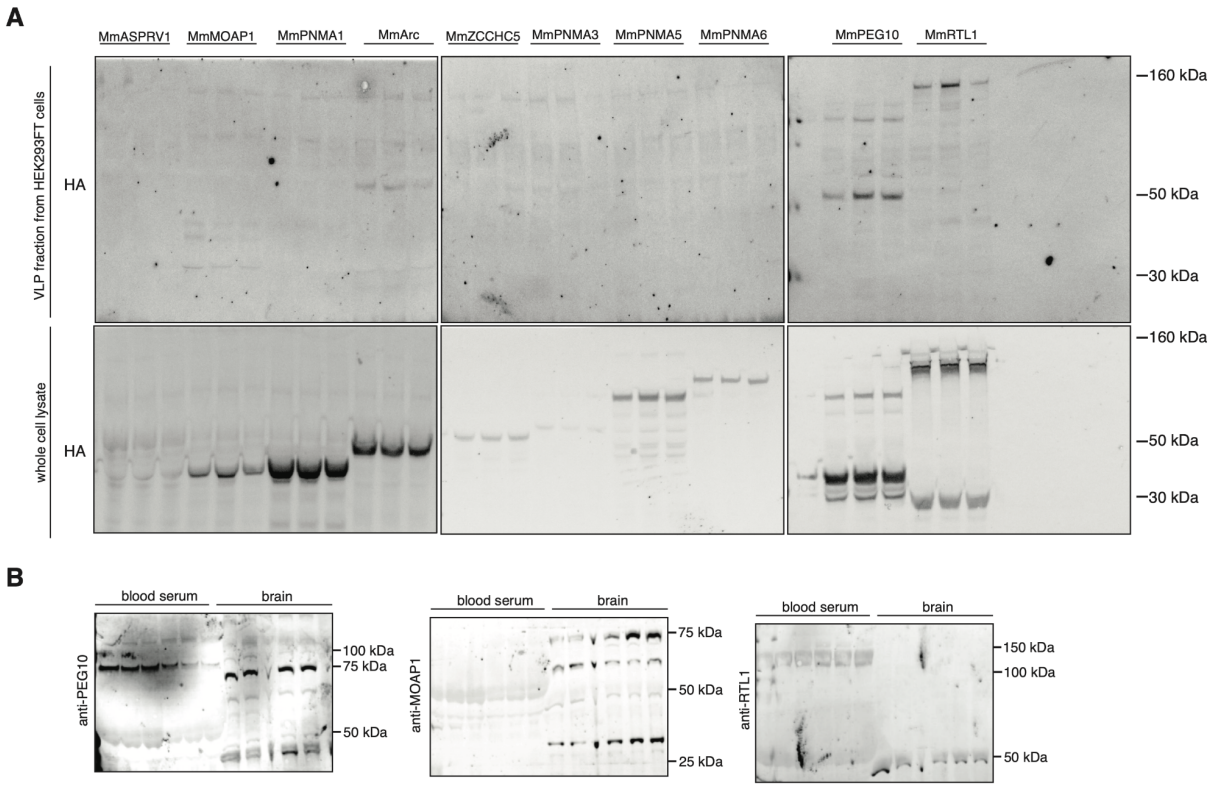

Fig. S4. Some CA-containing proteins are secreted into the supernatant and detectable in mouse tissues.

A. Whole-cell lysate and VLP fraction from HEK293FT cells in which each of the N-terminal HA-tagged CA-containing mouse proteins are overexpressed with transient plasmid transfection in n = 3 replicates (quantified in Fig. 1G and representative western blot with n = 1 replicate displayed in Fig. 1F).

B. Western blots for endogenous PEG10, RTL1, and MOAP1 from cell-free serum and blood of n = 6 adult female C57BL/6 mice.

Fig. S5

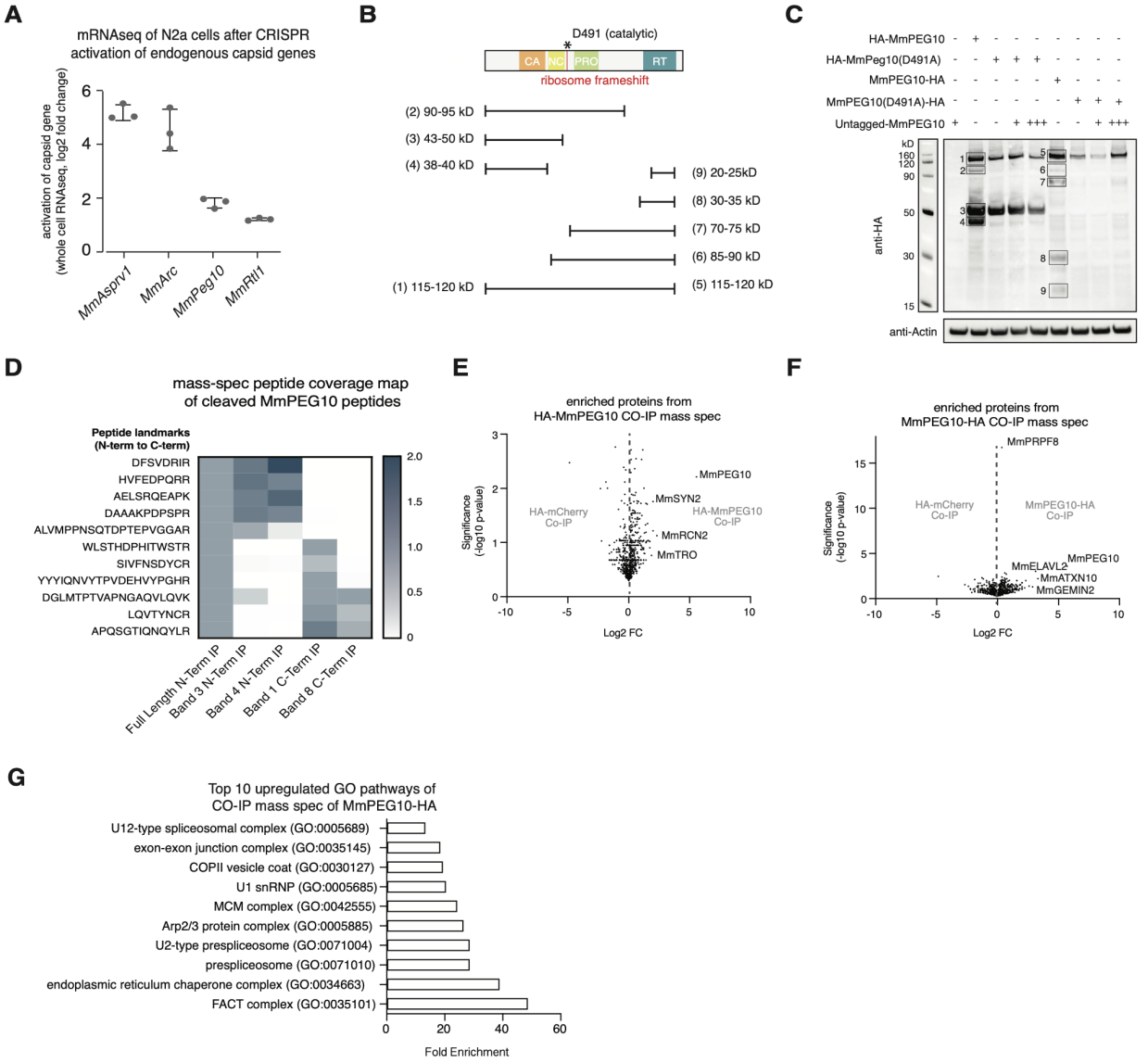

Fig. S5. MmPEG10 is cleaved by its own protease into its constitutive domains, which associate with a number of proteins inside the cell.

A. Log2 fold change in expression of each of the transcriptionally activated genes compared to cells with control non-targeting gRNAs, as determined by whole cell mRNA sequencing of n = 3 biological replicates.

B. The denoted bands from the western blot in (C) correspond to the annotated cleavage sites in this diagram which are approximately between the various domains of MmPEG10 including the CA, NC, PR, and RT.

C. Western blot for HA and actin of N-term and C-term HA tagged MmPEG10 with and without a protease mutation at D491A and with 3X molar excess (+++) of untagged MmPEG10 *in trans*.

D. Peptide landmarks from immunoprecipitation mass spectrometry of HA-tagged MmPEG10 corresponding to the approximate cleavage sites detailed in (C).
E. Log_2 fold change and significance of proteins that co-immunoprecipitated with N-term HA tagged MmPEG10 transfected into N2a cells. Log2 fold change represents the mean fold change across n = 3 replicates.
F. Log_2 fold change and significance of proteins that co-immunoprecipitated with C-term HA tagged MmPEG10 transfected into N2a cells. Log2 fold change represents the mean fold change across n = 3 replicates.
G. Top 10 gene ontology terms (GO) from significant enriched for proteins from the MmPEG10-HA Co-IP mass-spec results.

Fig. S6

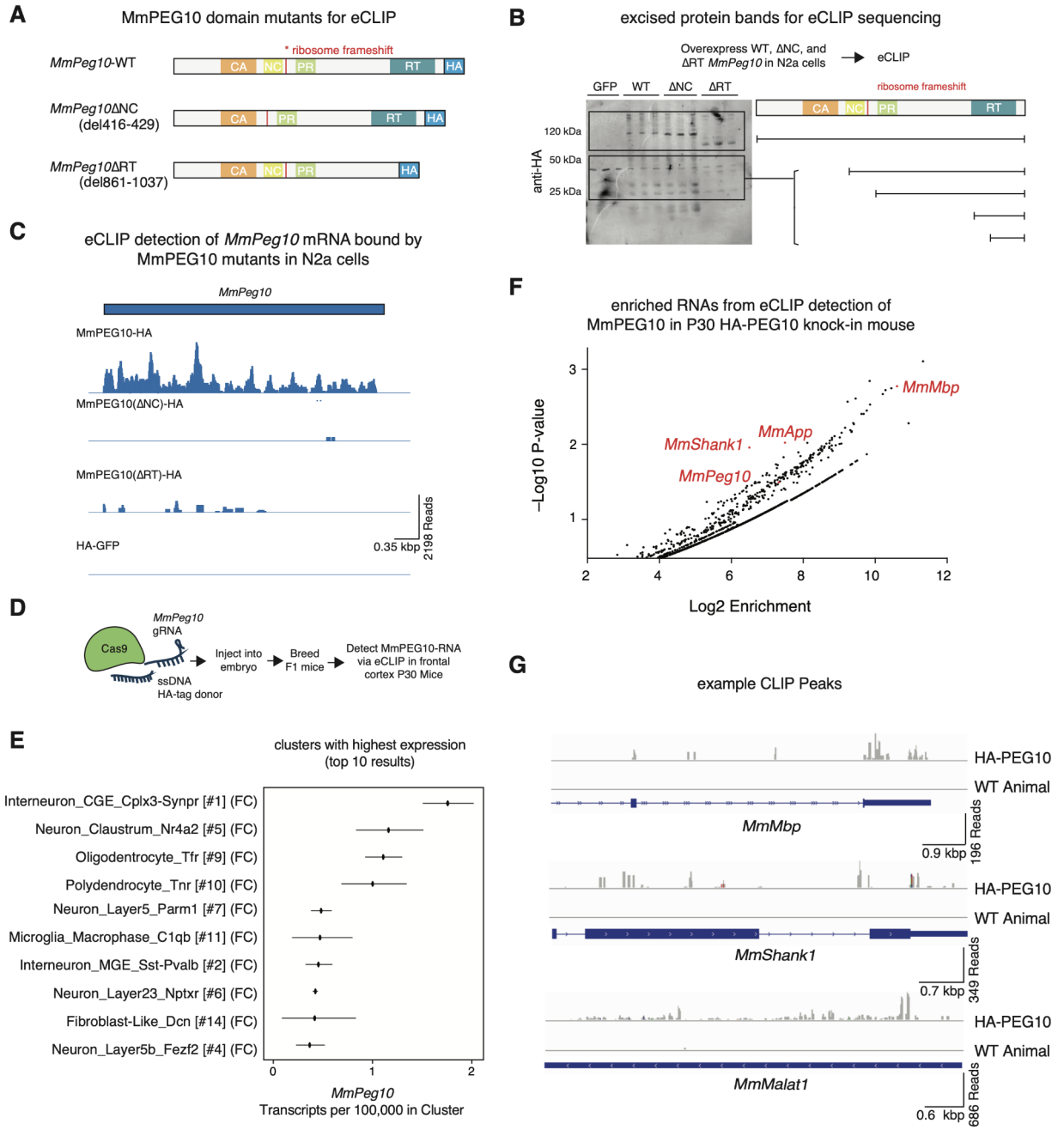

Fig. S6. MmPEG10 binding of mRNA is dependent on its nucleocapsid and reverse transcriptase domains.

A. Schematic outlining the strategy for determining the domains responsible for MmPEG10 nucleic acid binding in vitro by deleting each of the predicted nucleic acid binding domains.

B. Western blot against HA from each of the immunoprecipitated MmPEG10 mutants that were excised for eCLIP.
C. Sequencing alignment histogram at the *MmPeg10* locus from eCLIP of each of the MmPEG10 domain mutants.
D. Schematic for generating HA tagged MmPEG10 in embryonic mice to study MmPEG10 interactions in its native context in vivo.
E. Public gene expression data of *MmPeg10* in the mouse frontal cortex (*22*).
F. Plot of eCLIP results from UV cross linked immunoprecipitated HA tagged MmPEG10 in P30 frontal cortex. Fold enrichment is a comparison between n = 3 HA-tagged and n = 3 wild-type P27-P35 C57BL/6 mice.
G. Sequencing alignment histograms of *MmPeg10* bound mRNAs.

Fig. S7

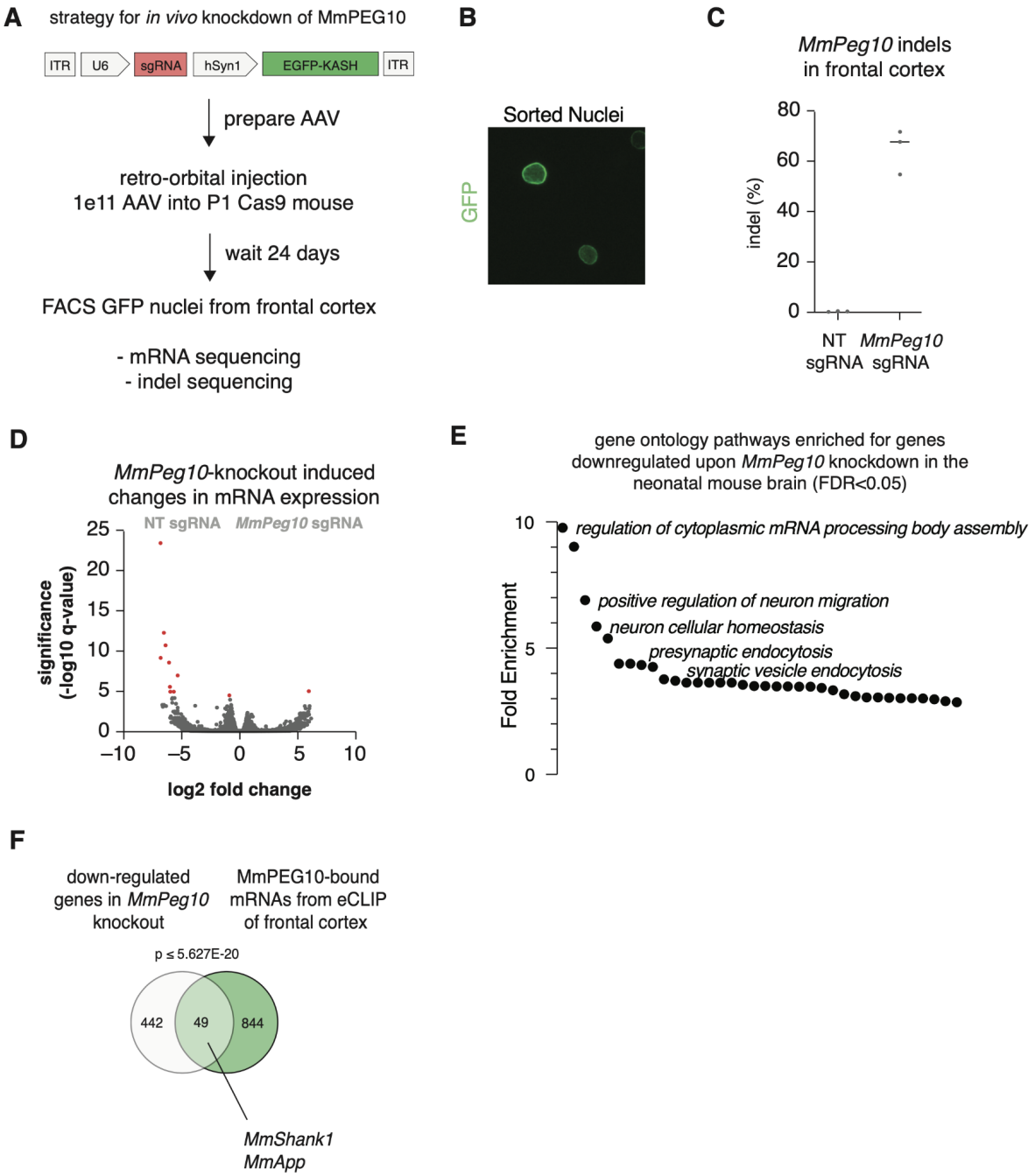

Fig. S7. Many genes are downregulated upon knockout of *MmPeg10* in neonatal mouse brains.

A. Schematic outlining approach for knocking down *MmPEG10* in the postnatal developing brain of SpCas9 mice using sgRNAs packaged by PHP.eb.

B. Representative image of GFP+ sorted neuronal nuclei from the cortex of P25 SpCas9 mice injected with PHP.eB carrying KASH-GFP under the hSyn1 promoter and sgRNAs against *MmPeg10*.
C. Indel rates at *MmPeg10* genomic locus from GFP+ sorted neuronal nuclei from n = 3 animals injected with AAV encoding *MmPeg10* targeting sgRNAs and n=e animals injected with AAVs encoding a non-targeting (NT) sgRNA.
D. Volcano plot of mRNA sequencing results from neuronal nuclei harvested from the frontal cortex of animals transduced with AAVs encoding *MmPeg10* targeting sgRNAs. Log2 fold change represents the mean fold expression between two cohorts of animals, each with of n = 3 injected mice
E. Gene ontology analysis of genes downregulated upon knock-down of MmPEG10 in the cortex of neonatal mice.
F. Venn diagram showing that, of the significant downregulated genes in *MmPeg10* knock-down neurons, 49 are significantly bound by MmPEG10 in the brain, demonstrated in eCLIP data. P-value represents significance of gene overlap, hypergeometric test.

Fig. S8.

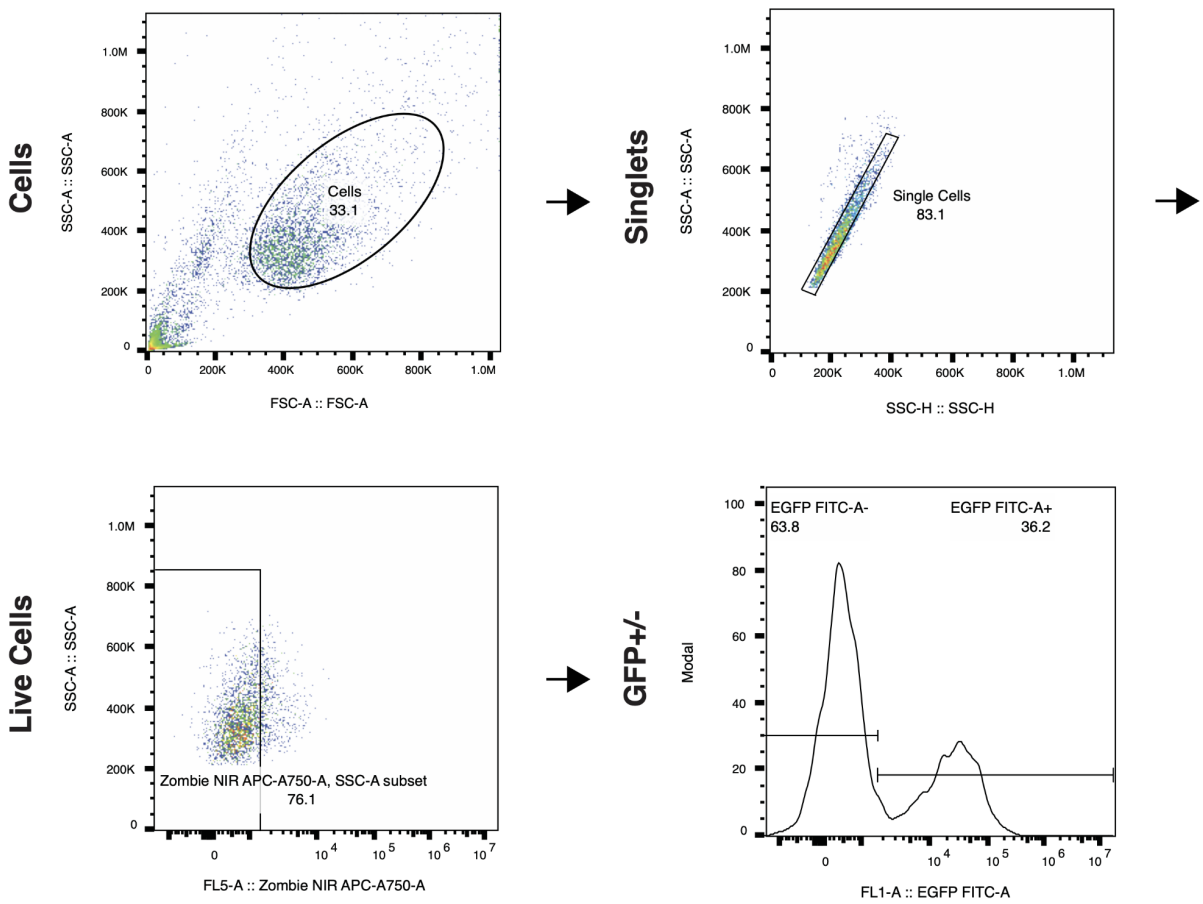

Fig. S8. Representative flow cytometry gating scheme for functional transfer experiments. Cells were first gated on FSC and SSC to remove debris. Following this singlets were gated on SSC, dead cells were removed by gating on the Zombie NIR live/dead stain. GFP+ cells were gated based on untreated controls.

Fig. S9

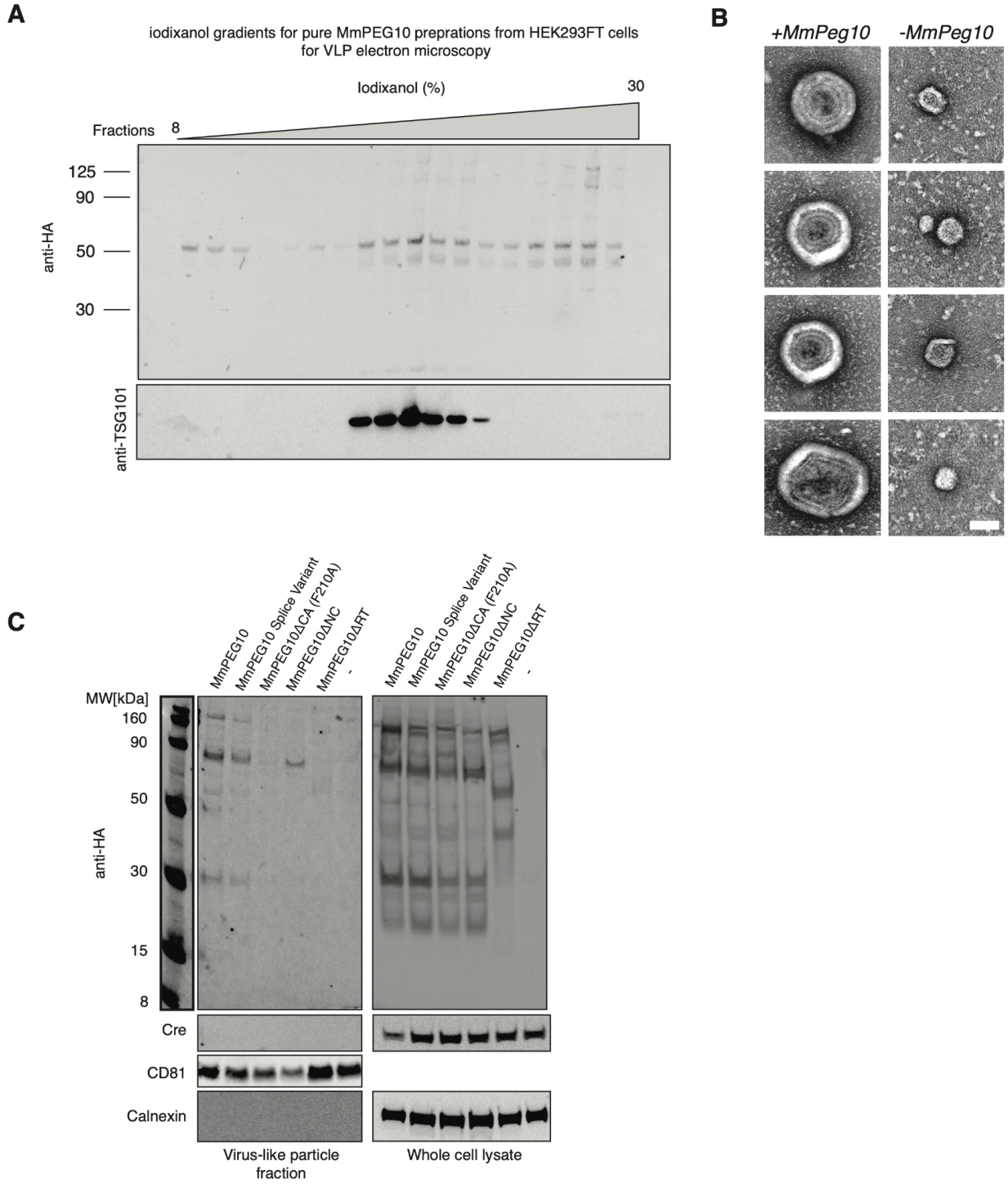

Fig. S9. MmPEG10 particles carry RNA cargo, not protein.

A. Western blot for HA and exosome marker TSG101 of MmPEG10 VLPs pelleted through a 20% sucrose cushion and then further purified using a 8-30% iodixanol step gradient for immunogold labeling and electron microscopy.

B. TEM micrographs of the VLP fraction derived from HEK293FT cells transfected with or without *MmPeg10,* both conditions included Mm.cargo(Cre) and VSVg. Scale bar represents 50 nm.
C. Western blot of MmPEG10 VLPs produced in HEK293FT cells with wild-type MmPEG10 and various domain mutants. Blots for HA show expression (right) and vesicular secretion of protein (left) but no Cre protein is present in the VLP fraction. CD81 is used as a loading control in the VLP fraction and calnexin as a loading control and marker of cell contamination.

Fig. S10

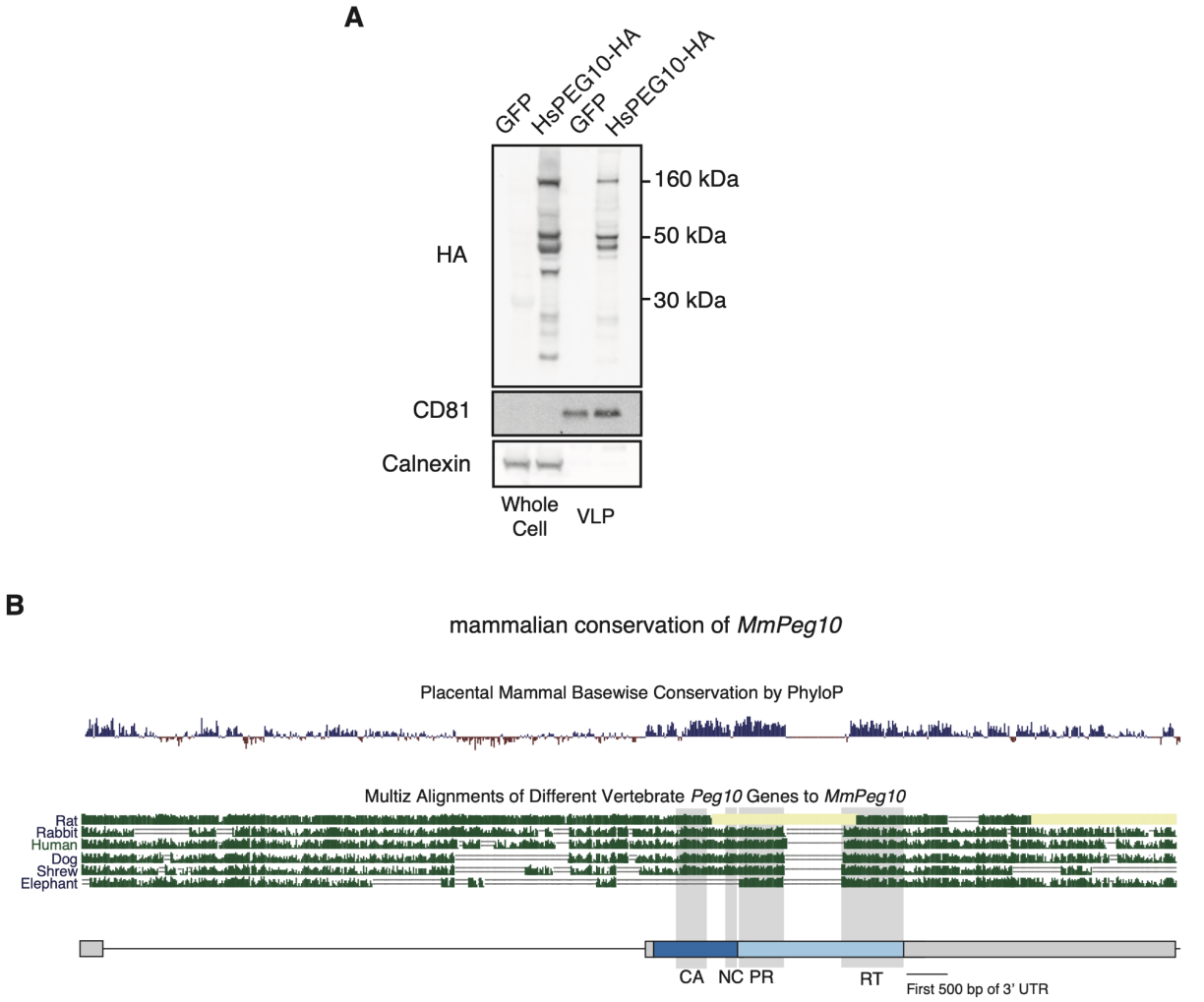

Fig. S10. HsPEG10 is also secreted and capable of mRNA transfer.

A. Western blot of the whole cell lysate and VLP fraction from HEK293FT cells transfected with *HsPEG10*. CD81 and calnexin shown as loading controls and markers of cell contamination, respectively.

B. Mammalian conservation map of the entire *MmPeg10* locus generated using UCSC genome browser.

Fig. S11

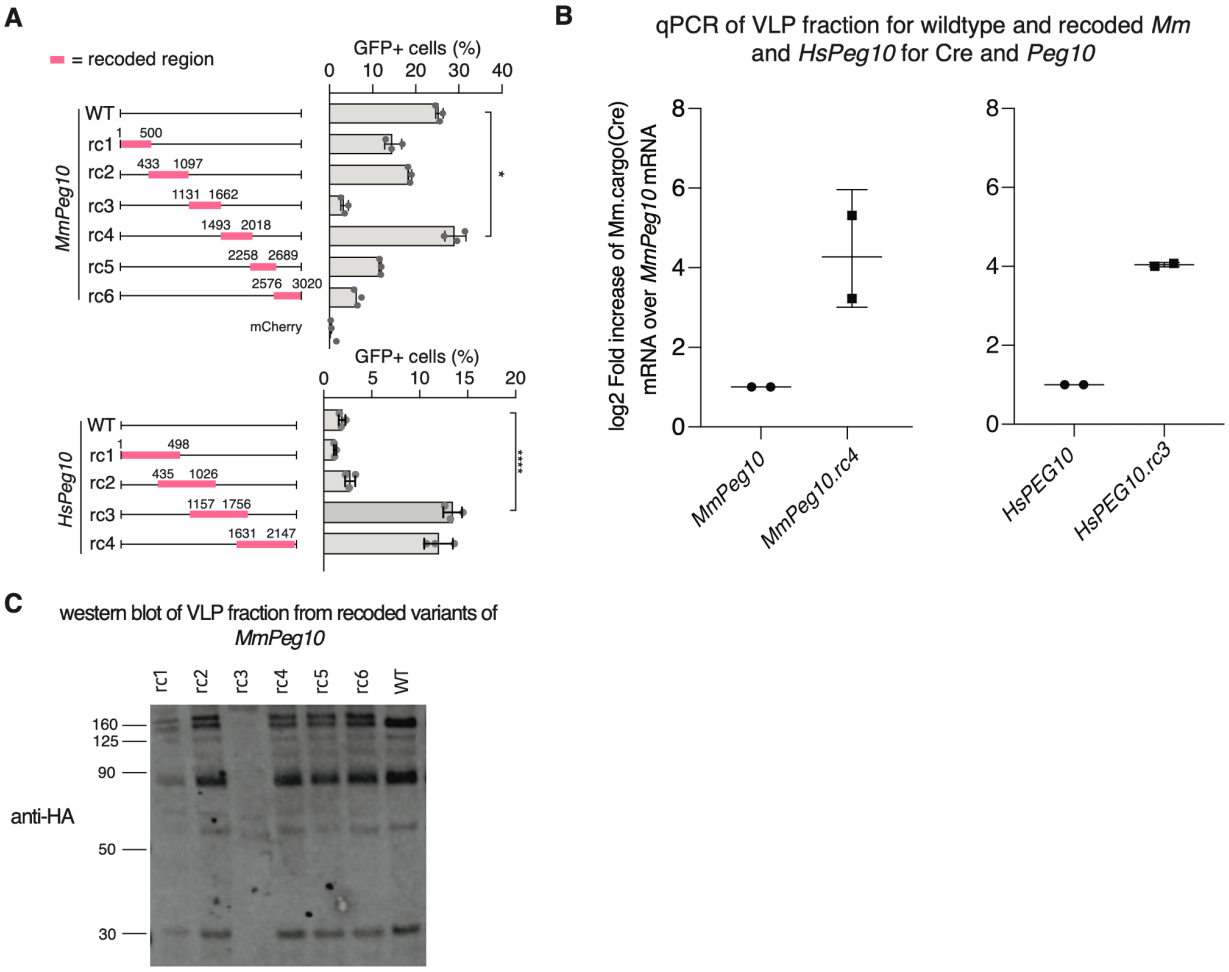

Fig. S11. *MmPeg10* and *HsPEG10* recoding boosts packaging and functional transfer of cargoRNA.

A. Functional transfer of Mm or Hs.cargo(Cre) using WT *MmPeg10*, *HsPEG10* or *Peg10/PEG10* with 500bp bins of recoded (rc) sequences. Data quantified by flow cytometry 72 hours after addition to loxP-GFP N2a reporter cells, n = 3 replicates per condition.

B. Log2 fold change (ΔΔCT) of Mm.cargo(Cre) over 'self' *Peg10/PEG10* in the VLP fraction of cells transfected with wild-type or recoded *MmPeg10* and *HsPeg10*. Means represent average fold change (ΔΔCT) from n=2 replicates. For all panels *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, One-Way ANOVA.

C. Western blot of the VLP fraction of recoded variants of *MmPeg10*.

Fig. S12.

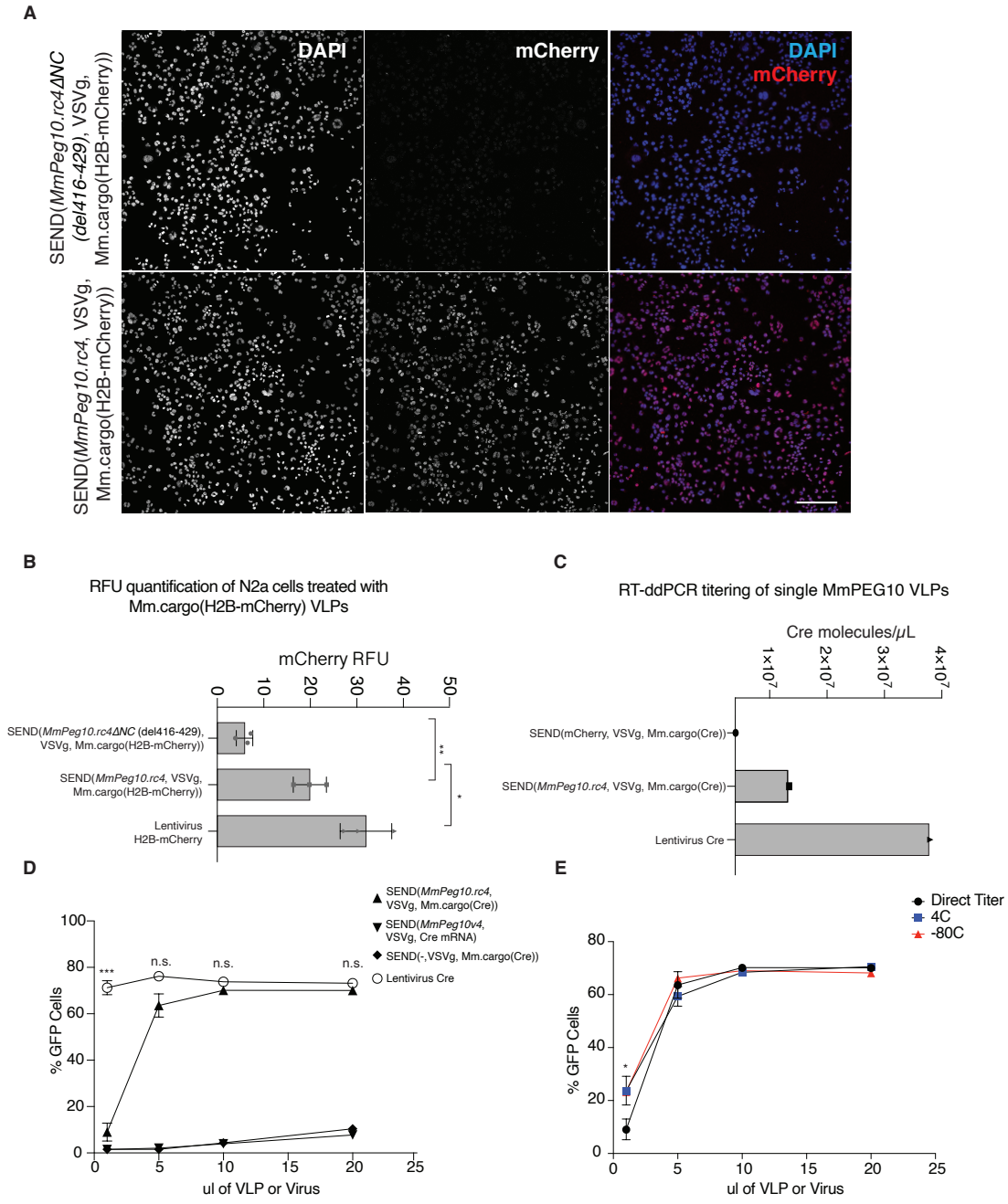

Fig. S12. Molecular and functional titration of SEND demonstrates it has a reduced titer compared to lentivirus.

A. Representative immunofluorescence images 72 hours following VLP delivery of Mm.cargo(H2B-mCherry) using SEND. Scale bar represents 200μm.

B. Quantification of mean fluorescence intensity across entire tile imaged wells from n=3 replicates 72 hours following lentiviral delivery of H2B-mCherry, rMmPEG10 mediated

delivery of Mm.cargo(H2B-mCherry), or rMmPEG10ΔNC mediated delivery of Mm.cargo(H2B-mCherry).

C. Digital droplet RT-PCR of equivalent volumes of MNAse treated VLP fractions showing RNA copy number of Cre mRNA in SEND VLPs versus lentivirus.
D. Titration of lentivirus and SEND delivering Cre on a volume per volume basis in loxP-GFP N2a cells. ***$p<0.001$, n.s. $p>0.05$.
E. Determination of the functional titer of SEND after overnight freezing at -80ºC or overnight storage at 4ºC. * $p<0.05$. For (D) and (E) data quantified by flow cytometry 72 hours after addition of SEND(*rMmPeg10*, VSVg, Mm.cargo(Cre)) or lentivirus. Two-Way ANOVA with Tukey correction, n=3 per condition per timepoint.

Fig. S13.

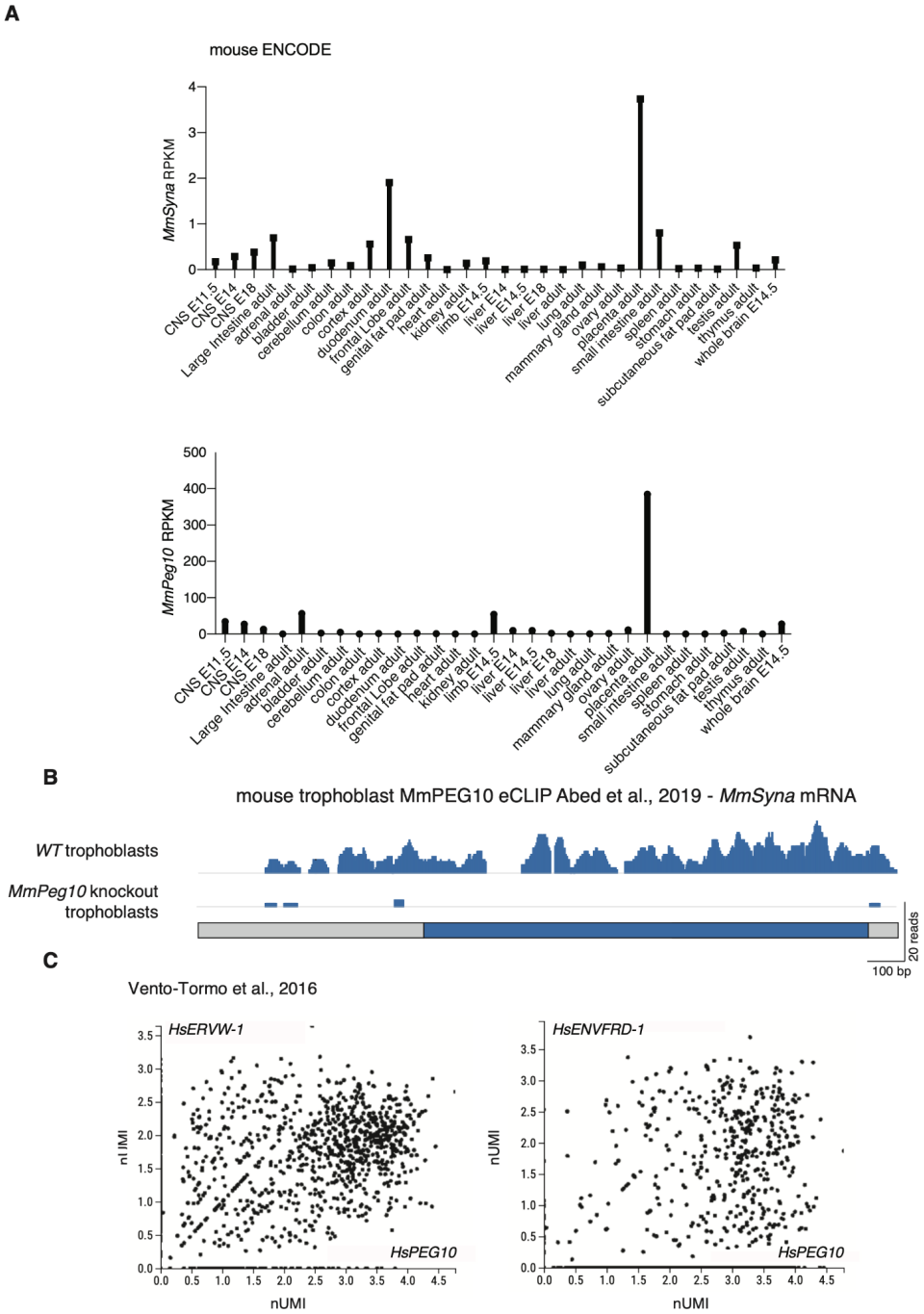

Fig. S13. Endogenous fusogens are co-expressed in *Peg10/PEG10* expressing cells.

A. ENCODE tissue wide mRNA sequencing data of *MmSyna* and *MmPeg10* across multiple mouse tissues.

B. Sequencing alignment histogram of *MmSynA* RNA from re-analyzed MmPEG10 eCLIP in trophoblast stem cells in vitro *(13)*.

C. Single cell sequencing scatter plots of human placental cells of *HsPEG10* co-expression with the endogenous fusogens *HsERVW-1* and *HsENVFRD-1* (*37*).

Fig. S14

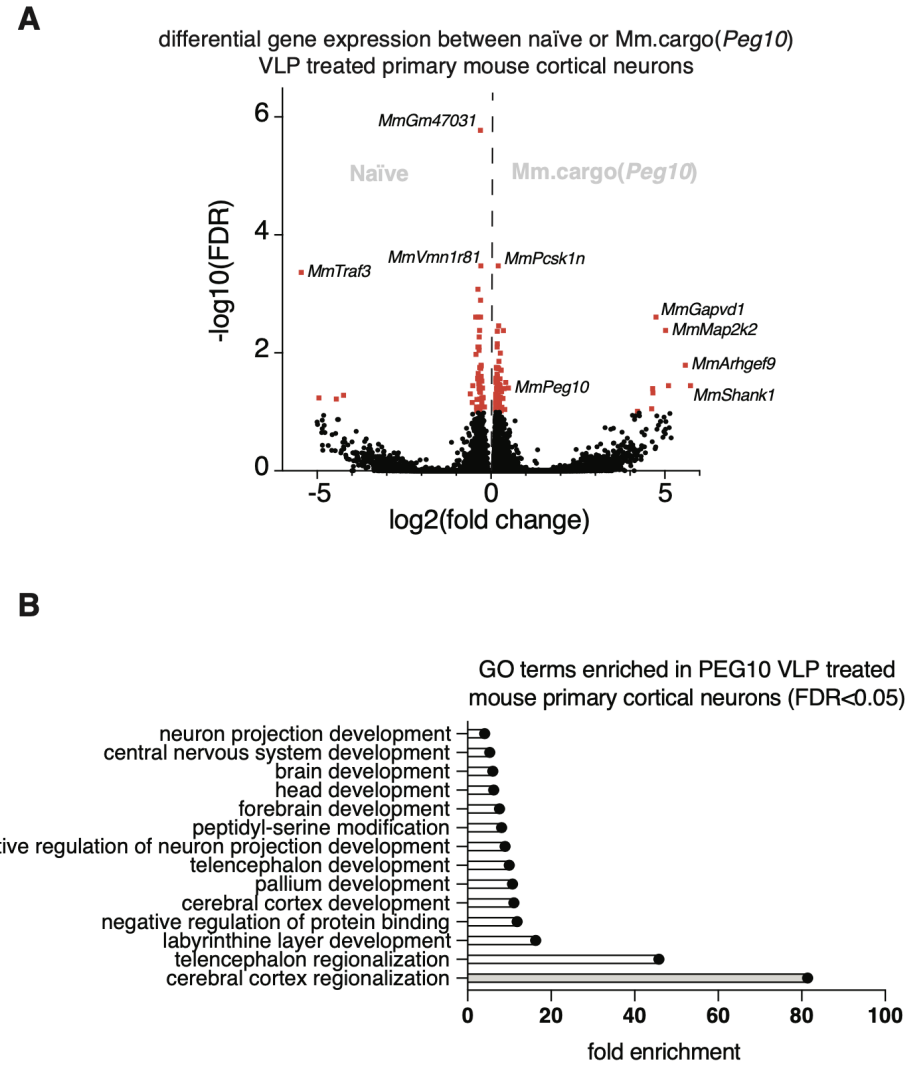

Fig. S14. *MmPeg10* VLPs carrying *MmPeg10* mRNA and pseudotyped with MmSynA induce transcriptional changes in primary mouse cortical neurons.

A. Volcano plot of mRNA transcriptome changes compared to naive cells after MmSYNA pseudotyped SEND VLPs containing cargo(Peg10) were transferred onto primary mouse cortical neurons for 72 hours. Data show $\log_2$ fold change and $-\log_{10}$(FDR) of n = 3 replicates. Red dots indicate FDR≤0.1.

B. GO pathway enrichment in neurons treated with cargo*Peg10* VLPs compared to naïve cells.

Fig. S15

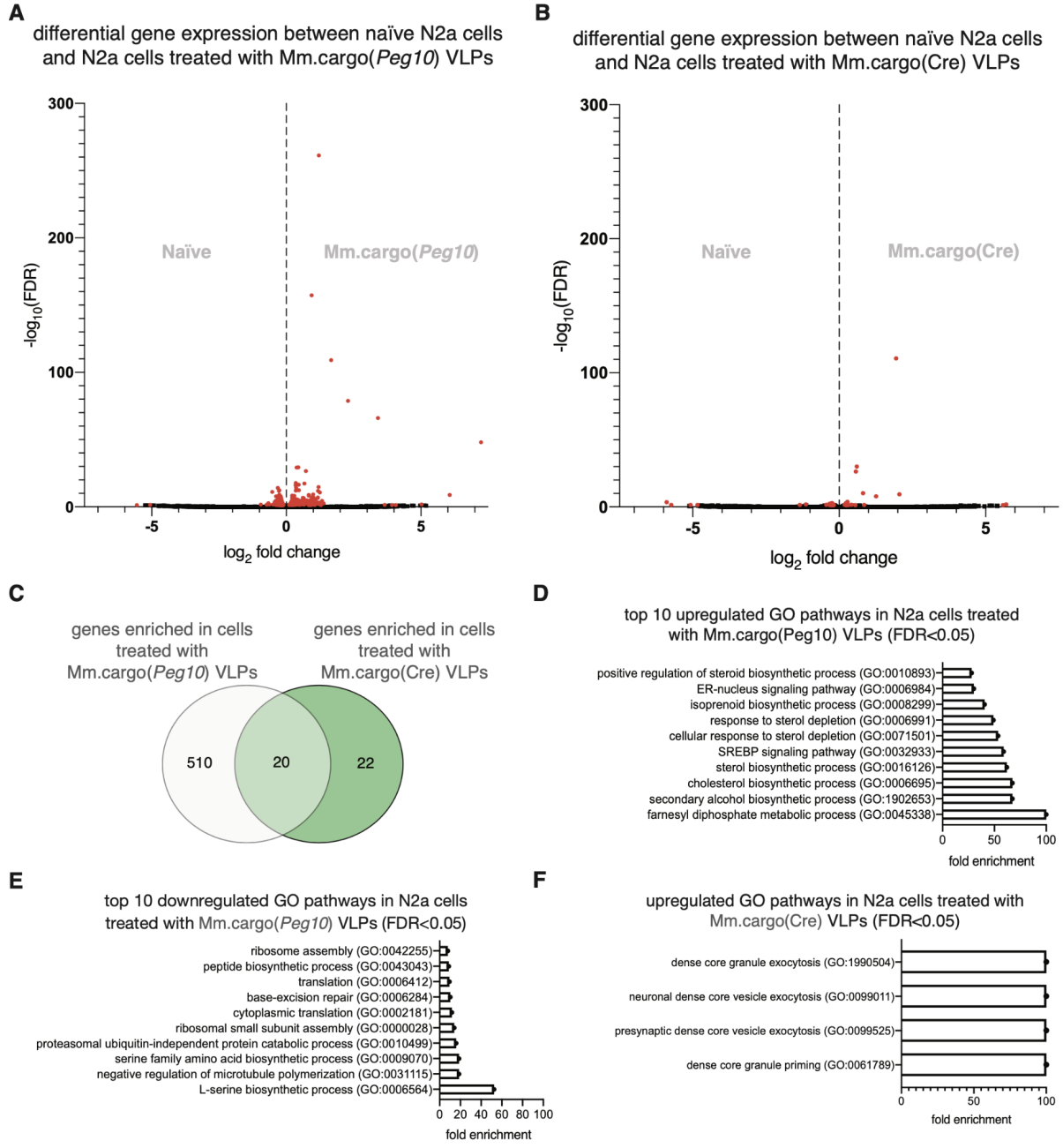

Fig. S15. Mm.cargo(Peg10) induces significant transcriptional changes in recipient cells.

A. Log_2 fold change and $-log_{10}$(FDR) of differentially expressed genes between naïve N2a cells and those treated with SEND(rMmPeg10, VSVg, Mm.cargo(Peg10)), red indicates genes with FDR<0.05 (Benjamini-Hochberg correction) n = 3 per condition.

B. Log_2 fold change and $-log_{10}$(FDR) of differentially expressed genes between naïve N2a cells and those treated with SEND(rMmPeg10, VSVg, Mm.cargo(Cre)), red indicates genes with FDR<0.05 (Benjamini-Hochberg correction) n = 3 per condition.

C. Venn diagram showing number of genes differentially expressed in N2a cells treated with SEND(rMmPeg10, VSVg, Mm.cargo(Peg10)) or SEND(rMmPeg10, VSVg, Mm.cargo(Cre)) shows 20 overlapping genes between the two conditions. This overlap is significant with $p<2.28e-32$, hypergeometric test.
D. Fold enrichment of top 10 gene ontology pathways (FDR<0.05) of genes enriched upon Mm.cargo(Peg10) VLP treatment (A).
E. Fold enrichment of top 10 gene ontology pathways (FDR<0.05) of genes depleted upon Mm.cargo(Peg10) VLP treatment (A).
F. Fold enrichment of top 10 gene ontology pathways (FDR<0.05) of genes enriched upon Mm.cargo(Cre) VLP treatment (B).

Fig. S16.

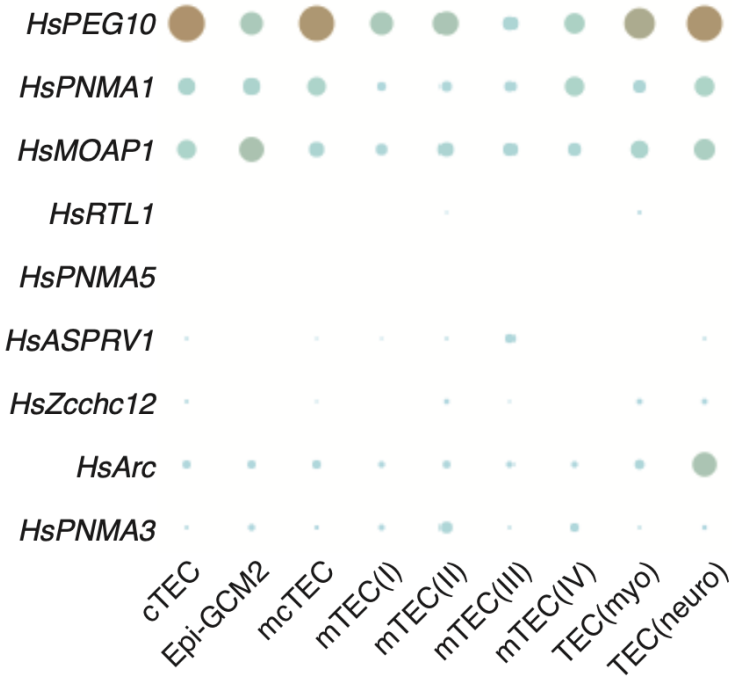

Fig. S16. *HsPEG10* is highly expressed in the human developing thymus compared to other endogenous CA-containing proteins. Dot plot showing expression of endogenous CA-containing genes across several epithelial cell types in the human developing thymus. Dot size represents relative expression level. *HsPeg10* is the most highly expressed endogenous CA-containing gene in many thymic epithelial cell types. Data were generated from the Human Cell Atlas Developmental web portal using data derived from (*33*). Plot was generated from the human fetal thymus epithelium dataset using the interactive_heatmap_dotplot tool developed by Dorin-Mirel Popescu.

Table S1. Summary of orthologous groups of proteins with domains homologous to Gag capsid protein in human and mouse based on sequence alignments. Most orthologous groups with proteins from both species include one human sequence and a few mouse sequences, in most cases these mouse sequences in one group correspond to the same gene but differ mostly by truncation or point mutations. In two groups, namely with human proteins NP_776159.1 and NP_001299820.1 the mouse proteins are clearly paralogous, but are practically equally close to human group members. Overall there are 19 orthologous groups with proteins from both species, two of which are with obvious mouse paralogs, and 15 human and 48 mouse singletons.

Candidates (*Homo sapiens*)			Orthologous candidates (*Mus musculus*)		
Identified protein	Pfam domain architecture	Accession	Identified protein	Pfam domain architecture	Accession
endogenous retrovirus group K member 5 Gag polyprotein	Retroviral GAG p10 protein; gag gene protein p24 (core nucleocapsid protein); Zinc knuckle; GAG-polyprotein viral zinc-finger	XP_011511643.1	-		-
endogenous retrovirus group K member 7 Gag polyprotein	Retroviral GAG p10 protein; gag gene protein p24 (core nucleocapsid protein); Zinc knuckle; GAG-polyprotein viral zinc-finger	XP_016863109.1	-		-
uncharacterized protein LOC107985332	Retroviral GAG p10 protein; gag gene protein p24 (core nucleocapsid protein); Zinc knuckle	XP_016883063.1	-		-
paraneoplastic antigen-like protein 5	PNMA	NP_001096620.1	-		-
retrotransposon-derived protein PEG10 isoform 3	Domain of unknown function (DUF4939); GAG-polyprotein viral zinc-finger; Retroviral aspartyl protease	NP_001165908.1	Retrotransposon-derived protein PEG10		sp|Q7TN75
retrotransposon-derived protein PEG10 isoform 2		NP_001035242.1	retrotransposon-derived protein PEG10 isoform 2	Domain of unknown function (DUF4939); GAG-polyprotein viral zinc-finger; Retroviral aspartyl protease	NP_001035701.1
retrotransposon-derived protein PEG10 isoform 4		NP_001165909.1	retrotransposon-derived protein PEG10 isoform 1		NP_570947.2
retrotransposon-derived protein PEG10 isoform 5		NP_001171890.1			
retrotransposon-derived protein PEG10 isoform 6		NP_001171891.1			
retrotransposon-derived protein PEG10 isoform 1		NP_055883.2			

retrotransposon Gag-like protein 8B	Domain of unknown function (DUF4939)	NP_001071639.1	-		-
retrotransposon Gag-like protein 8B	Domain of unknown function (DUF4939)	NP_001071641.1	-		-
retrotransposon Gag-like protein 8A isoform 1	Domain of unknown function (DUF4939)	NP_001071640.1	-		-
paraneoplastic antigen Ma3 isoform 2 [Homo sapiens]	PNMA	NP_001269464.1	unnamed protein product, partial	PNMA; Zinc knuckle	BAE24735.1
paraneoplastic antigen Ma3 isoform 1 [Homo sapiens]		NP_037496.4	paraneoplastic antigen Ma3 homolog		NP_694809.1
paraneoplastic antigen-like protein 6A	PNMA	NP_116271.3	mCG1032934	PNMA	EDL29902.1
paraneoplastic antigen-like protein 5	PNMA	NP_443158.1	-		-
paraneoplastic antigen Ma1	PNMA	NP_006020.4	paraneoplastic antigen Ma1 homolog	PNMA	NP_081714.2
modulator of apoptosis 1	PNMA	NP_071434.2	modulator of apoptosis 1	PNMA	NP_071718.1
			modulator of apoptosis 1		NP_001136409.1
paraneoplastic antigen Ma2	PNMA	NP_009188.1	mKIAA0883 protein, partial		BAD90245.1
paraneoplastic antigen Ma2 isoform X1		XP_011542667.1	paraneoplastic antigen MA2, isoform CRA_a, partial		EDL35986.1
			paraneoplastic antigen MA2, isoform CRA_b	PNMA	EDL35987.1
			paraneoplastic antigen Ma2 homolog		NP_780707.1
			paraneoplastic antigen Ma2 homolog isoform X1		XP_006519052.1
			paraneoplastic antigen Ma2 homolog isoform X1		XP_006519053.1
paraneoplastic antigen-like protein 8B	PNMA	NP_065760.1	PNMA-like protein 2	PNMA	NP_001093106.1
protein Bop	Domain of unknown function (DUF4939)	NP_078903.3	-		-

retrotransposon Gag-like protein 3	Domain of unknown function (DUF4939); Zinc knuckle	NP_689907.1	retrotransposon Gag-like protein 3	Domain of unknown function (DUF4939); Zinc knuckle	NP_955762.1
retrotransposon Gag-like protein 5	Domain of unknown function (DUF4939)	NP_001019626.1	retrotransposon Gag-like protein 5	Domain of unknown function (DUF4939)	NP_001265463.1
			mKIAA2001 protein, partial		BAD90267.1
retrotransposon-like protein 1	Domain of unknown function (DUF4939); RNase H-like domain found in reverse transcriptase	NP_001128360.1	-		-
retrotransposon Gag-like protein 6	Domain of unknown function (DUF4939)	NP_115663.2	retrotransposon Gag-like protein 6	Domain of unknown function (DUF4939)	NP_808298.2
protein LDOC1	Domain of unknown function (DUF4939)	NP_036449.1	-		-
zinc finger CCHC domain-containing protein 12	PNMA	NP_001299820.1	zinc finger CCHC domain-containing protein 12		NP_001345405.1
zinc finger CCHC domain-containing protein 12		NP_776159.1	zinc finger CCHC domain-containing protein 12		NP_001345406.1
			zinc finger CCHC domain-containing protein 12		NP_001345407.1
			zinc finger CCHC domain-containing protein 12		NP_001345408.1
			zinc finger CCHC domain-containing protein 12		NP_001345409.1
			zinc finger CCHC domain-containing protein 12 isoform X1		XP_006541462.1
			zinc finger CCHC domain-containing protein 12 isoform X1	PNMA	XP_011249276.1
			zinc finger CCHC domain-containing protein 12 isoform X1		XP_017174119.1
			zinc finger CCHC domain-containing protein 12 isoform X1		XP_036017996.1

			zinc finger CCHC domain-containing protein 12 isoform X1		XP_036017997.1
			zinc finger CCHC domain-containing protein 12 isoform X1		XP_036017998.1
			zinc finger CCHC domain-containing protein 12		NP_082601.1
zinc finger CCHC domain-containing protein 18 isoform X1	PNMA; Zf-CCHC	XP_011529314.1	zinc finger, CCHC domain containing 18, isoform CRA_a, partial		EDL23888.1
zinc finger CCHC domain-containing protein 18		NP_001137450.1	unnamed protein product		BAB23950.1
			zinc finger CCHC domain-containing protein 18		NP_001030586.1
			zinc finger CCHC domain-containing protein 18		NP_001030587.1
			zinc finger CCHC domain-containing protein 18		NP_001345366.1
			zinc finger CCHC domain-containing protein 18	PNMA; Zf-CCHC	NP_001345368.1
			zinc finger CCHC domain-containing protein 18		NP_001345369.1
			zinc finger CCHC domain-containing protein 18		NP_001345370.1
			zinc finger CCHC domain-containing protein 18		NP_080169.2
			zinc finger CCHC domain-containing protein 18 isoform X1		XP_030107344.1
			zinc finger CCHC domain-containing protein 18 isoform X1		XP_030107345.1
retrotransposon Gag-like protein 9	Retrotransposon gag protein domain	NP_065820.1	retrotransposon Gag-like protein 9 isoform X1		XP_011246109.1

			retrotransposon Gag-like protein 9 isoform X1	Retrotransposon gag protein domain	XP_011246110.1
			retrotransposon Gag-like protein 9		NP_001035524.2
activity-regulated cytoskeleton-associated protein	Arc C-lobe	NP_056008.1	activity-regulated cytoskeleton-associated protein		NP_001263613.1
			activity-regulated cytoskeleton-associated protein	Arc C-lobe	NP_061260.1
retrotransposon Gag-like protein 4	Domain of unknown function (DUF4939); Zinc knuckle	NP_001004308.2	-		-
endogenous retrovirus group K member 8 Gag polyprotein-like	Retroviral GAG p10 protein	XP_011526763.1	-		-
paraneoplastic antigen-like protein 5	PNMA	NP_001171853.1	paraneoplastic antigen-like protein 5		NP_001093931.1
paraneoplastic antigen-like protein 5		NP_001096620.1			
paraneoplastic antigen-like protein 5		NP_001096621.1		PNMA	
paraneoplastic antigen-like protein 5		XP_016884741.1			
paraneoplastic antigen-like protein 5		XP_016884742.1			
paraneoplastic antigen-like protein 5		NP_443158.1			
retroviral-like aspartic protease 1	PNMA; gag-polyprotein putative aspartyl protease	NP_690005.2	Asprv1 protein, partial	PNMA; gag-polyprotein putative aspartyl protease	AAH57938.1
paraneoplastic antigen Ma6F	PNMA	NP_001341909.1	mCG1032934	PNMA	EDL29902.1
natural cytotoxicity triggering receptor 3 ligand 1 precursor	Immunoglobulin V-set domain; Immunoglobulin C1-set domain; Matrix protein (MA), p15	NP_001189368.1			
natural cytotoxicity triggering receptor 3 ligand 1 isoform X1		XP_011518374.1			
natural cytotoxicity triggering receptor 3 ligand 1 isoform X1		XP_011518375.1			

natural cytotoxicity triggering receptor 3 ligand 1 isoform X1		XP_011518376.1			
natural cytotoxicity triggering receptor 3 ligand 1 isoform X1		XP_011518377.1			

Table S2. Summary of additional proteins with domains homologous to Gag capsid protein in mouse based on sequence alignments. In addition to those listed in Table S1., 48 additional proteins with domains homologous to Gag capsid protein are identified in mouse.

Candidates (*Mus musculus*)		
Identified protein	Domain	Accession
protein LDOC1	DUF4939	NP_001018097.1
PREDICTED: agouti-signaling protein isoform X1	Gag_p24	XP_011237991.1
gag protein	Gag_p24	AAC12789.1
Gag	Gag_p24	AAC52922.1
BC005685 protein, partial	Gag_p24	AAH05685.1
unnamed protein product	Gag_p24	BAC38137.1
gag	Gag_p24	BAC79170.1
gag	Gag_p24	BAF81988.1
TPA_exp: gag protein	Gag_p24	DAA01924.1
TPA_exp: gag protein	Gag_p24	DAA01925.1
TPA_exp: gag protein	Gag_p24	DAA01928.1
mCG142377, partial	Gag_p24	EDL00544.1
PREDICTED: uncharacterized protein LOC108167332	Gag_p24	XP_011239845.1
IgE-binding protein	Gag_p24	sp|P03975.1|IGEB_MOUSE
mCG1044120, partial	Gag_p24	EDL07694.1
PREDICTED: endogenous retrovirus group K member 24	Gag_p24	XP_011245081.1
PREDICTED: endogenous retrovirus group K member 8 Gag polyprotein-like	Gag_p24	XP_017167946.1
gag-myb protein, partial	Gag_p30	AAA39784.1
putative	Gag_p30	AAA51041.1
Gag-Pol polyprotein	Gag_p30	AAB06450.1
gag protein	Gag_p30	AAN46638.1
truncated polyprotein	Gag_p30	AAY27069.1
gag polyprotein pr65	Gag_p30	ABD14432.1
gag-pro-pol polyprotein	Gag_p30	ABD14433.1
gag polyprotein pr65	Gag_p30	ABD14435.1
gag-pro-pol polyprotein	Gag_p30	ABD14436.1

glyco-gag polyprotein	Gag_p30	AID54952.1
gag polyprotein	Gag_p30	AID54953.1
gag-pro-pol polyprotein	Gag_p30	AID54954.1
gag, partial	Gag_p30	AMK48512.1
putative gag-pro-pol polyprotein	Gag_p30	ARB03507.1
unnamed protein product	Gag_p30	BAC41106.1
unnamed protein product	Gag_p30	BAC41107.1
truncated gag-pro-pol polyprotein	Gag_p30	CCD57102.1
gag-pro-pol polyprotein	Gag_p30	CCD57104.1
gag protein	Gag_p30	CCD57105.1
mCG144922, isoform CRA_b, partial	Gag_p30	EDL00999.1
LOC72520 protein, partial	Gag_p30	AAH21868.1
BC040756 protein, partial	Gag_p30	AAH40756.1
LOC72520 protein, partial	Gag_p30	AAH44668.2
PREDICTED: uncharacterized protein LOC108167440 isoform X1	Gag_p30	XP_017167935.1
PREDICTED: uncharacterized protein LOC108167440 isoform X2	Gag_p30	XP_017167936.1
PREDICTED: uncharacterized protein LOC108167440 isoform X3	Gag_p30	XP_017167937.1
unnamed protein product, partial	PNMA	BAC37719.1
mCG1050067, isoform CRA_a	PNMA	EDL42061.1
coiled-coil domain-containing protein 8 homolog	PNMA	NP_001095005.1
predicted gene, 42372	PNMA	NP_001357780.1
PREDICTED: paraneoplastic antigen Ma2 homolog	PNMA	XP_011249051.1

Table S3. List of reagents used in this study.

Antibodies		
Items	Vendor	Dilution (for WB)
anti HA (mouse)	Cell Signalling, #2367	1:1000
anti HA (rabbit)	Abcam, ab910	1:1000
anti CD81	Santa Cruz, sc-166029	1:500
anti beta-Actin	Abcam, ab8227	1:5000
anti Moap1	Sigma, SAB1411249	1:1000
anti Peg10	Proteintech, A4412	1:1000, 1:25 for immunogold
anti Rtl1	ThermoFisher, PA566887	1:1000
IRDye 680RD Donkey anti-Mouse IgG (H+L)	LI-COR 926-68072	1:10,000
IRDye 680RD Donkey anti-Rabbit IgG (H+L)	LI-COR 926-68073	1:10,000
IRDye® 800CW Goat anti-Rabbit IgG Secondary Antibody	LI-COR 926-32211	1:10,000
Goat anti-rabbit 6 nm gold conjugate	Electron Microscopy Sciences 25103	1:40
Plasmids		
Items	Source	Application
PB-UniSAM	Addgene, 99866	CRISPR A
pUCmini-iCAP-PHP.eB	Addgene, 103005	CRISPR KO
CAG-GFP-IRES-CRE	Addgene, 48201	Cre cargoRNA cloning
pHelper	pHelper was a gift from CRUK	AAV production
PX552	Addgene, 60958	AAV production
pMD2.G	Addgene, 12259	Lentivirus
psPAX	Addgene, 12260	Lentivirus
RV-Cag-Dio-GFP	Addgene, 87662	Cre reporter N2a cell line
pBS-Cas9-Bsd	Custom	Cas9 cell line
pGuide-H2B-mCherry	Custom	sgRNA cell lines
Oligos		

Target	Sequence	Application
Asprv1	F: AGGAACCCTGGGGGCCCA R: GTGGGAGCCCTCCGGTGC	Gibson assembly of the coding sequence from species-specific cDNA library into mammalian overexpression vectors
Moap1	F: ACACTGAGACTTCTAGAAGACTGG R: AGTGCAATAGCCTTCTAATTCG	
Pnma1	F: GCTATGACACTATTGGAAGACTGGTGC R: GAAGTGCCCCTCCAGGCC	
Arc	F: GAGCTGGACCATATGACCACC, R: TTCAGGCTGGGTCCTGTCACT	
Zcchc12	F: GCTAGCATCCTTTCACGTTTGG R: CTGTGGTTCAGATAGGCCAATG	
Pnma3	F: ATGAAACAGCGAAGGAAGCCTC R: ATGTGCTGGATGCAGTGGCT	
Pnma5	F: GCCGTGGCTCTATTAGATGA R: CTCACGAAAGGACTCAAGGG	
Pnma6	F: GTTATCACATTCCTCCAGGACG R: ATGGCGGTGACCATGCTG	
Peg10	F: GCTGCTGCGGGTGGTTCC R:CGCAGCACTGCAGGATGA	
Rtl1	F: GATAGAACCCTCTGAAGACT R: GTCAAGTTCATCATCTGAGT	
PEG10	F: GCTGCTGCGGGTGGTTCC R: CGCAGCACTGCAGGATGA	
Peg10	F: GCAGCCCCTATCCCAAACTT R: CGATCAGCATGCTTGTCACG	RT-qPCR
Peg10 CDS A1468C (for PEG10 D491A)	F: GCTTCTGGTGCATCTGGCAAC R: AATCATAGCTCGGACAAACAGGGT	Mutagenesis and cloning via KLD enzyme mix
*Peg10 Δ*NC	F:CCAGCGAAAGCCTCCAAG R: CAAATTCATTTTGCGGCGTCTC	
Peg10 ΔRT	F: TGTGCCTGTTGTAATCACCTGGTCT R: CATGTGGTAGAAGAATGGTGGCTG	
Peg10	F: TGTTTACAGTGCCACAACCGAATT R: AGATGCTCATGCTGATCTGGAG	Indel sequencing
Mm Kras	F:TCTTTTTCAAAGCGGCTGGC R:ACTTGTGGTGGTTGGAGCT	
Hs Vegfa		
CRISPR-mediated perturbations		
Target	Sequence (of spacers unless noted otherwise)	Application

Asprv1	AGGTGTCCCGTAGGTACTGA	CRISPRa
Asprv1	GGGTGGAGCTTCTAGAACAA	
Arc	GCGAGTAGGCGCGGAAGGCG	
Arc	GGCCCGTGGGCGGCAGCTCG	
Peg10	AGCGTGCTTCGCGAGCAGCG	
Peg10	CGCTGCTCGCGAAGCACGCT	
Rtl1	GGGCGCGGCATGCACTGCTT	
Rtl1	AGCAATTTAGGTTCTCAAGA	
Peg10	TGCAGATGCTGATGCATATG	CRISPR KO
Peg10	TCTGTATCCGGTTATGCACC	
Hs Vegfa	GGTGAGTGAGTGTGTGCGTG	
Ms Kras	GCAGCGTTACCTCTATCGTA	
Non-targeting	GCTTTCACGGAGGTTCGACG	CRISPRa and CRISPR KO
Non-targeting	ATGTTGCAGTTCGGCTCGAT	
Peg10	AGAGGGGCTTCACTCCCCTG	CRISPR Knock-in
ssDNA donor for knock-in	GCTAATAGCGACTGCTCTGAATGAATATGTTGAATGTATGCTTCTGTTGTCATTTACAGGAACAGGCGGGTTTTAAGAACCAAAAGACGCCAACCACGAGGGTCCCAGGATCCAGGGCTCCCTCCCCAGGCCACCATGTATCCCTATGACGTGCCCGATTATGCCGCTGCTGCGGGTGGTTCCTCCAACTGCCCGCCCCCTCCCCCTCCCCCTCCTCCCAACAACAACAACAACAACAACACCCCAAAGAGCCCAGGCGTGCCTGACGCCGAAGATGATGATGAACGCAGACACG	

Table S4. Table of plasmid sequences provided as genbank files in Data S1.

ID	plasmid name	description
1	cmv_Hs.annotated.recoded.tiles.gb	Annotated map of recoded regions for human *Peg10*
2	cmv_Hs.cargorna_cas9.gb	Plasmid encoding optimized human cargoRNA for SpCas9 (Hs.cargo[Cas9])
3	cmv_Hs.cargorna_cre.gb	Plasmid encoding optimized human cargoRNA for Cre recombinase (Hs.cargo[Cre])
4	cmv_Hs.cre.utr.tiling.annotated.gb	Annotated map of UTR tiles tested for human *Peg10*
5	cmv_Mm.annotated.recoded.tiles.gb	Annotated map of recoded regions for human *Peg10*
6	cmv_Mm.cargorna_cas9.gb	Plasmid encoding optimized mouse cargoRNA for SpCas9 (Mm.cargo[SpCas9])
7	cmv_Mm.cargorna_cre.gb	Plasmid encoding optimized mouse cargoRNA for Cre recombinase (Mm.cargo[Cre])
8	cmv_Mm.cargorna_h2b_mcherry.gb	Plasmid encoding optimized mouse cargoRNA for H2B-mCherry (Mm.cargo[H2B-mCherry])
9	cmv_Mm.cre.utr.tiling.annotated.gb	Annotated mouse cargo(Cre) with the various tiles annotated of the 3'UTR
10	cmv_Mm.Peg10_ctermHA_delnc_mut.gb	Mutant mouse Peg10 with deletion in the nucleocapsid domain
11	cmv_Mm.Peg10_ctermHA_delrt_mut.gb	Mutant mouse Peg10 with deletion in the reverse transcriptase-like domain
12	HsPeg10.rc3.gb	Optimized and recoded human Peg10 overexpression plasmid
13	MmPeg10.rc4.gb	Optimized and recoded mouse Peg10 overexpression plasmid
14	plv-cmv-cre.gb	Lentiviral overexpression vector used to deliver Cre
15	pmd2-g-syna.gb	Mouse endogenous fusogen used for functional transfer
16	pmd2-g-synb.gb	Mouse endogenous fusogen used for functional transfer

Data S1. (separate file)
Genbank files of annotated plasmids listed in Supplementary Table 4

References

1. J. L. Goodier, H. H. Kazazian Jr, Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*. **135**, 23–35 (2008).

2. A. F. Smit, Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).

3. M. R. Patel, M. Emerman, H. S. Malik, Paleovirology - ghosts and gifts of viruses past. *Curr. Opin. Virol.* **1**, 304–309 (2011).

4. L. Guio, J. González, New Insights on the Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans. *Methods Mol. Biol.* **1910**, 505–530 (2019).

5. C. Feschotte, C. Gilbert, Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* **13**, 283–296 (2012).

6. F. J. Kim, J.-L. Battini, N. Manel, M. Sitbon, Emergence of vertebrate retroviruses and envelope capture. *Virology*. **318**, 183–191 (2004).

7. A. Dupressoir, C. Vernochet, O. Bawa, F. Harper, G. Pierron, P. Opolon, T. Heidmann, Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12127–12132 (2009).

8. C. Myrum, A. Baumann, H. J. Bustad, M. I. Flydal, V. Mariaule, S. Alvira, J. Cuéllar, J. Haavik, J. Soulé, J. M. Valpuesta, J. A. Márquez, A. Martinez, C. R. Bramham, Arc is a flexible modular protein capable of reversible self-oligomerization. *Biochem. J.* **468**, 145–158 (2015).

9. J. Ashley, B. Cordy, D. Lucia, L. G. Fradkin, V. Budnik, T. Thomson, Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. *Cell*. **172**, 262–274.e11 (2018).

10. E. D. Pastuzyn, C. E. Day, R. B. Kearns, M. Kyrke-Smith, A. V. Taibi, J. McCormick, N. Yoder, D. M. Belnap, S. Erlendsson, D. R. Morado, J. A. G. Briggs, C. Feschotte, J. D. Shepherd, The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell*. **173**, 275 (2018).

11. S. F. Erica Korb, Arc in synaptic plasticity: from gene to behavior. *Trends Neurosci.* **34**, 591 (2011).

12. P. Barragan-Iglesias, J. B. De La Pena, T. F. Lou, S. Loerch, Intercellular Arc signaling regulates vasodilation (2020) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3684856).

13. M. Abed, E. Verschueren, H. Budayeva, P. Liu, D. S. Kirkpatrick, R. Reja, S. K. Kummerfeld, J. D. Webster, S. Gierke, M. Reichelt, K. R. Anderson, R. J. Newman, M. Roose-Girma, Z. Modrusan, H. Pektas, E. Maltepe, K. Newton, V. M. Dixit, The Gag protein PEG10 binds to RNA and regulates trophoblast stem cell lineage specification. *PLoS One*. **14**, e0214110 (2019).

14. R. Ono, K. Nakamura, K. Inoue, M. Naruse, T. Usami, N. Wakisaka-Saito, T. Hino, R.

Suzuki-Migishima, N. Ogonuki, H. Miki, T. Kohda, A. Ogura, M. Yokoyama, T. Kaneko-Ishino, F. Ishino, Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**, 101–106 (2006).

15. C. Henke, P. L. Strissel, M.-T. Schubert, M. Mitchell, C. C. Stolt, F. Faschingbauer, M. W. Beckmann, R. Strick, Selective expression of sense and antisense transcripts of the sushi-ichi-related retrotransposon--derived family during mouse placentogenesis. *Retrovirology*. **12**, 9 (2015).

16. M. Krupovic, E. V. Koonin, Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2401–E2410 (2017).

17. S. O. Dodonova, S. Prinz, V. Bilanchone, S. Sandmeyer, J. A. G. Briggs, Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10048–10057 (2019).

18. M. Campillos, T. Doerks, P. K. Shah, P. Bork, Computational characterization of multiple Gag-like human proteins. *Trends Genet.* **22**, 585–589 (2006).

19. S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, F. Zhang, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. **517**, 583–588 (2015).

20. C. S. Wallace, G. L. Lyford, P. F. Worley, O. Steward, Differential intracellular sorting of immediate early gene mRNAs depends on signals in the mRNA sequence. *J. Neurosci.* **18**, 26–35 (1998).

21. M. Golda, J. A. Mótyán, M. Mahdi, J. Tőzsér, Functional Study of the Retrotransposon-Derived Human PEG10 Protease. *Int. J. Mol. Sci.* **21** (2020), doi:10.3390/ijms21072424.

22. A. Saunders, E. Z. Macosko, A. Wysoker, M. Goldman, F. M. Krienen, H. de Rivera, E. Bien, M. Baum, L. Bortolin, S. Wang, A. Goeva, J. Nemesh, N. Kamitaki, S. Brumbaugh, D. Kulp, S. A. McCarroll, Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell*. **174** (2018), pp. 1015–1030.e16.

23. K. Clemens, V. Bilanchone, N. Beliakova-Bethell, Sequence requirements for localization and packaging of Ty3 retroelement RNA. *Virus Res.* (2013) (available at https://www.sciencedirect.com/science/article/pii/S0168170212003784).

24. K. Ridder, A. Sevko, J. Heide, M. Dams, A.-K. Rupp, J. Macas, J. Starmann, M. Tjwa, K. H. Plate, H. Sültmann, P. Altevogt, V. Umansky, S. Momma, Extracellular vesicle-mediated transfer of functional RNA in the tumor microenvironment. *Oncoimmunology*. **4**, e1008371 (2015).

25. Y. Coquin, M. Ferrand, A. Seye, L. Menu, A. Galy, Syncytins enable novel possibilities to transduce human or mouse primary B cells and to achieve well-tolerated in vivo gene transfer. *Cold Spring Harbor Laboratory* (2019), p. 816223.

26. P. S. Kowalski, A. Rudra, L. Miao, D. G. Anderson, Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. *Mol. Ther.* **27**, 710–728 (2019).

27. U. Mock, K. Riecken, B. Berdien, W. Qasim, E. Chan, T. Cathomen, B. Fehse, Novel lentiviral vectors with mutated reverse transcriptase for mRNA delivery of TALE nucleases. *Sci. Rep.* **4**, 6409 (2014).

28. S. J. Kaczmarczyk, K. Sitaraman, H. A. Young, S. H. Hughes, D. K. Chatterjee, Protein delivery using engineered virus-like particles. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16998–17003 (2011).

29. P. E. Mangeot, V. Risson, F. Fusil, A. Marnef, E. Laurent, J. Blin, V. Mournetas, E. Massouridès, T. J. M. Sohier, A. Corbin, F. Aubé, M. Teixeira, C. Pinset, L. Schaeffer, G. Legube, F.-L. Cosset, E. Verhoeyen, T. Ohlmann, E. P. Ricci, Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).

30. R. Kojima, D. Bojar, G. Rizzi, G. C.-E. Hamri, M. D. El-Baba, P. Saxena, S. Ausländer, K. R. Tan, M. Fussenegger, Designer exosomes produced by implanted cells intracerebrally deliver therapeutic cargo for Park\inson's disease treatment. *Nat. Commun.* **9**, 1305 (2018).

31. M. E. Hung, J. N. Leonard, A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery. *J Extracell Vesicles*. **5**, 31027 (2016).

32. J. L. Shirley, Y. P. de Jong, C. Terhorst, R. W. Herzog, Immune Responses to Viral Gene Therapy Vectors. *Mol. Ther.* **28**, 709–722 (2020).

33. J.-E. Park, R. A. Botting, C. Domínguez Conde, D.-M. Popescu, M. Lavaert, D. J. Kunz, I. Goh, E. Stephenson, R. Ragazzini, E. Tuck, A. Wilbrey-Clark, K. Roberts, V. R. Kedlian, J. R. Ferdinand, X. He, S. Webb, D. Maunder, N. Vandamme, K. T. Mahbubani, K. Polanski, L. Mamanova, L. Bolt, D. Crossland, F. de Rita, A. Fuller, A. Filby, G. Reynolds, D. Dixon, K. Saeb-Parsy, S. Lisgo, D. Henderson, R. Vento-Tormo, O. A. Bayraktar, R. A. Barker, K. B. Meyer, Y. Saeys, P. Bonfanti, S. Behjati, M. R. Clatworthy, T. Taghon, M. Haniffa, S. A. Teichmann, A cell atlas of human thymic development defines T cell repertoire formation. *Science*. **367** (2020), doi:10.1126/science.aay3224.

34. R. Ono, S. Kobayashi, H. Wagatsuma, K. Aisaka, T. Kohda, T. Kaneko-Ishino, F. Ishino, A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. *Genomics*. **73**, 232–237 (2001).

35. J. Volff, C. Körting, M. Schartl, Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they evolve into new cellular functions? *Mol. Biol. Evol.* **18**, 266–270 (2001).

36. K. Shigemoto, J. Brennan, E. Walls, C. J. Watson, D. Stott, P. W. Rigby, A. D. Reith, Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting. *Nucleic Acids Res.* **29**, 4079–4088 (2001).

37. R. Vento-Tormo, M. Efremova, R. A. Botting, M. Y. Turco, M. Vento-Tormo, K. B. Meyer, J.-E. Park, E. Stephenson, K. Polański, A. Goncalves, L. Gardner, S. Holmqvist, J. Henriksson, A. Zou, A. M. Sharkey, B. Millar, B. Innes, L. Wood, A. Wilbrey-Clark, R. P. Payne, M. A. Ivarsson, S. Lisgo, A. Filby, D. H. Rowitch, J. N. Bulmer, G. J. Wright, M. J. T. Stubbington, M. Haniffa, A. Moffett, S. A. Teichmann, Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*. **563**, 347–353 (2018).

38. H. Okabe, S. Satoh, Y. Furukawa, T. Kato, S. Hasegawa, Y. Nakajima, Y. Yamaoka, Y. Nakamura, Involvement of PEG10 in human hepatocellular carcinogenesis through interaction with SIAH1. *Cancer Res.* **63**, 3043–3048 (2003).

39. J. Liang, N. Liu, H. Xin, Knockdown long non-coding RNA PEG10 inhibits proliferation, migration and invasion of glioma cell line U251 by regulating miR-506. *Gen. Physiol. Biophys.* **38**, 295–304 (2019).

40. Y. Kawai, K. Imada, S. Akamatsu, F. Zhang, R. Seiler, T. Hayashi, J. Leong, E. Beraldi, N. Saxena, A. Kretschmer, H. Z. Oo, A. Contreras-Sanz, H. Matsuyama, D. Lin, L. Fazli, C. C. Collins, A. W. Wyatt, P. C. Black, M. E. Gleave, Paternally Expressed Gene 10 (PEG10) Promotes Growth, Invasion, and Survival of Bladder Cancer. *Mol. Cancer Ther.* **19**, 2210–2220 (2020).

41. S. Kim, D. Thaper, S. Bidnur, P. Toren, S. Akamatsu, J. L. Bishop, C. Colins, S. Vahid, A. Zoubeidi, PEG10 is associated with treatment-induced neuroendocrine prostate cancer. *J. Mol. Endocrinol.* **63**, 39–49 (2019).

42. R. J. Platt, S. Chen, Y. Zhou, M. J. Yim, L. Swiech, H. R. Kempton, J. E. Dahlman, O. Parnas, T. M. Eisenhaure, M. Jovanovic, D. B. Graham, S. Jhunjhunwala, M. Heidenreich, R. J. Xavier, R. Langer, D. G. Anderson, N. Hacohen, A. Regev, G. Feng, P. A. Sharp, F. Zhang, CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*. **159**, 440–455 (2014).

43. A. Shevchenko, O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14440–14445 (1996).

44. L. Käll, J. D. Storey, W. S. Noble, Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics*. **24**, i42–8 (2008).

45. A. I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).

46. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

47. R. C. Challis, S. R. Kumar, K. Y. Chan, C. Challis, M. J. Jang, P. S. Rajendran, J. D. Tompkins, K. Shivkumar, B. E. Deverman, V. Gradinaru, Widespread and targeted gene expression by systemic AAV vectors: Production, purification, and administration. *Cold Spring Harbor Laboratory* (2018), p. 246405.

48. N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, D. A. Weitz, O. Rozenblatt-Rosen, F. Zhang, A. Regev, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods*. **14** (2017), pp. 955–958.

49. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

50. S. Andrews, FastQC: A quality control tool for high throughput sequence data. Online publication (2010).

51. H. Pimentel, N. L. Bray, S. Puente, P. Melsted, L. Pachter, Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods*. **14**, 687–690 (2017).

52. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

53. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

54. J. P. Connelly, S. M. Pruett-Miller, CRIS.py: A Versatile and High-throughput Analysis Program for CRISPR-based Genome Editing. *Scientific Reports*. **9** (2019), , doi:10.1038/s41598-019-40896-w.

55. M. J. Moore, C. Zhang, E. C. Gantman, A. Mele, J. C. Darnell, R. B. Darnell, Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nature Protocols*. **9** (2014), pp. 263–293.

56. E. L. Van Nostrand, T. B. Nguyen, C. Gelboin-Burkhart, R. Wang, S. M. Blue, G. A. Pratt, A. L. Louie, G. W. Yeo, Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). *Methods Mol. Biol.* **1648**, 177–200 (2017).