Supplemental Information

Plug and Play Protein Modification using Homology-independent Universal Genome Engineering

Yudong Gao[1], Erin Hisey[1], Tyler W.A. Bradshaw[1,2], Eda Erata[1], Walter E. Brown[1], Jamie L. Courtland[1,2], Akiyoshi Uezu[1], Yu Xiang[1], Yarui Diao[1], and Scott H. Soderling[1,2]*

[1]Department of Cell Biology, Duke University Medical School, Durham, North Carolina, USA.

[2]Department of Neurobiology, Duke University Medical School, Durham, North Carolina, USA.

* Corresponding Author & Lead Contact: Scott H. Soderling, E-mail: scott.soderling@duke.edu

Figure S1

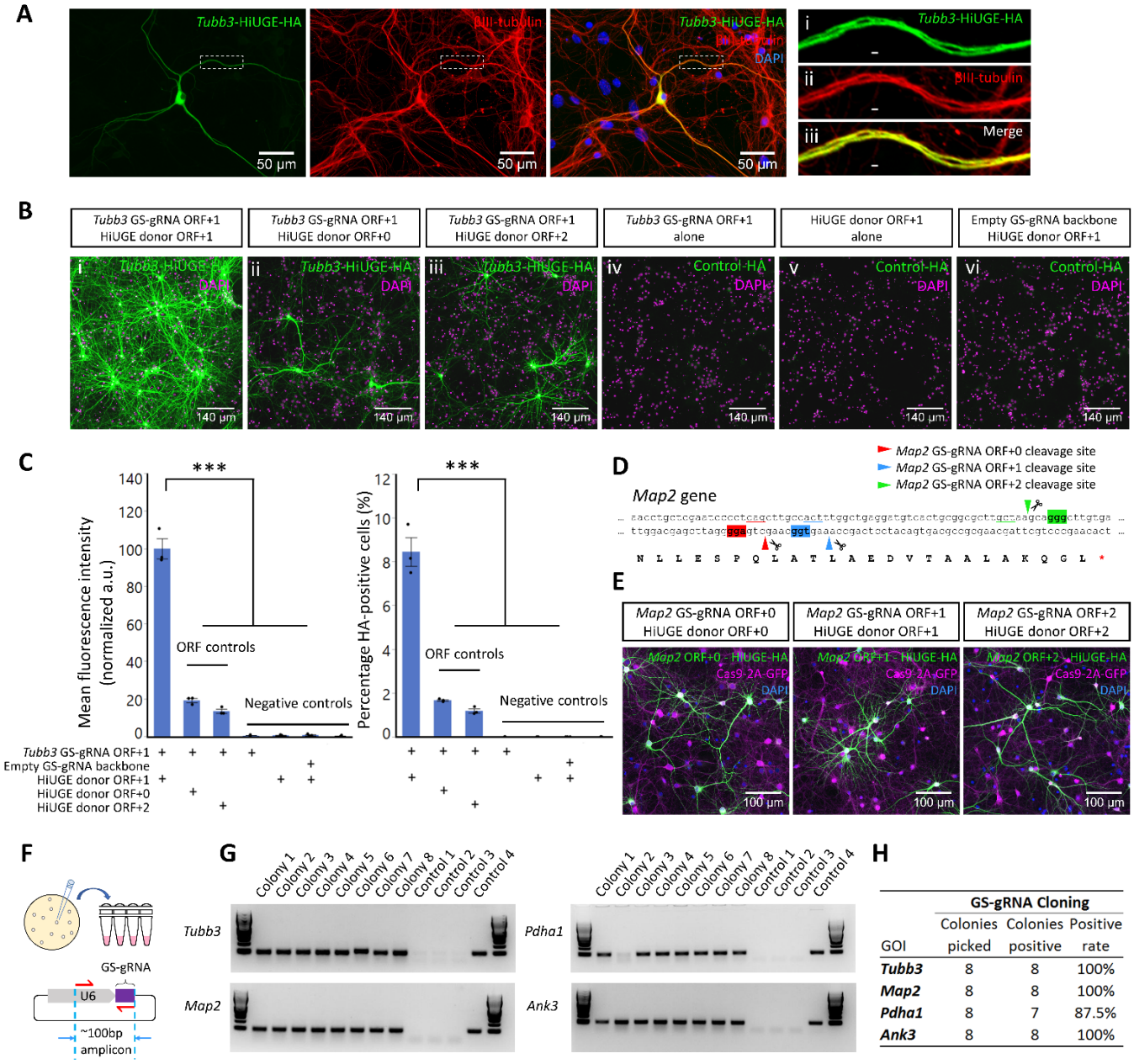

Figure S1. **Additional Qualitative and Quantitative Data of Proof-of-principle Experiments, Related to Figure 1 and Figure 2.**

(**A**) Immunolabeling of HiUGE-mediated HA-epitope KI to mouse *Tubb3* (green), counterstained with an anti-βIII-tubulin antibody (red). Colocalization of the fluorescent signal (yellow) is evident. Subpanels (i-iii) show enlarged views of the boxed region from the left. Scale bar is indicated in each panel, or within insets (2μm). (**B**) Representative immunolabeling of HiUGE-mediated HA-epitope KI to *Tubb3* in a control experiment after transduction with equal amount of AAV (5×10^{10} GC / mL per virus in the culture medium), showing HA-epitope immunoreactivity of (i) correct pairing of GS-gRNA and HiUGE donor open reading frame (ORF); (ii-iii) incorrect pairing of GS-gRNA and HiUGE donor ORFs; (iv-vi) negative controls. Nuclei labeling with DAPI and the scale bar are indicated in each panel. (**C**) Quantification of (**B**), showing the normalized mean fluorescence intensities (arbitrary units, a.u.) and the

HA-positive neurons as a percentage of all cells (DAPI positive, including neuronal and non-neuronal cells) across experimental groups. Error bars represent standard error of the mean (SEM). Correct ORF pairing of *Tubb3* GS-gRNA and HiUGE donor showed significantly higher HA-epitope fluorescence intensity over ORF-mismatched controls and negative controls (***: $p < 0.001$, one-way ANOVA followed by Tukey-Kramer HSD *post hoc* test, n=3). Also, HA-labeling in the correct ORF was approximately five fold more efficient (8.4 ± 0.7 % of all cells) compared to occasional HA-positive cells observed from out-of-frame ORFs (ORF+0: 1.7 ± 0.04 %; ORF+2: 1.2 ± 0.1 %. ***: $p < 0.001$, one-way ANOVA followed by Tukey-Kramer HSD *post hoc* test, n=3). Of note, due to the presence of glia cells in the co-culture that do not express the neuronal specific ßIII-tubulin, the cellular labeling efficiency is under-estimated. (**D**) To demonstrate the flexibility of GS-gRNA selection and the capability of HiUGE donor vectors of all three open reading frames (ORFs) to facilitate protein modification, three different GS-gRNAs, one for each ORF (ORF+0, ORF+1, ORF+2), were designed to target mouse *Map2* gene. Shaded DNA triplets denote the protospacer adjacent motif (PAM) of the target sequence for each GS-gRNA. Underlined DNA triplets denote the last codon before the Cas9 cleavage site for each GS-gRNA. (**E**) Representative images showing comparable detection of HA-epitope KI when GS-gRNAs of each ORF were paired with HiUGE donors of the corresponding ORF. GFP fluorescence of the Cas9-2A-GFP, nuclei labeling with DAPI, and the scale bar are indicated in each panel. **(F)** Illustration of colony PCR screening to demonstrate GS-gRNA cloning efficiency. 23-24mer GS-gRNA oligos of four exemplary genes (mouse *Tubb3*, *Map2*, *Pdha1*, and *Ank3*) were annealed and ligated into the SapI sites of GS-gRNA plasmids. Colony PCRs were performed using a common forward primer in the upstream U6 promotor region paired with the reverse GS-gRNA oligo of each gene as the reverse primer to detect a ~100bp amplicon when the GS-gRNAs were successfully integrated. (**G**) Eight colonies (Colony 1-8) were randomly picked from each LB plate and used as PCR template to screen for successful GS-gRNA integration. Control conditions (Control 1-4) include template materials from: (1) non-colony areas on the LB plate, (2) empty GS-gRNA backbone plasmid, (3) water, and (4) positive control with previously sequenced GS-gRNA plasmid for each gene. (**H**) Positive rates of GS-gRNA integration are calculated, showing the cloning method is highly robust and efficient.

Figure S2

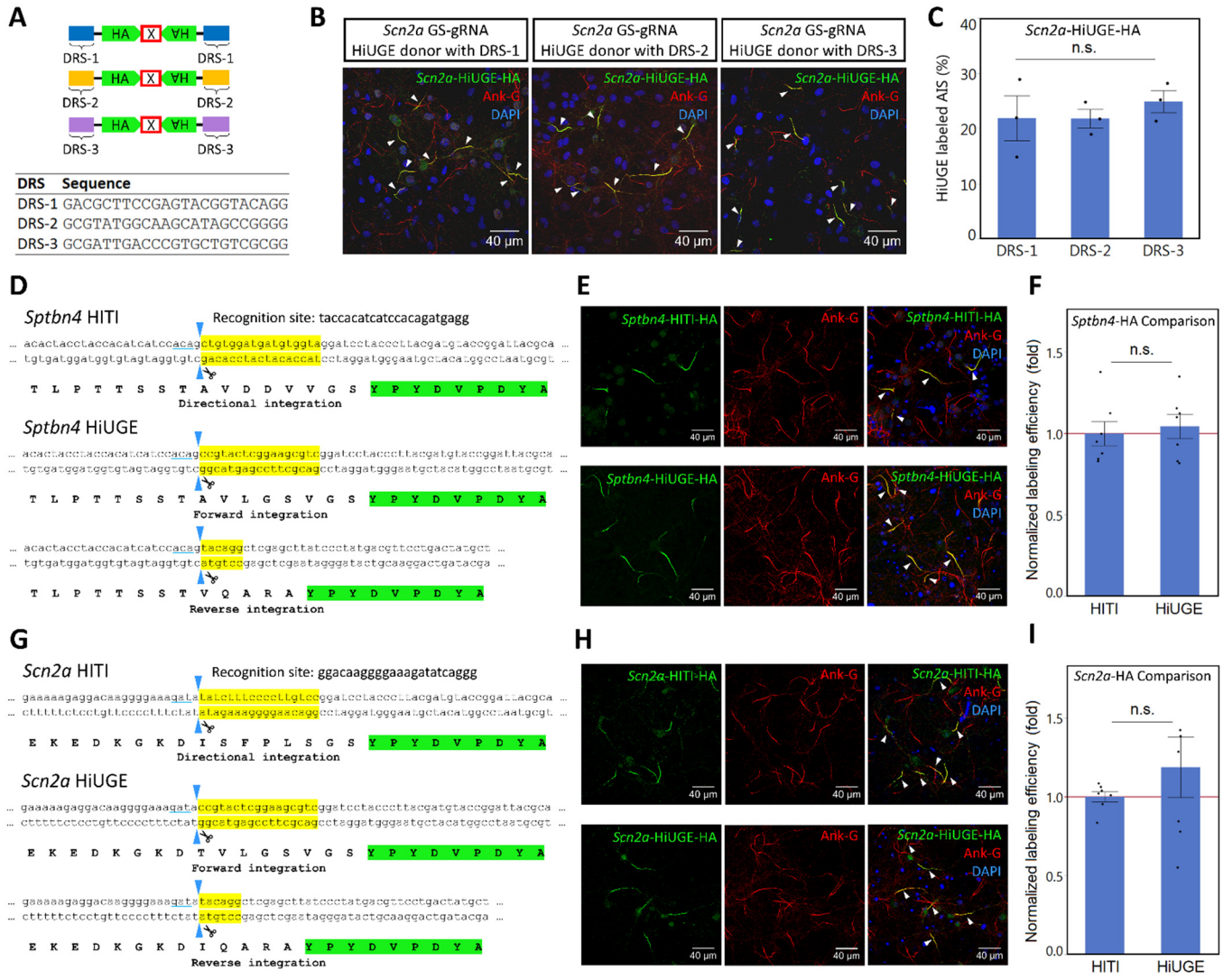

Figure S2. Additional Data of the Comparison Between Different Donor Recognition Sequences and Between HiUGE and HITI, Related to Figure 1 and Figure 2.

(**A**) Three HA-epitope donor vectors containing different HiUGE DRS sequences are compared in an experiment to assess their effectiveness for protein labeling. (**B**) Representative images of HiUGE-mediated HA-epitope labeling of NaV1.2 (encoded by mouse *Scn2a*) using donor vectors with three different DRS sequences (DRS1-3) are shown. Equal amount of virus (5×10^{10} GC / mL per virus in the culture medium) was used. Consistent labeling pattern specific to the axonal initial segment (AIS) is evident. Total AIS structures stained with an AIS-marker (Ank-G) is shown (red), which colocalizes with the HA-labeling (green, arrowheads). (**C**) Quantification result showing similar cellular labeling efficiencies using donor vectors with DRS1-3 (n.s.: $p > 0.05$, one-way ANOVA, n=3). (**D**) Illustration of the genomic sequences following HITI or HiUGE editing of mouse *Sptbn4*. Parts of donor recognition sequences (yellow), inserted HA-epitope peptide (green), and the last coding triplets before the junction (underlined) are highlighted. (**E**) Representative immunofluorescent images showing labeling pattern specific to the AIS following both HITI and HiUGE editing of mouse *Sptbn4*. (**F**) Quantification results

showing comparable cellular labeling efficiencies after HITI or HiUGE editing of mouse *Sptbn4* (n.s.: $p > 0.05$, two-tailed *t*-test, n=7). (**G**) Illustration of the genomic sequences following HITI or HiUGE editing of mouse *Scn2a*. Parts of donor recognition sequences (yellow), inserted HA-epitope peptide (green), and the last coding triplets before the junction (underlined) are highlighted. (**H**) Representative immunofluorescent images showing labeling pattern specific to the AIS following both HITI and HiUGE editing of mouse *Scn2a*. (**I**) Quantification results showing comparable cellular labeling efficiencies after HITI or HiUGE editing of mouse *Scn2a* (n.s.: $p > 0.05$, two-tailed *t*-test, n=7). Error bars represent standard error of the mean (SEM). Scale bar is indicated in each panel.

Figure S3

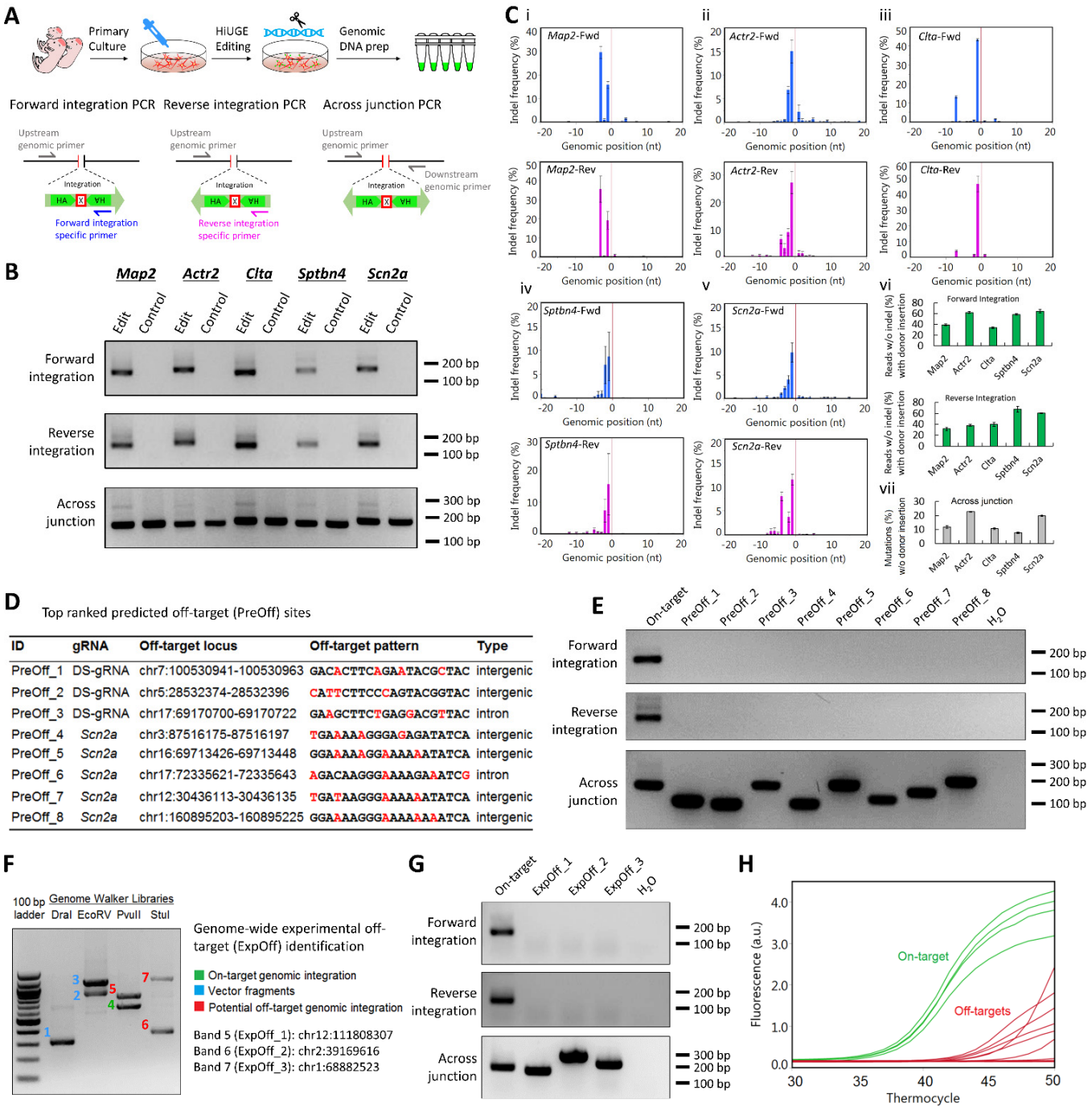

Figure S3. **Assessment of the Indel Rate and the Off-target Integration of the Donor, Related to Figure 2.**

(**A**) Schematic design of genomic PCR to detect dual-orientation HA-epitope payload integration into various genomic loci. Genomic DNAs were extracted from primary neuronal cultures either with or without HiUGE editing. PCR reactions were performed using upstream genomic forward primers of 5 genes (mouse *Map2*, *Actr2*, *Clta*, *Sptbn4*, and *Scn2a*), paired with either a reverse primer specific for the forward payload integration event, or a reverse primer specific for the reverse payload integration event. (**B**) Insert-specific PCRs for both forward and reverse payload integration showed positive bands (~150-

200 bp) in edited samples, compared to no band in negative controls. Across junction PCR of each genomic locus is also shown. Predicted sizes of amplicons (*Map2*, *Actr2*, *Clta*, *Sptbn4*, and *Scn2a*, from left to right respectively) were: forward integration: 168, 193, 156, 168, 174 bp; reverse integration: 157, 182, 145, 157, 163 bp; across junction: 159, 157, 166, 159, 182 bp. (**C**) Analysis of indel frequencies by deep sequencing the PCR products from (**B**). Detections of indel positions (i-v) showed that most indel events initiated immediately adjacent to the edited junction (denoted by 0 on the x-axis, red line). The values on x-axis denote the onset positions of indels called by GATK HaplotypeCaller, with the inserted or deleted sequences immediately succeeding that position. Proportions of reads without an indel from the insert-specific amplicons are also plotted (vi), showing the estimated frequencies of payload integrations without an indel across five genes in both orientations. In addition, the frequencies of allelic mutations without donor integration are estimated at each edited locus by deep sequencing the across junction PCR products (vii). (**D**) Top ranked CRISPOR-predicted off-target loci for both DS-gRNA and *Scn2a* GS-gRNA. (**E**) Genomic PCR reactions using gene-specific primers paired with payload-specific primers successfully detected on-target integrations, while the genomic integrations of the payload were undetected for the predicted off-target sites (PreOff_1-8). Across junction PCR reactions showed robust and specific amplifications using these genomic primers. Predicted sizes of amplicons (On-target and PreOff_1 - 8, from left to right respectively) were: forward integration: 174, 147, 164, 194, 173, 223, 178, 216, 255 bp; reverse integration: 163, 136, 153 ,183, 162, 212, 167, 205, 244 bp; across junction: 182, 117, 105, 188, 110, 199, 131, 165, 227 bp. (**F**) Genome Walker experiment using nested two-round PCR detected on-target integration (band 4), and 3 potential off-target integrations into the non-coding genomic regions (band 5-7). Vector fragments were also detected as expected (band 1-3). (**G**) Genomic PCR reactions using gene-specific primers paired with insert-specific primers successfully detected on-target integrations, while the genomic integrations of the payload were undetected for the experimentally identified potential off-target sites (ExpOff_1-3). Across junction PCR reactions showed robust and specific amplifications using these genomic primers. Predicted sizes of amplicons (On-target and ExpOff_1 - 3, from left to right respectively) were: forward integration: 174, 172, 287, 183 bp; reverse integration: 163, 161, 276, 172 bp; across junction: 182, 158, 309, 220 bp. (**H**) Comparison of real-time PCR amplification curves of on-target integrations (green) versus potential off-target integrations (red). These results are qualitative due to the possible bias of different primers. DNA sizes are marked in reference to the ladder. Genomic positions are in reference to the GRCm38 (mm10) assembly.

Figure S4

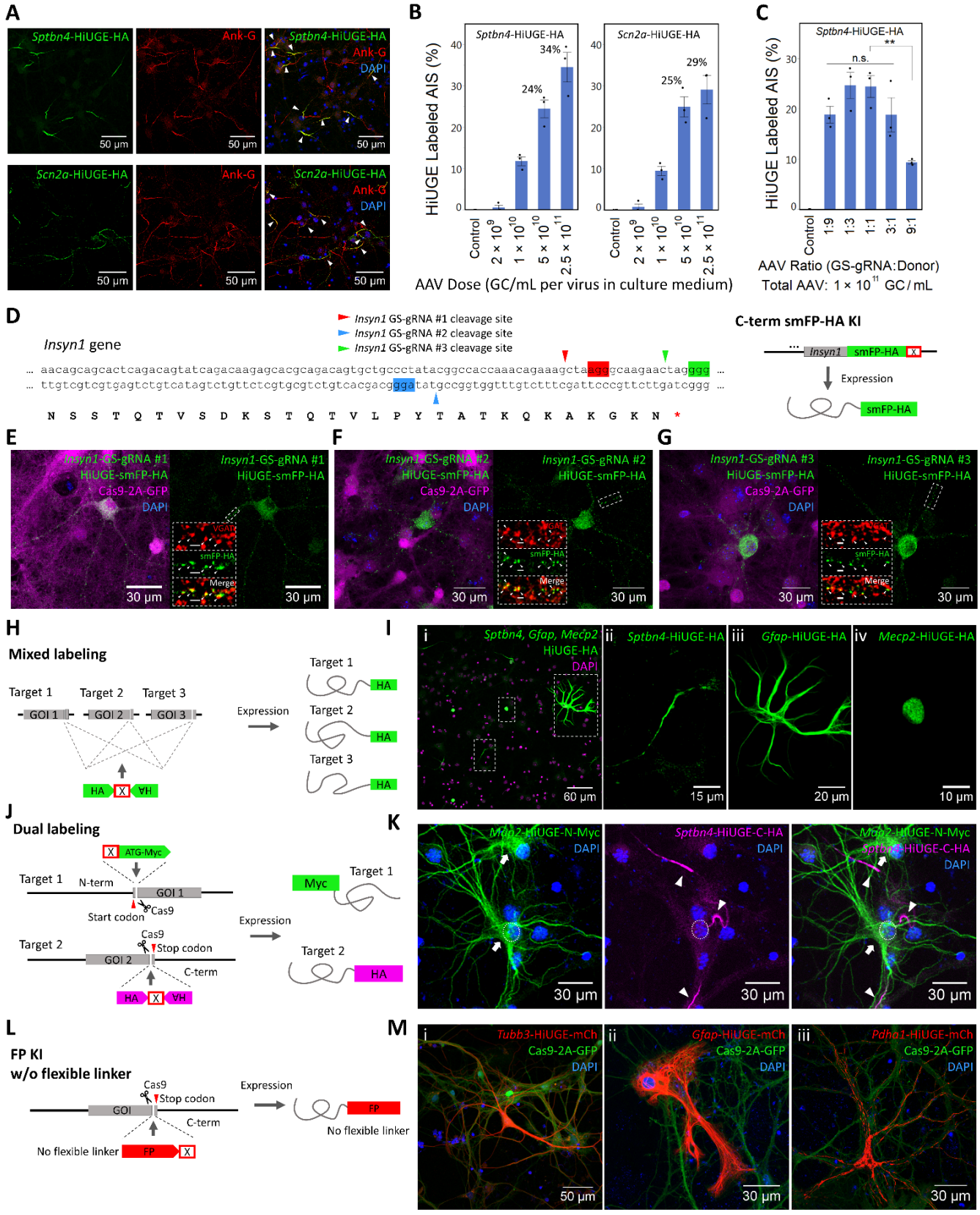

Figure S4. Additional Data of the Localization Mapping Application Using HiUGE, Related to Figure 2 and Figure 6.

(**A**) Representative images of HiUGE labeling of AIS proteins βIV-Spectrin and NaV1.2 by C-term HA-epitope KI to mouse *Sptbn4* and *Scn2a*, at high AAV concentrations in primary neuronal culture (2.5 ×

10^{11} GC / mL per virus in culture medium). Immunofluorescent staining with an antibody against AIS-marker Ankyrin-G (Ank-G) is also shown. Arrowheads represent the HiUGE-labeled AIS structures. (**B**) Quantification results showing the estimated efficiencies of cellular labeling across several AAV concentrations at 1:1 virus ratio (GS-gRNA : donor). Efficient labeling was achieved at a dose of 5×10^{10} GC / mL per virus (*Sptbn4*: 24.4 ± 2.2 %; *Scn2a:* 24.9 ± 2.4 %), or 2.5×10^{11} GC / mL per virus (*Sptbn4*: 34.5 ± 3.6 %; *Scn2a:* 29.0 ± 3.4 %). (**C**) Quantification results showing the estimated efficiencies of cellular labeling across several ratios of AAVs (GS-gRNA : donor) at 1×10^{11} GC / mL combined viral concentration in the culture medium. Ratios of 1:9, 1:3, 1:1 and 3:1 were largely equivalent in labeling efficiency (n.s.: $p > 0.05$, one-way ANOVA followed by Tukey-Kramer HSD *post hoc* test, n=3), suggesting a broad range of acceptable viral ratios. However, cellular labeling efficiency of 9:1 ratio is significantly lower compared to 1:1 ratio (**: $p < 0.01$, one-way ANOVA followed by Tukey-Kramer HSD *post hoc* test, n=3), suggesting that sufficient donor AAV is required for efficient labeling. (**D**) Schematic illustration of HiUGE C-term smFP-HA KI to mouse *Insyn1*, directed by three different GS-gRNAs. Shaded DNA triplets denote the protospacer adjacent motif (PAM) of the target sequence for each GS-gRNA. Arrowheads denote the Cas9 cleavage sites. (**E-G**) Sucessful and comparable punctate labeling at inhibitory synapses was obsereved across all three GS-gRNAs. Colocalization of the HA-immunoreactivity with the juxtaposed inhibitory presynaptic marker vesicular GABA transporter (VGAT) immunosignal is shown in the insets (arrowheads). (**H**) Various GS-gRNAs targeting different genes can be combined and applied simultaneously with a single HiUGE donor to create mixed labeling. (**I**) Simultaneous labeling of ßIV-spectrin, GFAP, and MeCP2 by pooled application of GS-gRNAs targeting mouse *Sptbn4, Gfap,* and *Mecp2* genes, together with the HA-epitope donor. Panels (ii-iv) are enlarged views of the boxed regions in (i). (**J**) Schematic illustration of HiUGE amino-terminus (N-term) KI construct to achieve dual-labeling of two different targets by differentially targeting the N-term or the C-term of two different proteins. The stop codon cassette within the N-term HiUGE donor vector ensures that in the event of integration into the C-term, the translation will terminate upstream of the Myc-epitope, thus this payload is selective for N-term expression. (**K**) Representative dual-labeled immunostaining image of mouse *Map2* and *Sptbn4* encoded proteins by N-term KI of Myc-epitope to *Map2* (MAP2, arrows), and C-term KI of HA-epitope to *Sptbn4* (ßIV-spectrin, arrowheads). Dashed circle represents the soma of the dual-labeled neuron. (**L**) Fluorescent protein (FP) payloads can be directly fused with endogenous targets (in contrast to being separated by a flexible $(GGGGS)_4$ linker, Figure 6A), enabling flexible selection of linker sequences using HiUGE. (**M**) Representative images of direct mCherry (mCh) fluorescence of HiUGE labeled (i) βIII-tubulin, (ii) GFAP, and (iii) Pyruvate dehydrogenase E1 alpha, using a mCh payload without flexible linker. Scale bar is indicated in each panel, or within insets (2µm).

Figure S5

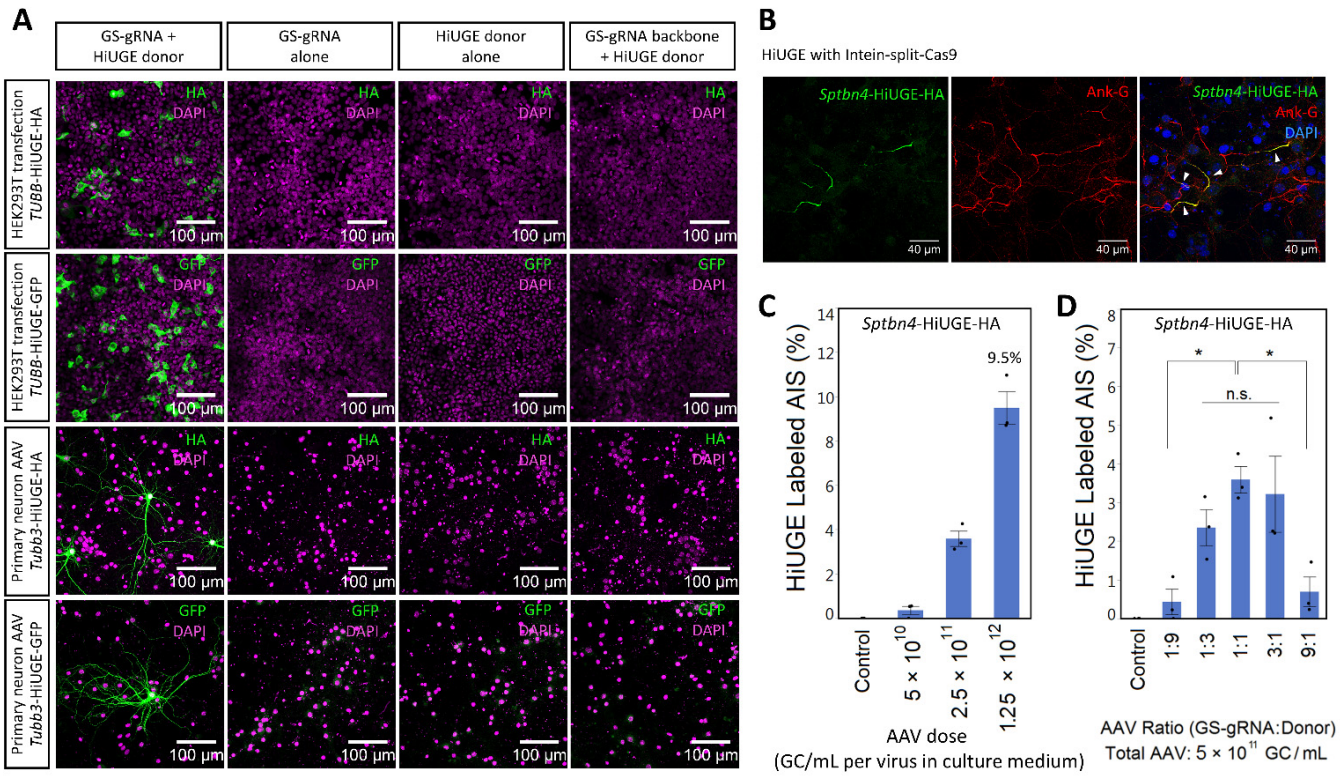

Figure S5. **Additional Data of HiUGE Vectors with Intein-Split-Cas9, Related to Figure 7.**

(**A**) Control experiments for intein-split-Cas9 mediated HiUGE vectors in HEK293T cells and WT primary neurons. HEK293T Cells were plasmid transfected and primary neurons were AAV transduced with HiUGE GS-gRNAs and donors to knock-in HA-epitope or GFP payloads to human *TUBB* or mouse *Tubb3* genomic loci. Experimental conditions are indicated in the figure. Positive HA-epitope or GFP immunoreactivity showing tubulin-like expression pattern was only found when GS-gRNAs were paired with the corresponding HiUGE donors (*leftmost column*). No HA-epitope or GFP KI was detected when GS-gRNA or HiUGE donor was applied alone, or when an empty GS-gRNA backbone was paired with the donor. (**B**) Representative images showing labeling of βIV-spectrin (*Sptbn4*) using HiUGE with intein-split-Cas9 at the viral dose of 1.25×10^{12} GC / mL per virus in the culture medium. Arrowheads represent the HiUGE-labeled AIS structures. (**C**) Quantification results showing the estimated efficiencies of cellular labeling across several AAV concentrations at a 1:1 virus ratio (GS-gRNA : donor). Labeling efficiency of 9.5 ± 0.7 % was achieved at the dose of 1.25×10^{12} GC / mL per virus in culture medium. (**D**) Quantification results showing the estimated efficiencies of cellular labeling across several ratios of AAVs (GS-gRNA : donor) at 5×10^{11} GC / mL combined viral concentration in the culture medium. At ratios of 1:3, 1:1 and 3:1 there were no significant differences in labeling efficiency (n.s.: $p > 0.05$, one-way ANOVA followed by Tukey-Kramer HSD *post hoc* test, n=3). However, cellular labeling efficiencies at 1:9 or 9:1 ratio were significantly lower compared to the 1:1 ratio (*: $p < 0.05$, one-way ANOVA

followed by Tukey-Kramer HSD *post hoc* test, n=3). Error bars represent standard error of the mean (SEM). Scale bar is indicated in each panel.