RNA DELIVERY

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], Feng Zhang[1,2,3,4,5]*

Eukaryotic genomes contain domesticated genes from integrating viruses and mobile genetic elements. Among these are homologs of the capsid protein (known as Gag) of long terminal repeat (LTR) retrotransposons and retroviruses. We identified several mammalian Gag homologs that form virus-like particles and one LTR retrotransposon homolog, PEG10, that preferentially binds and facilitates vesicular secretion of its own messenger RNA (mRNA). We showed that the mRNA cargo of PEG10 can be reprogrammed by flanking genes of interest with *Peg10*'s untranslated regions. Taking advantage of this reprogrammability, we developed selective endogenous encapsidation for cellular delivery (SEND) by engineering both mouse and human PEG10 to package, secrete, and deliver specific RNAs. Together, these results demonstrate that SEND is a modular platform suited for development as an efficient therapeutic delivery modality.

More than 8% of the human genome is composed of sequences derived from long terminal repeat (LTR) retroelements, including retroviruses, that have integrated into mammalian genomes throughout evolution (*1*–*5*). Retroviruses and retrotransposons have many common mechanistic features, including the core structural gene (known as *gag*); however, whereas retrotransposons replicate intracellularly, the acquisition of the envelope (*env*) gene by retroviruses has enabled intercellular replication (*6*). Most endogenous retroelements have lost their original functions, but some of their genes have been recruited for diverse roles in normal mammalian physiology. For example, the fusogenic syncytins evolved from retroviral *env* proteins (*7*). The *gag* homolog Arc, which forms capsids and has been reported to transfer mRNA (*8*–*10*), is involved in memory consolidation and regulates inflammation in the skin (*11*, *12*). Another *gag* homolog, the LTR retrotransposon–derived protein PEG10, which has been reported to bind RNA and also forms capsids (*13*), is involved in mammalian placenta formation (*14*, *15*). These examples raise the possibility that retroelement-derived proteins encoded in the mammalian genome may be harnessed to transfer specific nucleic acids, providing a potentially programmable mechanism for intercellular communication.

[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.
*Corresponding author. Email: zhang@broadinstitute.org

Computational survey of mammalian capsid-forming *gag* homologs

To identify genes with the potential to transfer specific nucleic acids, we focused on homologs of *gag* that contain the core capsid (CA) domain, which protects the genome of both retrotransposons and exogenous retroviruses (*16*, *17*). Previous genome analyses identified many endogenous *gag* homologs in mammalian genomes (*18*), and experimental efforts have validated the ability of some of these proteins, including *Mus musculus* Arc (MmArc) and MmPEG10, to form capsid-like particles that are secreted within extracellular vesicles (EVs) (*10*, *13*). To ensure a complete list of candidates, we searched the human and mouse genomes for *gag* homologs. This search identified 48 *gag*-derived genes in the human genome and 102 *gag* homologs in the mouse genome; for 19 human genes, an orthologous relationship between human and mouse was readily traced (in several cases, with additional mouse paralogs), whereas the remaining ones appeared to be species-specific (tables S1 and S2).

Canonical genomes of both LTR retrotransposons and retroviruses encode a long polyprotein consisting of several conserved domains: The matrix (MA), CA, and nucleocapsid (NC) form the *gag* subdomain and are responsible for membrane attachment, capsid formation, and genome binding, respectively. The *pol* subdomain contains the protease (PRO), which is responsible for cleaving the polyprotein; the reverse transcriptase (RT), which converts retroelement RNA into DNA; and the integrase (IN) domain, which integrates the genome into that of the host. Some families of endogenous *gag*-containing proteins, such as the PNMA (paraneoplastic antigen Ma) family, contain only the CA and NC subdomains of *gag*, whereas others, such as RTL1 and PEG10 (also known as RTL2 or Mart2), additionally include subdomains of *pol*, namely a PRO domain and a predicted RT-like domain (Fig. 1A and table S1). Phylogenetic analysis of *Peg10* and its homologs supports the origin of this gene from LTR retrotransposons (fig. S1, A and B).

Among these genes, *Arc* is the most well studied. *Drosophila Arc1* (*darc1*) is a *gag* homolog that contains the MA, CA, and NC domains. It has been shown to form capsids, bind its own mRNA, and transfer it from motor neurons to muscles at the neuromuscular junction (*9*). *darc1* mRNA binding is dependent on its own 3′-untranslated region (3′ UTR), and fusion of this sequence to heterologous mRNAs can initiate their export and transfer as well. *MmArc*, by contrast, contains only the CA domain and has also been shown to form capsids and transfer *Arc* and other mRNAs across synapses (*10*).

To narrow down the scope of our analysis, we focused on CA domain–containing proteins that are conserved between human and mouse and have detectable levels of mRNA in adult human tissues, reasoning that such proteins were most likely to have been co-opted for important physiological roles in mammals (fig. S2). We produced mouse versions of the selected CA-containing proteins in *Escherichia coli* and found that a number of these formed higher molecular weight oligomers that were identified by size exclusion (Fig. 1B and fig. S3A), as previously noted for some of these proteins, such as MmArc (*10*). Electron microscopy of these aggregated proteins showed that MmMOAP1, MmZCCHC12, MmRTL1, MmPNMA3, MmPNMA5, MmPNMA6a, and MmPEG10 self-assemble into capsid-like particles, many of which appear spherical (Fig. 1, C and D, and fig. S3, B and C).

MmPEG10 binds and secretes its own mRNA

To determine whether these proteins are secreted within an EV, we overexpressed an epitope-tagged mouse ortholog of each CA-containing gene in human embryonic kidney (HEK) 293 FT cells and harvested both the whole-cell lysate and the virus-like particle (VLP) fraction by clarification and ultracentrifugation of the culture media (Fig. 1E). We found that MmMOAP1, MmArc, MmPEG10, and MmRTL1 were all present in the VLP fraction (Fig. 1F and fig. S4A), but MmPEG10 was the most abundant protein in the VLP fraction (Fig. 1G). Additionally, endogenous MmPEG10, but not MmMOAP1 or MmRTL1, was readily detectable in cell-free adult mouse serum (fig. S4B).

We next tested whether any of the capsid-like particles formed by Gag homologs contained specific mRNAs using RNA sequencing.

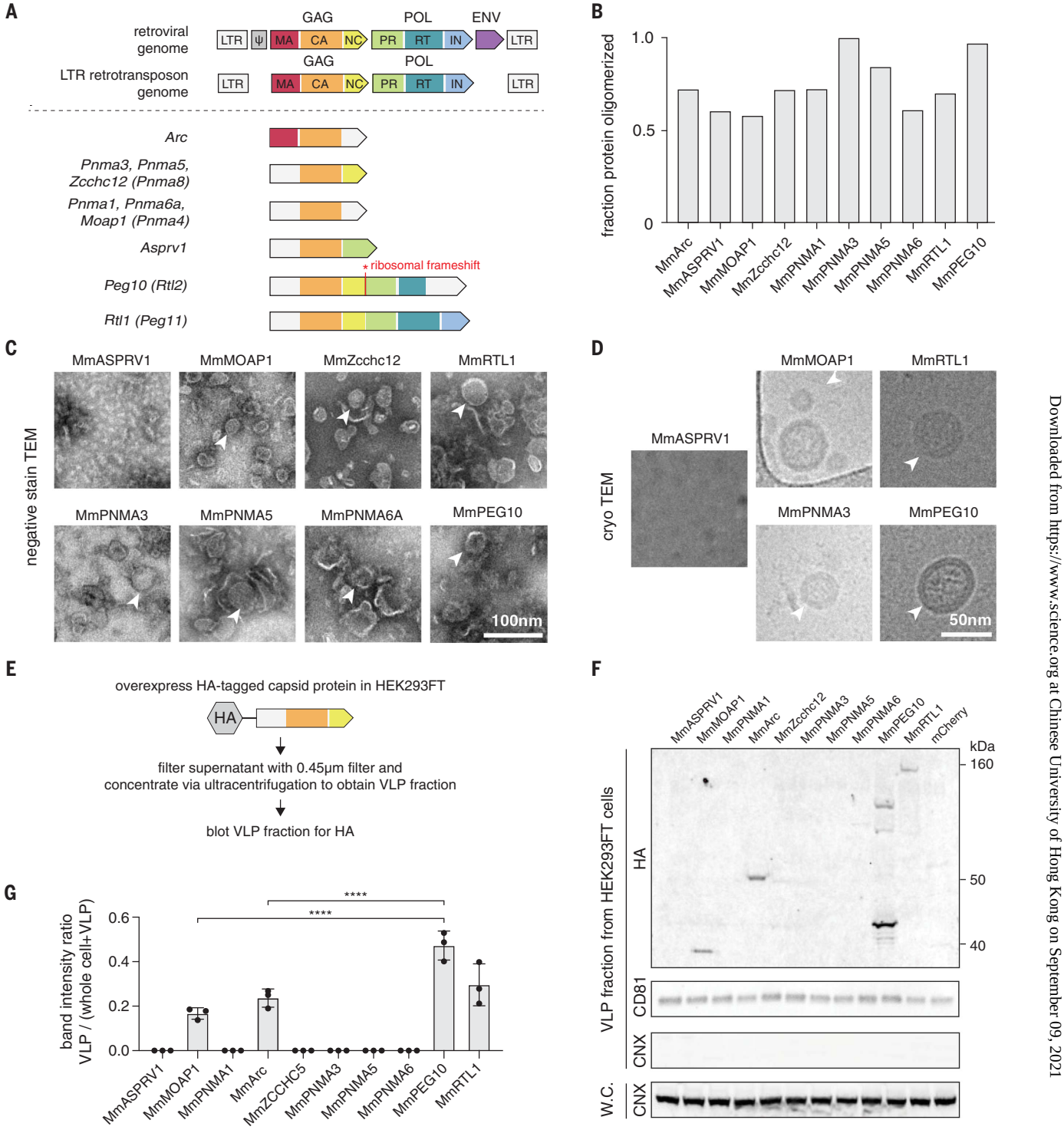

Fig. 1. Identification of mammalian retroelement-derived Gag homologs that form capsids and are secreted. (**A**) Domain architectures of selected capsid (CA)–containing mammalian Gag homologs compared with that of typical retrovirus and LTR retrotransposons. Each group of Gag homologs contains a distinct combination of predicted CA, nucleocapsid (NC), protease (PR), and reverse transcriptase (RT) domains. LTR, long terminal repeat; MA, matrix; IN, integrase. (**B**) Fraction of the total bacterially produced protein that forms oligomers (>600 kD), as determined by size exclusion chromatography. (**C**) Representative negative stain transmission electron micrographs (TEMs) of the Mm orthologs of the CA domain–containing proteins. Scale bar, 100 nm. (**D**) Representative electron micrographs using cryogenic electron microscopy (cryoTEM) of a selected subset of the identified CA domain–containing proteins. Scale bar, 50 nm. (**E**) Method for detecting extracellular forms of CA domain–containing homologs. (**F**) Representative blots of CA domain–containing proteins in the cell-free fraction. CD81 was used as loading control for the ultracentrifuged cell-free fraction. Whole-cell (W.C.) and VLP fraction blots for the endoplasmic reticulum marker CALNEXIN (CNX) ensure equal loading of whole cell protein and the purity of cell-free VLP fraction. (**G**) Quantification of extracellular CA domain–containing proteins [as in (F)] on the basis of $n = 3$ replicates. ****$P < 0.0001$.

To avoid the possibility of transfected Gag homolog expression plasmids contributing to high background signal during sequencing, we used CRISPR activation (*19*) to induce expression of endogenous genes in mouse N2a cells (Fig. 2A and fig. S5A). We performed mRNA sequencing on whole-cell lysate and the VLP fraction (after nuclease treatment to remove any residual, unencapsidated RNA) to identify RNA species in the VLP fraction. We found that *MmPeg10* transcriptional activation led to the accumulation of appreciable amounts of full-length *MmPeg10* mRNA transcripts in the VLP fraction (Fig. 2, B and C). Previous work on MmPEG10 demonstrated that it binds a number of mRNAs inside trophoblast stem cells, including itself (*13*); however, here we further show that MmPEG10

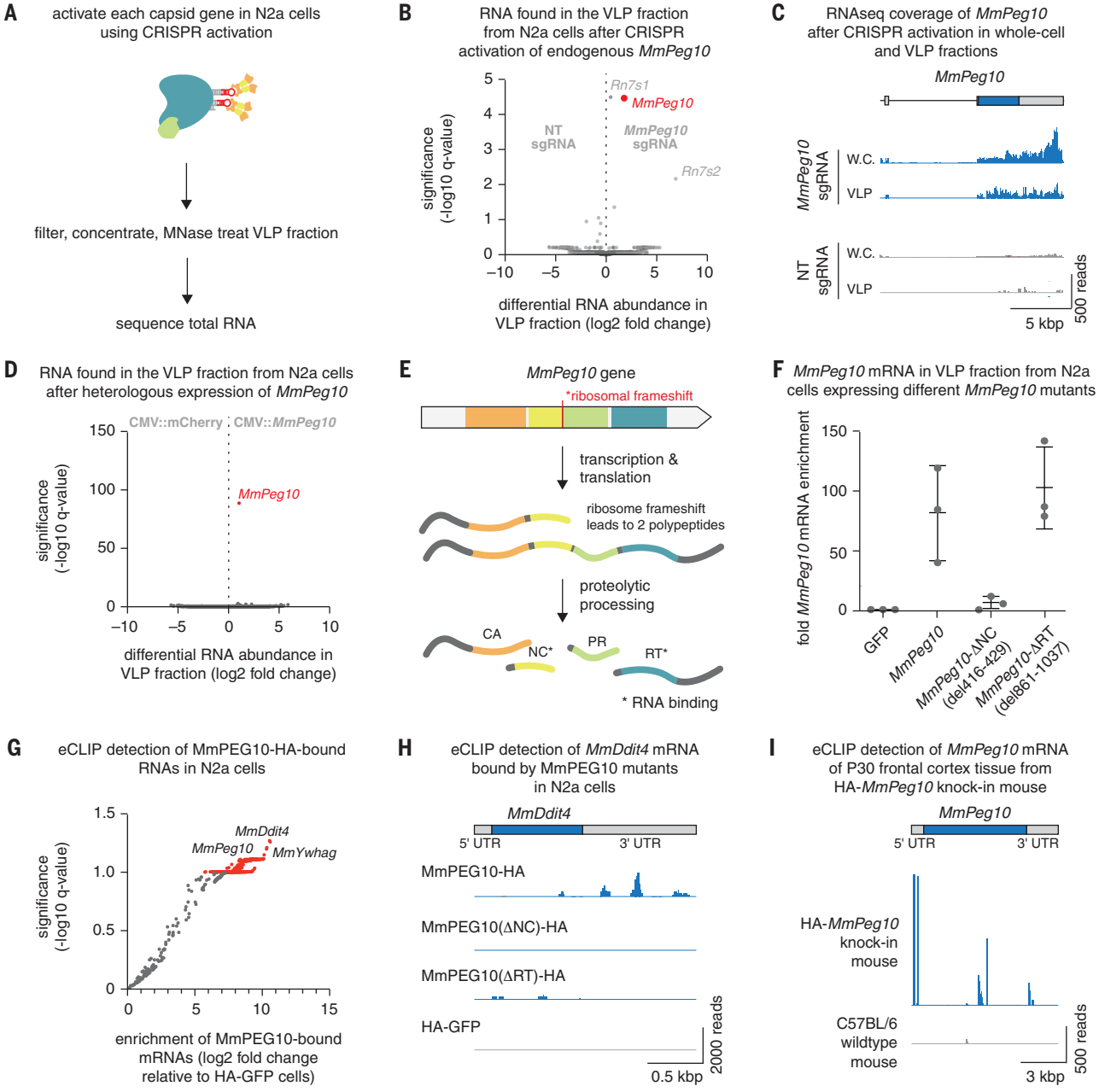

Fig. 2. MmPEG10 protein and mRNA are secreted in vesicles by cells in vitro. (**A**) Method for identifying nucleic acids that are secreted in the VLP fraction upon gene activation of CA domain–containing proteins. (**B**) Differential RNA abundance and significance in the VLP fraction from N2a cells after CRISPR activation of endogenous *MmPeg10*. NT, nontargeting gRNA. (**C**) Alignment of sequencing reads showing sequencing coverage of the *MmPeg10* mRNA from (B). (**D**) Differential RNA abundance and significance in the VLP fraction from N2a cells after heterologous transfection of *MmPeg10*. $n = 3$ replicates. CMV, cytomegalovirus. (**E**) Four domains of MmPEG10 are translated into two isoforms. These are self-processed by the PEG10 protease into separate domains, of which the NC and RT bind RNA. (**F**) Fold enrichment of *MmPeg10* mRNA compared with GFP in the VLP fraction from N2a cells transfected with wild-type *MmPeg10* or deletions of the predicted nucleocapsid (ΔNC) and reverse transcriptase (ΔRT) domains. (**G**) Log_2 fold change and significance of bound RNAs from eCLIP data comparing HA-GFP with wild-type MmPEG10-HA. (**H**) Representative sequencing alignment histogram of the *MmDdit4* locus generated from eCLIP of N2a cells transfected with wild-type or mutant *MmPeg10*. (**I**) Representative sequencing alignment histogram of the *MmPeg10* locus generated from eCLIP data of $n = 3$ HA-PEG10 and $n = 3$ untagged animals.

binds and secretes its own mRNA into the VLP fraction. An important caveat of this experiment is that some of these proteins, particularly MmArc, are subject to regulation at the level of translation, so the lack of enrichment in the VLP fraction could be due to low protein expression (*20*).

To confirm our observation for *MmPeg10*, we transiently transfected overexpression plasmids of UTR-flanked *MmPeg10* into N2a cells and found only enrichment for *MmPeg10* mRNA in the VLP fraction (Fig. 2D) under this overexpression condition. PEG10 contains two putative nucleic acid-binding domains, namely the NC and RT, which are released from the polypeptide upon PEG10 self-processing (*21*) (Fig. 2E, supplementary text 1, and fig. S5, B to D). We generated deletions of these domains and found that mRNA export depends on the MmPEG10 NC, as loss of the nucleic acid–binding zinc finger CCHC motif (residues 416 to 429) from the MmPEG10 NC substantially reduced export of its mRNA (Fig. 2F).

To better understand the roles of the nucleic acid–binding domains of MmPEG10 in RNA binding, we performed enhanced cross-linking and immunoprecipitation (eCLIP) in N2a cells after transient transfection with hemagglutinin (HA)–tagged MmPeg10 as well as the NC and RT mutants (fig. S6, A and B). Compared with the control, MmPEG10 strongly bound a number of mRNAs in N2a cells, including its own mRNA (Fig. 2G). Notably, both the NC and the RT domains are required for the binding of these mRNAs by MmPEG10 (Fig. 2H and fig. S6C). To confirm MmPEG10's cellular role in an in vivo context, we generated knockin mice carrying an N-terminal HA tag on the endogenous MmPEG10 protein (fig. S6D). Expression of *MmPeg10* in cortical neurons has been demonstrated previously (fig. S6E) (*22*). Endogenous MmPEG10 was also found to bind its own mRNA as well as other transcripts abundant in neurons (fig. S6, F and G); in contrast to previous datasets, we detected strong MmPEG10 binding in the 5′ UTR, as well as some additional binding near the boundary between the NC and PRO coding sequences and in the beginning of the 3′ UTR (Fig. 2I) (*13*).

Binding of mRNA by MmPEG10 has been reported to increase the cellular abundance of target transcripts (*13*). To confirm this role of MmPEG10 in its native context in vivo, we perturbed *MmPeg10* gene expression in the postnatal mouse brain and assessed the expression changes of MmPEG10-bound transcripts (supplementary text 2). We found that the mRNAs of 49 genes that are down-regulated in the brain upon *MmPeg10* knockout are bound to MmPEG10 in the age-matched mouse brain (fig. S7F), suggesting that one of the functions of MmPEG10 is to bind and stabilize mRNAs with fundamental roles in neurodevelopment.

Pseudotyped PEG10 VLPs can deliver engineered cargo mRNAs bearing RNA packaging signals from PEG10 UTRs

To reprogram MmPEG10 to bind and package heterologous RNA, we tested whether a cargo mRNA consisting of both the 5′ and 3′ UTR of *MmPeg10* flanking a gene of interest would be efficiently packaged, exported, delivered, and translated in recipient cells (Fig. 3A). This UTR grafting approach has been demonstrated for the Ty3 retroelement and *darc1* (*9*, *23*). We first used a Cre-loxP system, a highly sensitive system for tracking RNA exchange that has been used previously with exosomes in vivo (*24*). We flanked the Cre recombinase coding sequence with the *MmPeg10* UTRs and cotransfected it with *MmPeg10* with and without a fusogen, the vesicular stomatitis virus envelope protein (VSVg) (Fig. 3A). We found that MmPEG10 VLPs pseudotyped with VSVg are secreted within EVs that mediate transfer of Cre mRNA, not protein, into target loxP–green fluorescent protein (GFP) reporter N2a cells in a VSVg- and *MmPeg10* UTR–dependent manner (Fig. 3, B to D; fig. S8; and supplementary text 3). This result suggests that addition of the *Peg10* UTRs enables the functional intercellular transfer of an mRNA via VLPs and that these VLPs require a fusogenic protein for cell entry.

We next examined whether there is a minimal UTR packaging signal for mediating efficient packaging and functional transfer. The 3′ UTR of *MmPeg10* is ~4 kb long, but eCLIP indicates that only portions of the 3′ UTR are bound by MmPEG10 (Fig. 2I). We created constructs that encode the *MmPeg10* 5′ UTR, Cre, and 500–base pair (bp) segments of the *MmPeg10* 3′ UTR. We found that the proximal 500 bp of the *MmPeg10* 3′ UTR are sufficient for efficient functional transfer of Cre mRNA into target reporter cells (Fig. 3E). Notably, no efficient functional mRNA transfer was observed for non–UTR-flanked Cre or for Cre without the proximal 500 bp of the 3′ UTR. Henceforth, we refer to RNA cargo flanked by the *MmPeg10* 5′ UTR and the proximal 500 bp of the 3′ UTR as Mm.cargo(RNA), where "(RNA)" specifies the cargo being flanked [e.g., Mm.cargo(Cre)].

Like the mouse ortholog, human PEG10 (HsPEG10) is an abundantly secreted protein in the VLP fraction (fig. S10A). Using the same approach that we employed with *MmPeg10*, we identified that the 5′ UTR and the first 500 bp of the *HsPEG10* 3′ UTR are sufficient to mediate functional transfer of Cre mRNA, hereafter denoted as Hs.cargo(RNA) (Fig. 3F). Notably, these functional regions of the UTRs are highly conserved across mammals (fig. S10B). Similar to its mouse ortholog, the human system is specific and requires *HsPEG10* UTR sequences for functional mRNA transfer, whereas nonflanked Cre produced only minimal reporter cell activity (Fig. 3F).

To further boost the packaging of a cargo RNA by PEG10, we explored the impact of removing any additional PEG10 *cis* binding elements within the *MmPeg10*/*HsPEG10* coding sequence. For both human and mouse orthologs, transfer was increased as a result of recoding the sequence between the NC and the PRO domains, which corresponds to the MmPEG10-bound region in the eCLIP experiments (Fig. 2I and supplementary text 4).

Combining these optimizations, we produced VLPs with the recoded mouse and human PEG10 (rMmPEG10 or rHsPEG10), VSVg, and the optimized cargoRNA containing the first 500 bp of the 3′ UTR; we refer to this system as selective endogenous encapsidation for cellular delivery (SEND). With SEND, we detected a substantial (up to 60%) increase in the functional transfer of cargo(Cre) into N2a cells for both human and mouse PEG10 (Fig. 3, G and H). Furthermore, we showed that VLPs produced with rMmPEG10 can mediate the functional transfer of H2B-mCherry (fig. S12, A and B). A comparison of SEND with previously developed delivery vectors showed that SEND is four to five times less potent than an integrating lentiviral vector, as assayed by digital droplet polymerase chain reaction and functional titration (fig. S12, B to E). However, given that SEND delivers mRNA rather than integrating an overexpression cassette, we expect it to perform competitively against other mRNA delivery vehicles.

PEG10 is a modular platform for RNA delivery

To generate a fully endogenous SEND system, we tested whether VSVg can be replaced with an endogenous fusogenic transmembrane protein. Given the overlapping tissue expression of *MmPeg10*/*HsPEG10* and syncytin genes (supplementary text 5), we tested the feasibility of pseudotyping the mouse SEND system with MmSYNA or MmSYNB compared with pseudotyping with VSVg. Pseudotyped particles were incubated with tail-tip fibroblasts from loxP-tdTomato reporter mice, a cell type that we have found amenable to transduction by these fusogens. Based on previous reports, we added the transduction enhancer vectofusin-1 to the supernatant for MmSYNA and MmSYNB particles to enhance in vitro transduction (*25*). In these primary cells, both VSVg and MmSYNA enabled SEND-mediated functional transfer of Mm.cargo(Cre), whereas MmSYNB did not (Fig. 4, A and B). Again, this packaging was highly specific, as only UTR-flanked mRNA [i.e., Mm.cargo(Cre)] was functionally transferred. Together with MmSYNA, SEND can be configured as a fully endogenous system for functional gene transfer.

Supported by our understanding of the minimal requirements for PEG10-mediated mRNA delivery (i.e., UTRs and an endogenous fusogen), we could begin to probe the endogenous role of

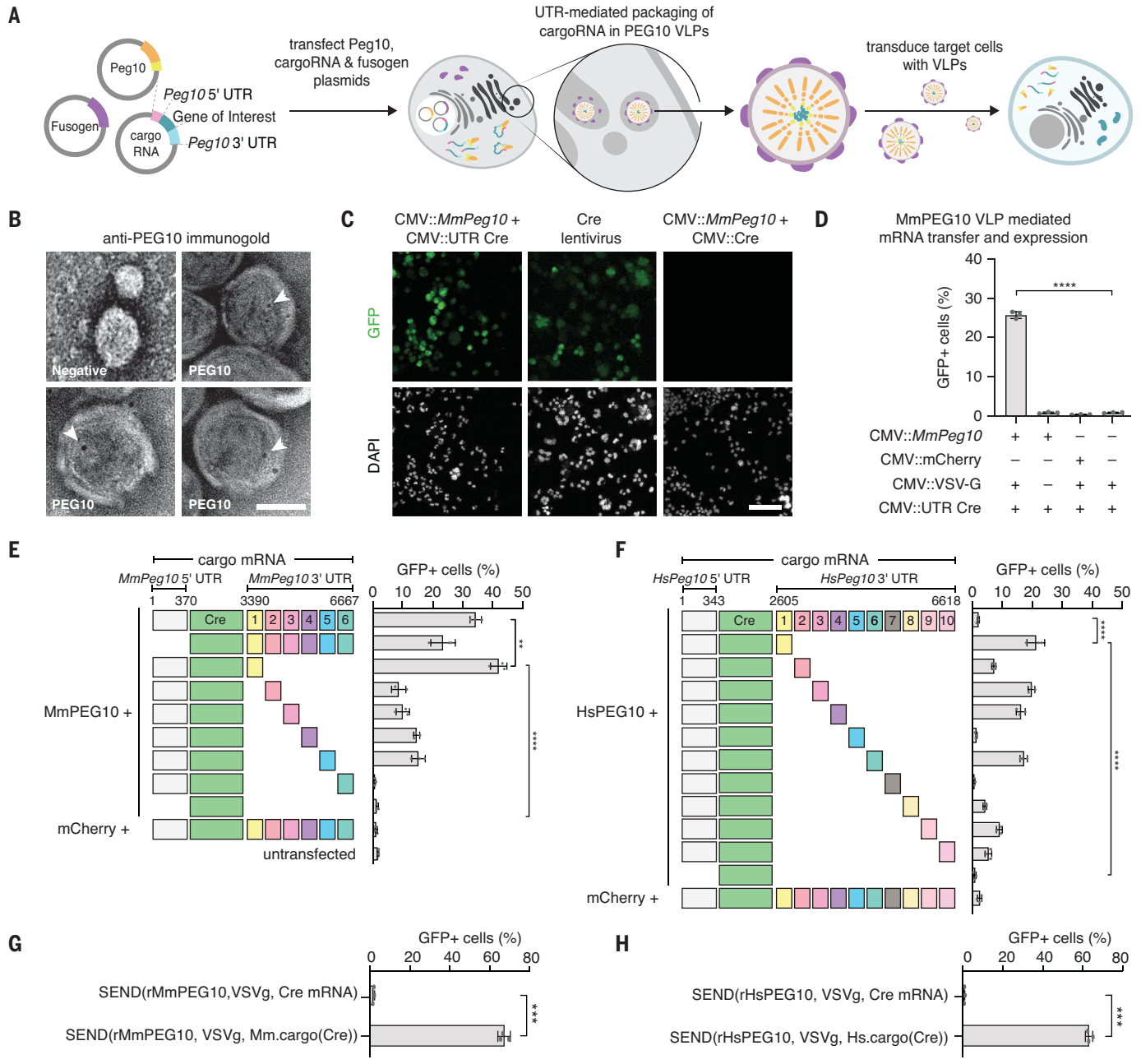

Fig. 3. Flanking mRNA with *MmPeg10* 5′ and 3′ UTRs enables functional intercellular transfer of mRNA into a target cell. (**A**) Schematic showing reprogramming MmPEG10 for functional delivery of a cargo RNA flanked with the *MmPeg10* 5′ and 3′ UTRs [hereafter "cargo(RNA)"]. (**B**) Representative TEMs of VLP fraction immunogold labeled for MmPEG10. Text labels indicate transfection of cells with *MmPeg10* or mock (negative). Arrowheads indicate gold labeling. Scale bar, 50 nm. (**C**) Representative images of loxP-GFP N2a cells treated with VSVg-pseudotyped MmPEG10 VLPs, which were produced by transfecting Mm.cargo(Cre) or Cre mRNA, and a lentivirus encoding Cre. Scale bar, 100 μm. DAPI, 4′,6-diamidino-2-phenylindole. (**D**) Functional transfer of RNA into loxP-GFP N2a cells mediated by VSVg-pseudotyped MmPEG10 VLPs. Data were quantified by flow cytometry 72 hours after VLP addition, $n = 3$ replicates. (**E**) Functional transfer of RNA into loxP-GFP N2a cells mediated by VSVg-pseudotyped VLPs that were produced with *MmPeg10* or mCherry and Mm.cargo(Cre) constructs that encoded tiles of the *MmPeg10* 3′ UTR. Data were quantified by flow cytometry 72 hours after VLP addition, $n = 3$ replicates. (**F**) Functional transfer of RNA into loxP-GFP N2a cells mediated by VSVg-pseudotyped VLPs that were produced with *HsPEG1010* or mCherry and Hs.cargo(Cre) constructs that encoded tiles of the *HsPeg10* 3′ UTR. Data were quantified by flow cytometry 72 hours after VLP addition, $n = 3$ replicates. (**G**) Functional transfer of RNA into loxP-GFP N2a cells mediated by VSVg-pseudotyped VLPs that were produced with r*MmPeg10* and Mm.cargo(Cre) or Cre mRNA. Data were quantified by flow cytometry 72 hours after VLP addition, $n = 3$ replicates. (**H**) Functional transfer of RNA into loxP-GFP N2a cells mediated by VSVg-pseudotyped VLPs that were produced with r*HsPeg10* and Hs.cargo(Cre) or Cre mRNA. Data were quantified by flow cytometry 72 hours after VLP addition, $n = 3$ replicates. For all panels, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, one-way analysis of variance.

MmPEG10-mediated *MmPeg10* RNA delivery in neurons. The functional transfer of MmSYNA-pseudotyped VLPs that carry the native PEG10 transcript into primary mouse cortical neurons led to up-regulation of a number of genes involved in neurodevelopment (supplementary text 6). This finding reinforces the notion that one role of endogenous *MmPeg10* delivery is binding and stabilizing specific mRNA transcripts in recipient cells. RNA sequencing of N2a cells receiving Mm.cargo(Peg10) revealed substantial gene expression changes upon

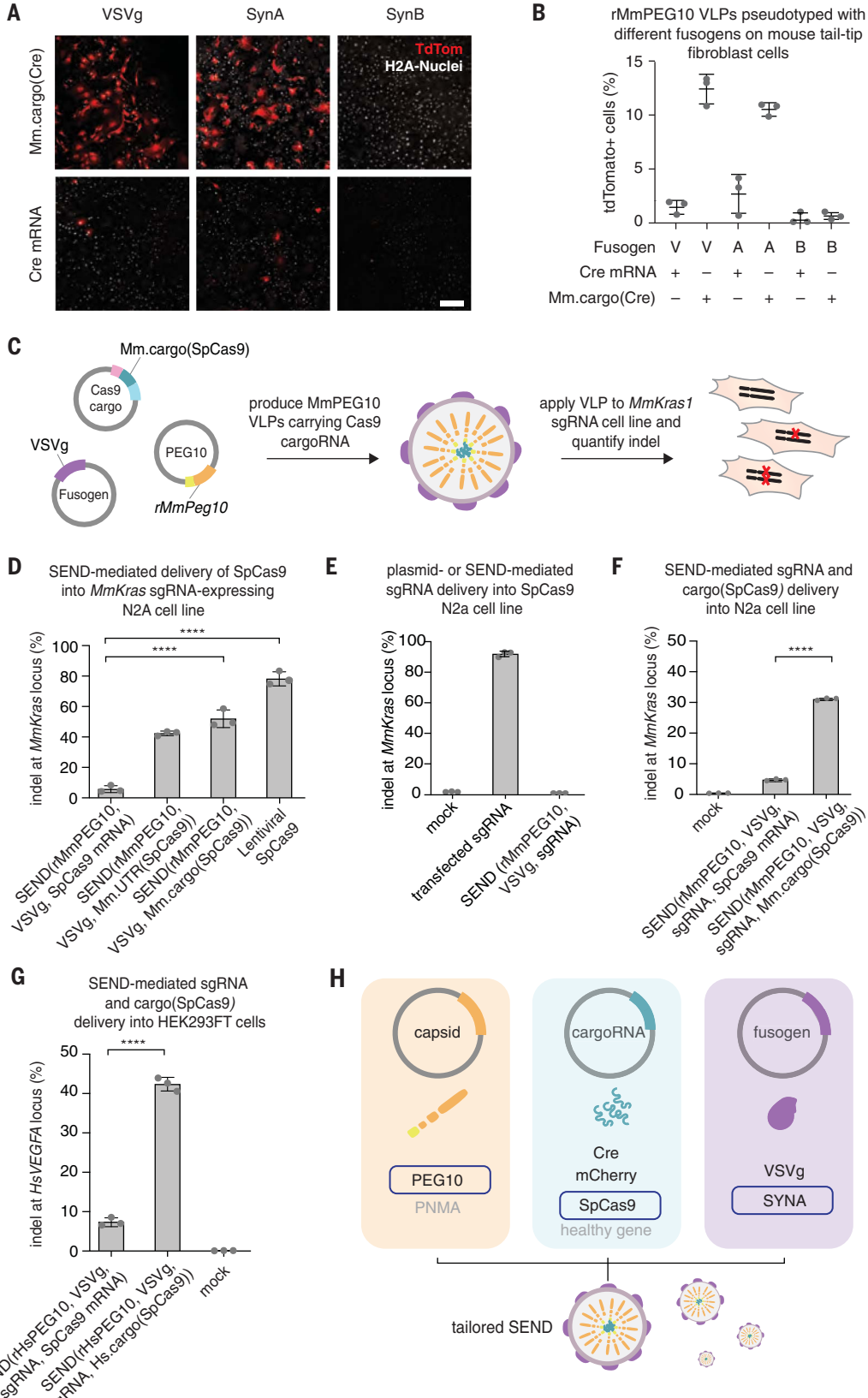

Fig. 4. SEND is a modular system capable of delivering gene editing tools into human and mouse cells. (**A**) Representative images demonstrating functional transfer of Mm.cargo(Cre) or Cre mRNA in rMmPEG10 VLPs pseudotyped with VSVg (V), MmSYNA (A), or MmSYNB (B) in Ai9 (loxP-tdTomato) tail-tip fibroblasts. Scale bar, 200 μm. (**B**) Percent of tdTomato-positive cells out of the total number of H2A-stained nuclei from high content imaging of $n = 3$ replicates of (A). (**C**) Schematic representing the retooling of SEND for genome engineering. (**D**) Indels at the *MmKras* locus in *MmKras1*-sgRNA-N2a cells treated with SEND (VSVg-pseudotyped rMmPEG10 VLPs) containing SpCas9 mRNA, Mm.UTR (SpCas9), or Mm.cargo(SpCas9) and a lentivirus encoding SpCas9. Indels were quantified by NGS 72 hours after VLP or lentivirus addition, $n = 3$ replicates. (**E**) Indels at the mouse *MmKras* locus in a constitutively expressing SpCas9 N2a cell line either transfected with a plasmid carrying the *MmKras* sgRNA or treated with SEND (rMmPEG10, VSVg, or *MmKras* sgRNA). Indels were quantified by NGS after 72 hours, $n = 3$ replicates. (**F**) Indels at the *MmKras* locus in N2a cells treated with SEND (VSVg-pseudotyped rMmPeg10 SEND VLPs) containing either SpCas9 mRNA or Mm.cargo (SpCas9) and sgRNA. Indels were quantified by NGS 72 hours after VLP addition, $n = 3$ replicates. (**G**) Indels at the *HsVEGFA* locus in HEK293FT cells treated with SEND (VSVg-pseudotyped rHsPEG10 VLPs) containing either SpCas9 mRNA or Hs.cargo(SpCas9) and an unmodified sgRNA. Indels were determined by NGS 72 hours after VLP addition, $n = 3$ replicates. (**H**) SEND is a modular delivery platform combining an endogenous Gag homolog, cargo mRNA, and fusogen, which can be tailored for specific contexts.

MmPeg10 delivery that were largely abrogated with PEG10-mediated Mm.cargo(Cre) delivery (supplementary text 7). This suggests that transferring a reprogrammed cargo does not have the same impact on recipient cells as transferring *MmPeg10* and indicates that *MmPeg10* transcript delivery rather than the delivery of MmPEG10 protein is responsible for the observed gene expression changes. It remains unclear whether MmPEG10 VLPs are natively pseudotyped by the endogenous fusogen MmSYNA to enable cellular uptake of PEG10 VLPs in the central nervous system.

To further characterize the modularity of the components of this system, we tested different cargoRNAs. Using the same pipeline developed for cargo(Cre), we tested whether SEND could mediate the functional transfer of a large ~5-kb Mm.cargo(SpCas9) into N2a cell lines that constitutively express a single guide RNA (sgRNA) against *MmKras* (Fig. 4C). SEND was able to functionally transfer SpCas9, leading to ~60% insertions and deletions (indels) in recipient cells (Fig. 4D); similar to the results with Cre, SEND is specific and only able to efficiently functionally transfer SpCas9 flanked by either the full-length or optimized *Peg10* UTR sequences.

To create an all-in-one vector for delivery of sgRNA and SpCas9, we first tested whether an sgRNA can be efficiently delivered by SEND. We independently packaged an sgRNA targeting *Kras* into rMmPeg10 VLPs by coexpressing rMmPeg10 with VSVg and a U6-driven sgRNA and incubated them with Cas9-expressing N2a cells; we detected very little activity even though direct transfection of the guide showed robust indel formation (Fig. 4E). We found, however, that copackaging the guide alongside Mm.cargo(SpCas9) by coexpressing Mm.cargo (SpCas9) with a U6-driven sgRNA on a separate plasmid was sufficient to mediate 30% indels (Fig. 4F). To determine the reproducibility of this genome-editing approach, we repeated this copackaging strategy with the human SEND system and were able to generate ~40% indels in HEK293FT cells at the *HsVEGFA* locus (Fig. 4G).

The development of SEND (Fig. 4H) from an endogenous retroelement complements existing delivery approaches using lipid nanoparticles (*26*), VLPs derived from bona fide retroviruses (*27*–*29*), and active mRNA-loading approaches in EVs (*30*, *31*). Moreover, SEND may have reduced immunogenicity compared with currently available viral vectors (*32*) because of its use of endogenous human proteins. Supporting this are gene expression data from the developing human thymus, which demonstrate that *HsPEG10* is highly expressed compared with other CA-containing genes in the thymic epithelium (fig. S16) (*33*), which is responsible for T cell tolerance induction. As a modular, fully endogenous system, SEND has the potential to be extended into a minimally immunogenic delivery platform that can be repeatedly dosed, which greatly expands the applications for nucleic acid therapy.

REFERENCES AND NOTES

1. J. L. Goodier, H. H. Kazazian Jr., *Cell* **135**, 23–35 (2008).
2. A. F. Smit, *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
3. M. R. Patel, M. Emerman, H. S. Malik, *Curr. Opin. Virol.* **1**, 304–309 (2011).
4. L. Guio, J. González, *Methods Mol. Biol.* **1910**, 505–530 (2019).
5. C. Feschotte, C. Gilbert, *Nat. Rev. Genet.* **13**, 283–296 (2012).
6. F. J. Kim, J.-L. Battini, N. Manel, M. Sitbon, *Virology* **318**, 183–191 (2004).
7. A. Dupressoir *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12127–12132 (2009).
8. C. Myrum *et al.*, *Biochem. J.* **468**, 145–158 (2015).
9. J. Ashley *et al.*, *Cell* **172**, 262–274.e11 (2018).
10. E. D. Pastuzyn *et al.*, *Cell* **173**, 275–288 (2018).
11. E. Korb, S. Finkbeiner, *Trends Neurosci.* **34**, 591–598 (2011).
12. P. Barragan-Iglesias, J. B. De La Pena, T. F. Lou, S. Loerch, *Cell Rep.* 10.2139/ssrn.3684856 (2020). doi: 10.2139/ssrn.3684856.
13. M. Abed *et al.*, *PLOS ONE* **14**, e0214110 (2019).
14. R. Ono *et al.*, *Nat. Genet.* **38**, 101–106 (2006).
15. C. Henke *et al.*, *Retrovirology* **12**, 9 (2015).
16. M. Krupovic, E. V. Koonin, *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, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10048–10057 (2019).
18. M. Campillos, T. Doerks, P. K. Shah, P. Bork, *Trends Genet.* **22**, 585–589 (2006).
19. S. Konermann *et al.*, *Nature* **517**, 583–588 (2015).
20. C. S. Wallace, G. L. Lyford, P. F. Worley, O. Steward, *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**, 2424 (2020).
22. A. Saunders *et al.*, *Cell* **174**, 1015–1030.e16 (2018).
23. K. Clemens, V. Bilanchone, N. Beliakova-Bethell, *Virus Res.* **171**, 319–331 (2013).
24. K. Ridder *et al.*, *OncoImmunology* **4**, e1008371 (2015).
25. Y. Coquin, M. Ferrand, A. Seye, L. Menu, A. Galy, bioRxiv 816223 [Preprint]. 24 October 2019.
26. P. S. Kowalski, A. Rudra, L. Miao, D. G. Anderson, *Mol. Ther.* **27**, 710–728 (2019).
27. U. Mock *et al.*, *Sci. Rep.* **4**, 6409 (2014).
28. S. J. Kaczmarczyk, K. Sitaraman, H. A. Young, S. H. Hughes, D. K. Chatterjee, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16998–17003 (2011).
29. P. E. Mangeot *et al.*, *Nat. Commun.* **10**, 45 (2019).
30. R. Kojima *et al.*, *Nat. Commun.* **9**, 1305 (2018).
31. M. E. Hung, J. N. Leonard, *J. Extracell. Vesicles* **5**, 31027 (2016).
32. J. L. Shirley, Y. P. de Jong, C. Terhorst, R. W. Herzog, *Mol. Ther.* **28**, 709–722 (2020).
33. J.-E. Park *et al.*, *Science* **367**, eaay3224 (2020).

ACKNOWLEDGMENTS

We thank D. S. Yun for electron microscopy assistance, A. Koller for mass spectrometry assistance, L. Wu and the Harvard GMF for the generation of transgenic animals, A. Tang for illustration assistance, and the entire Zhang laboratory for support and advice. **Funding:** This work was supported by a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT (M.S.); National Institutes of Health Intramural Research Program (E.V.K.); National Institutes of Health grants 1R01-HG009761 and 1DP1-HL141201 (F.Z.); Howard Hughes Medical Institute (F.Z.); Open Philanthropy (F.Z.); G. Harold and Leila Y. Mathers Charitable Foundation (F.Z.); Edward Mallinckrodt, Jr. Foundation (F.Z.); Poitras Center for Psychiatric Disorders Research at MIT (F.Z.); Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT (F.Z.); Yang-Tan Center for Molecular Therapeutics at MIT (F.Z.); Lisa Yang (F.Z.); Phillips family (F.Z.); R. Metcalfe (F.Z.); and J. and P. Poitras (F.Z.). **Author contributions:** M.S. and F.Z. conceived the project. M.S., B.L., and F.Z. designed the experiments. M.S., B.L., J.S., A.L., X.J., and C.C.L. performed the experiments. M.S., B.L., J.S., and F.Z. analyzed the data. M.S., S.L.M., and E.V.K. performed bioinformatics analysis of Gag protein diversity. F.Z. supervised the research and experimental design with support from R.K.M., and M.S., B.L., R.K.M., and F.Z. wrote the manuscript with input from all authors. **Competing interests:** M.S., B.L., and F.Z. are co-inventors on a US provisional patent application filed by the Broad Institute related to this work. (U.S. Provisional Patent Application no. 63/191,067) F.Z. is a cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences. **Data and materials availability:** Expression plasmids are available from Addgene under a uniform biological material transfer agreement. Additional information is available through the Zhang Lab website (https://zlab.bio). Next-generation sequencing data generated are available from National Center for Biotechnology Information Sequence Read Archive with accession number PRJNA743280. All other data are available in the paper and supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/373/6557/882/suppl/DC1
Materials and Methods
Supplementary Text 1 to 7
Figs. S1 to S16
Tables S1 to S4
References (*34*–*56*)
Data S1

17 January 2021; resubmitted 26 April 2021
Accepted 6 July 2021
10.1126/science.abg6155

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

Michael SegelBlake LashJingwei SongAlim LadhaCatherine C. LiuXin JinSergei L. MekhedovRhiannon K. MacraeEugene V. KooninFeng Zhang

Science, 373 (6557),

Hitching a ride with a retroelement

Retroviruses and retroelements have inserted their genetic code into mammalian genomes throughout evolution. Although many of these integrated virus-like sequences pose a threat to genomic integrity, some have been retooled by mammalian cells to perform essential roles in development. Segel *et al*. found that one of these retroviral-like proteins, PEG10, directly binds to and secretes its own mRNA in extracellular virus–like capsids. These virus-like particles were then pseudotyped with fusogens to deliver functional mRNA cargos to mammalian cells. This potentially provides an endogenous vector for RNA-based gene therapy. —DJ

Science (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.