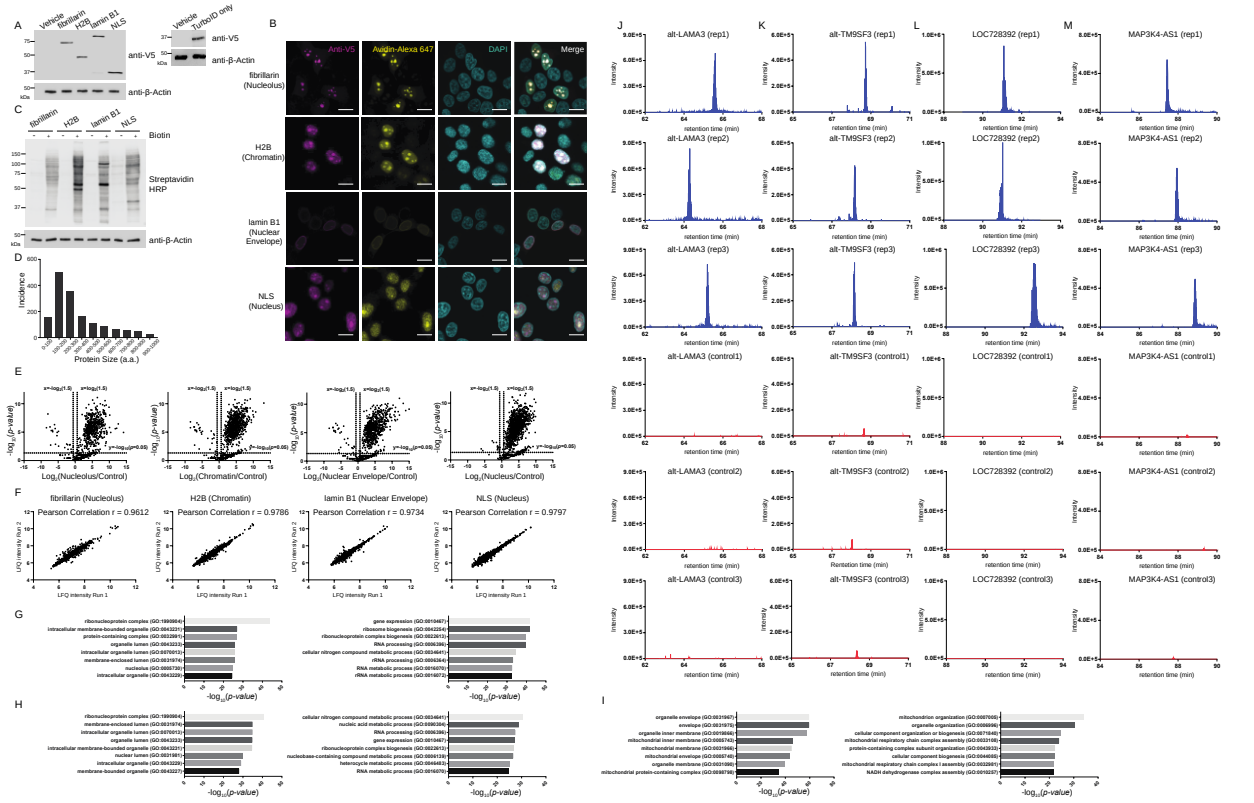

Figure S1. Proximity biotinylation of canonical proteins, and unannotated microproteins and alternative proteins, in cell culture by MicroID: validation of the method (related to Figure 1). (A) Western blotting of HEK 293T lysates following transduction with lentivirus targeting V5-tagged TurboID to different subcellular compartments via genetic fusions. Nucleolus: TurboID-fibrillarin; Chromatin: TurboID-H2B (histone H2B); Nuclear envelope: TurboID-lamin B1; Nucleus: TurboID-3xNLS. untargeted: TurboID only. (B) Immunofluorescence of biotin-treated HEK 293T cell lines stably expressing V5-tagged TurboID in the indicated subcellular compartment. Scale bar, 10 μm. (C) Streptavidin blotting of cell lysates from HEK 293T cell lines expressing TurboID fused to the indicated proteins for subcellular targeting, which were then treated with vehicle (–) or biotin (+). Actin serves as a loading control. (D) Cumulative length distribution of annotated proteins detected in combined MicroID experiments in HEK 293T cells (all TurboID constructs compared with untargeted control). 67% are smaller than 300 amino acids. (E) Volcano plot of proteins enriched by TurboID fusions with fibrillarin (nucleolus)/H2B (chromatin)/Lamin B1 (nuclear envelope)/3xNLS (nucleus) against control cells expressing untargeted TurboID protein, with label-free quantiative (LFQ) proteomics (N = 3, biological replicates). (F) Pearson correlation of label-free quantification protein intensities in two biological replicate run of each subcellar compartment, respective. (G)-(I) GO (cellular component and biological processes) analysis of proteins <300 amino acids enriched (fold change ⩾ 30) in each subnuclear compartment (nucleolus(G), chromatin(H) and nuclear envelope(I)). (J)-(M) Confirmation of differential expression of microproteins and alt-proteins with extracted ion chromatograms (EICs). (J) For alt-LAMA3, peptide: PGRGGEDLGHR, observed mass: 384.2028, mass window for EIC: 384.20-384.21, from TurboID-fibrillarin rep1/rep2/rep3 versus untargeted TurboID rep1/rep2/rep3 (control); (K) For alt-TM9SF3, peptide: AVAAAAAAAPDPGGR, observed mass: 633.3333, mass window for EIC: 633.33-633.34, from TurboID-H2B rep1/rep2/rep3 versus untargeted TurboID rep1/rep2/rep3 (control); (L) For LOC728392, peptide: GLEQIRPDPESEGLFDKPPPEDPPAAR,

observed mass: 740.1235, mass window for EIC: 740.12-740.13, from TurboID-lamin B1 rep1/rep2/rep3 versus untargeted TurboID rep1/rep2/rep3 (control); (M) For MAP3K4-AS1, peptide: PSGPTEFGPGPAPLSASDR, observed mass: 920.4468, mass window for EIC: 920.44-920.45, from TurboID-3xNLS rep1/rep2/rep3 versus untargeted TurboID rep1/rep2/rep3 (control).

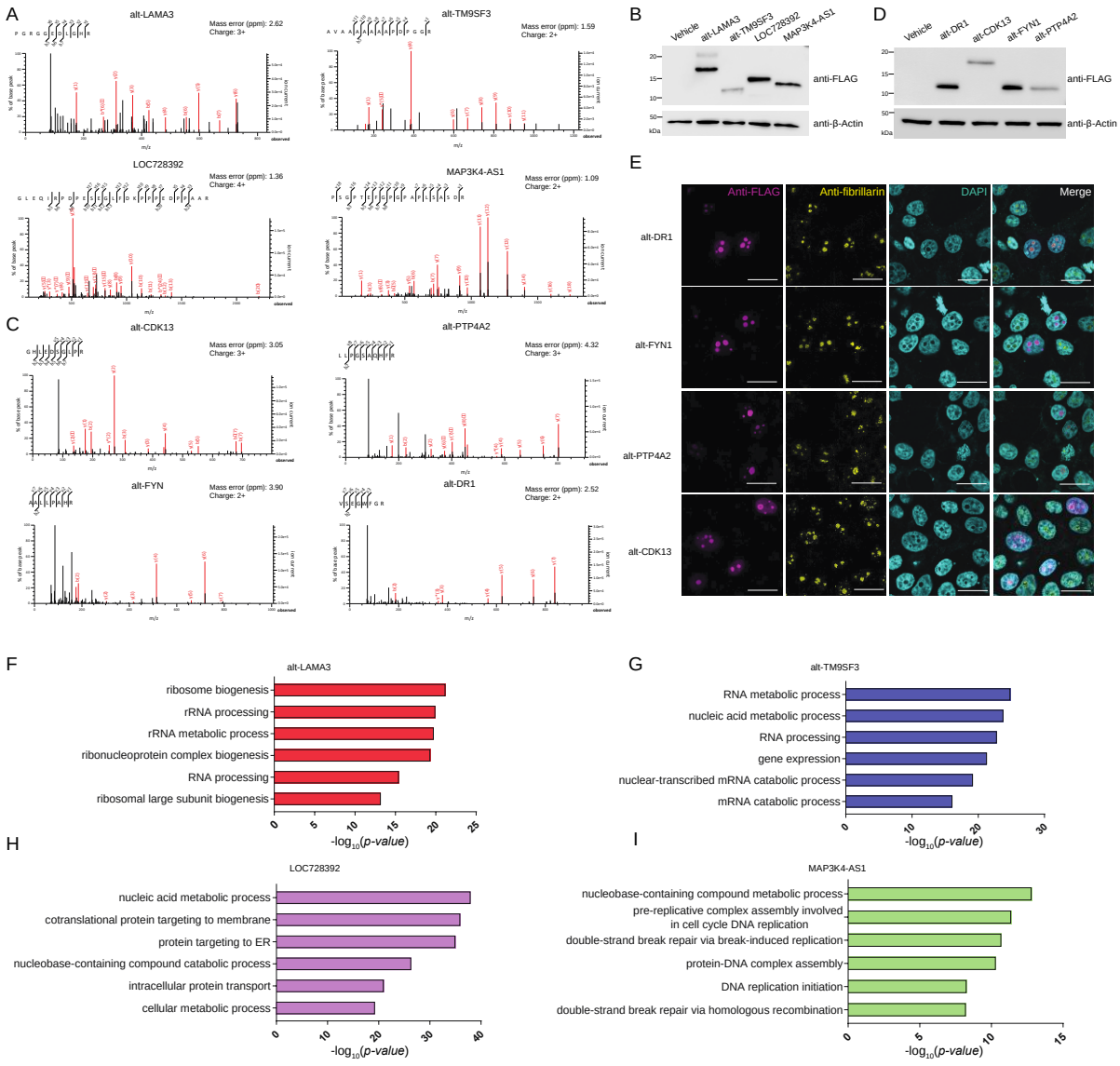

Figure S2. MS/MS spectra of unannotated microproteins and alt-proteins identified with MicroID (related to Figure 2). (A) MS/MS spectra of selected tryptic peptides matched to unannotated microproteins and alt-proteins detected by MicroID in each compartment profiled. Alt-LAMA3 was specifically identified using TurboID-fibrillarin, alt-TM9SF3 with TurboID-H2B, LOC728392 with TurboID-lamin B1 and MAP3K4-AS1 with TurboID-3xNLS. (B) Cas9-directed homology repair was applied to append epitope tags to a genomic copy of each indicated microprotein or alt-protein in independent HEK 293T cell lines, followed by Western blotting to confirm endogenous expression and molecular weight. Wild-type HEK 293T lysate served as a negative control. (C) MS/MS spectra of selected tryptic peptides (alt-DR1, alt-CDK13, alt-FYN and alt-PTP4A2) matched to unannotated alt-proteins detected by MicroID in the nucleolus (TurboID-fibrillarin). (D) Overexpression of constructs containing the full-length coding sequencing of alt-CDK13, alt-FYN, alt-DR1 and alt-PTP4A2 with a FLAG tag appended to the C-terminus of each putative alt-protein in HEK293T cells, followed by Western blotting to confirm expression and molecular weight. Untransfected HEK 293T lysate served as a negative control.

(E) HEK 293T cell lines overexpressing constructs containing the full-length coding sequencing of alt-CDK13, alt-FYN, alt-DR1 and alt-PTP4A2 with a FLAG tag appended to the C-terminus, and subjected to immunofluorescence with anti-FLAG tag (magenta), colocalization marker for nucleolus (yellow), and DAPI (cyan). Scale bar, 10 μm. Data are representative of three biological replicates. (F-I) GO (biological processes) analysis of genes enriched (fold change $\geqslant$ 30, PD/Control) in microprotein/alt-protein co-IPs.

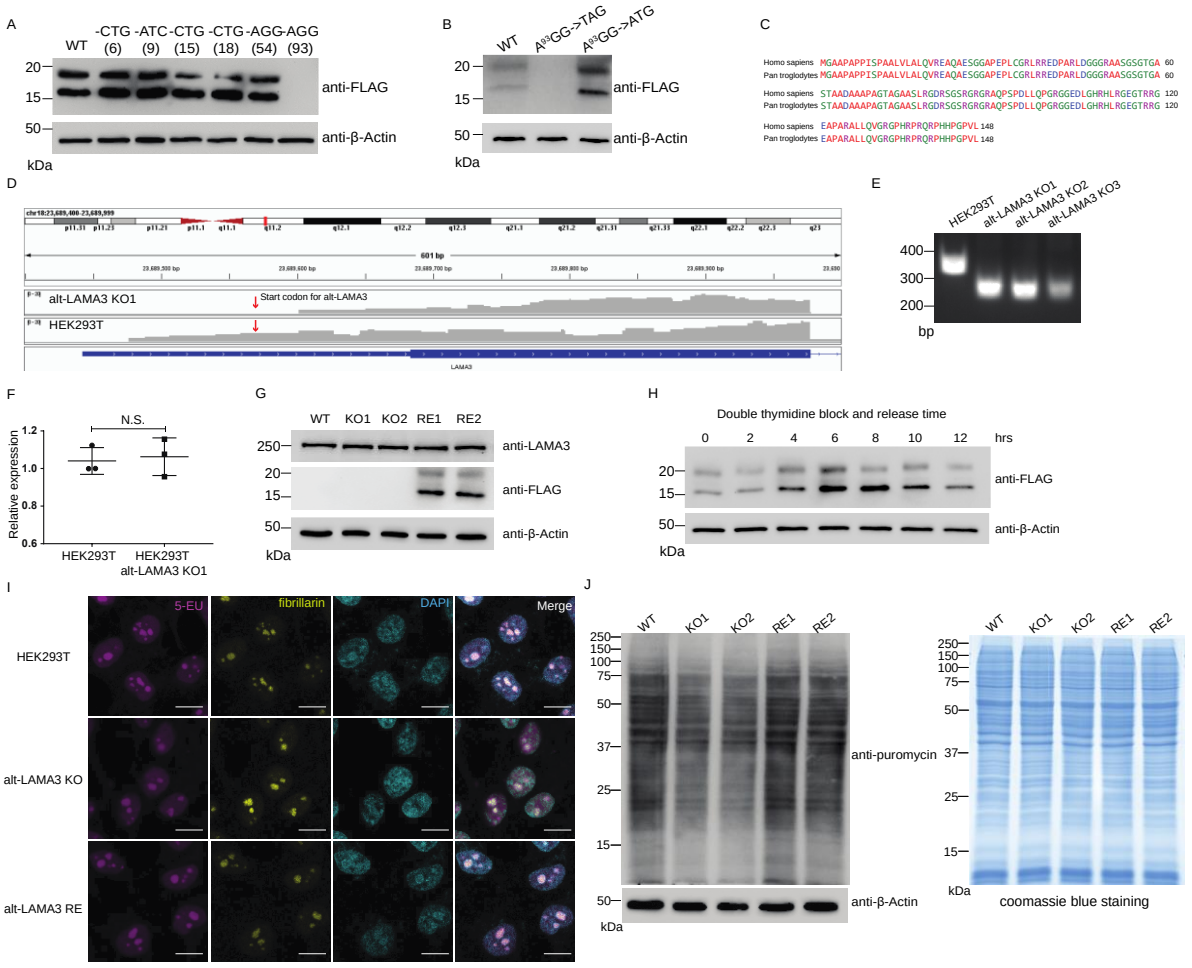

Figure S3. **Alt-LAMA3 is physically and functionally associated with pre-rRNA transcription and required for global protein synthesis (related to Figure 3).** (A) Transient over-expression of a construct containing the full 5′UTR and alt-LAMA3 coding sequence derived from human *LAMA3* transcript variant 1, with a FLAG tag appended to the C-terminus of alt-LAMA3, in HEK 293T cells, was followed by lysis and Western blotting with the antibodies indicated to the right. WT, wild-type alt-LAMA3 sequence; lanes 2-6 indicate candidate start codon sequences above and first nucleotide sequence positions below (number relative to the first nucleotide of the cDNA of *LAMA3* transcript variant 1, NM_198129.4) which were deleted. (B) Transient expression of a construct containing the full 5`UTR and alt-LAMA3 coding sequence derived from human LAMA3 transcript variant 1 ($A^{93}GG$ mutated to TAG and $A^{93}GG$ mutated to ATG), with a FLAG tag appended to the C-terminus of alt-LAMA3, in HEK 293T cells, was followed by lysis and Western blotting. (C) Predicted alt-LAMA3 homolog conservation between *Homo sapiens* and *Pan troglodytes*. (D) RNA-seq reads mapping to the *LAMA3* genomic locus in wild-type (WT) HEK 293T and alt-LAMA3 knock-out (KO) cell line 1 (alt-LAMA3 KO1) . The y-axis scale for RNA is reads per kilobase per million reads (y-axis scale, 0-50 RPKM). Above, genomic coordinates. (E) PCR confirmation of the genomic deletion within *LAMA3* exon 1 generating alt-LAMA3 KO cell lines. (F) Quantitative RT-PCR with primers specific to *LAMA3* transcript variant 1. (Error bars ± the standard deviation, N = 3 biologically independent samples, n.s. no significance). (G) Western blot analysis of wild-type (WT), alt-LAMA3 KO HEK 293T cells with the indicated antibodies. Data are representative of three biological replicates. (H)

Western blotting analysis of synchronized alt-LAMA3 KI cells released from the G1/S boundary at the indicated time points with antibodies indicated on the right. All data are representative of three biological replicates. (I) 5-ethynyluridine (5-EU) incorporation assay. HEK 293T, HEK 293T alt-LAMA3 KO and HEK 293T alt-LAMA3 RE cells were incubated with 1mM 5-EU for 1 h, followed by in situ click chemistry with Alexa Fluor™ 488 Azide and anti-fibrillarin immunofluorescence. (J) Global protein synthesis assay by puromycin tagging. HEK 293T, HEK 293T alt-LAMA3 KO and HEK 293T alt-LAMA3 RE cells were treated with 1 μM puromycin for 1 h to label nascent peptides, followed by Western blotting with an anti-puromycin antibody.

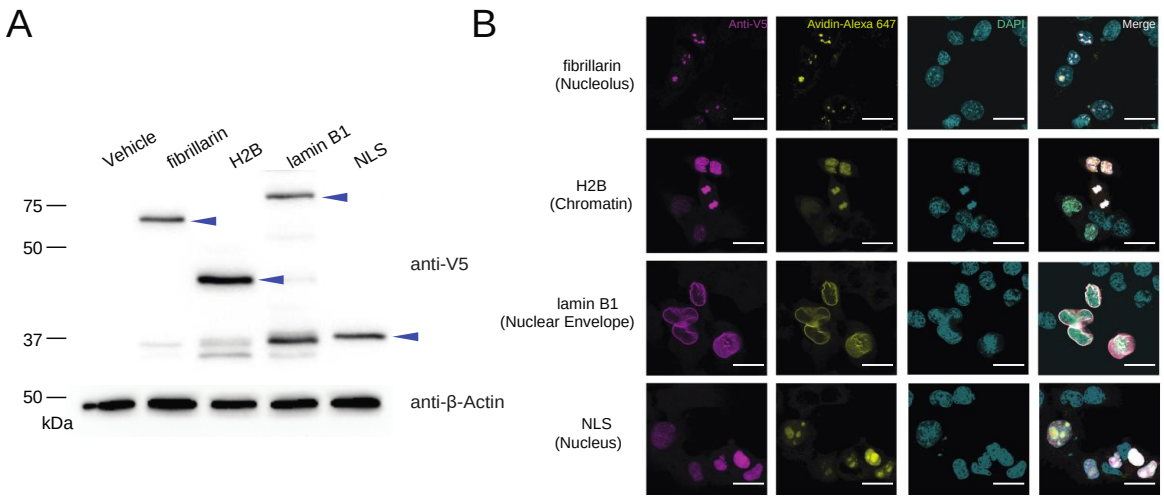

Figure S4. **Application of MicroID in mouse Hepa 1-6 cell line: validation of the method (related to Figure 4)**. (A) Western blotting of a panel of Hepa1-6 cells following transduction with AAV9 targeting V5-tagged TurboID to the indicated subcellular compartment via genetic fusion. Nucleolus: TurboID-fibrillarin; Chromatin: TurboID-H2B; Nuclear envelope: TurboID-lamin B1; Nucleus: TurboID-3xNLS. Blue carat indicates expected TurboID fusion protein size. (B) Immunofluorescence of stable cell lines transduced with AAV9 targeting TurboID to the indicated subcellular region. Scale bar, 10 μm.

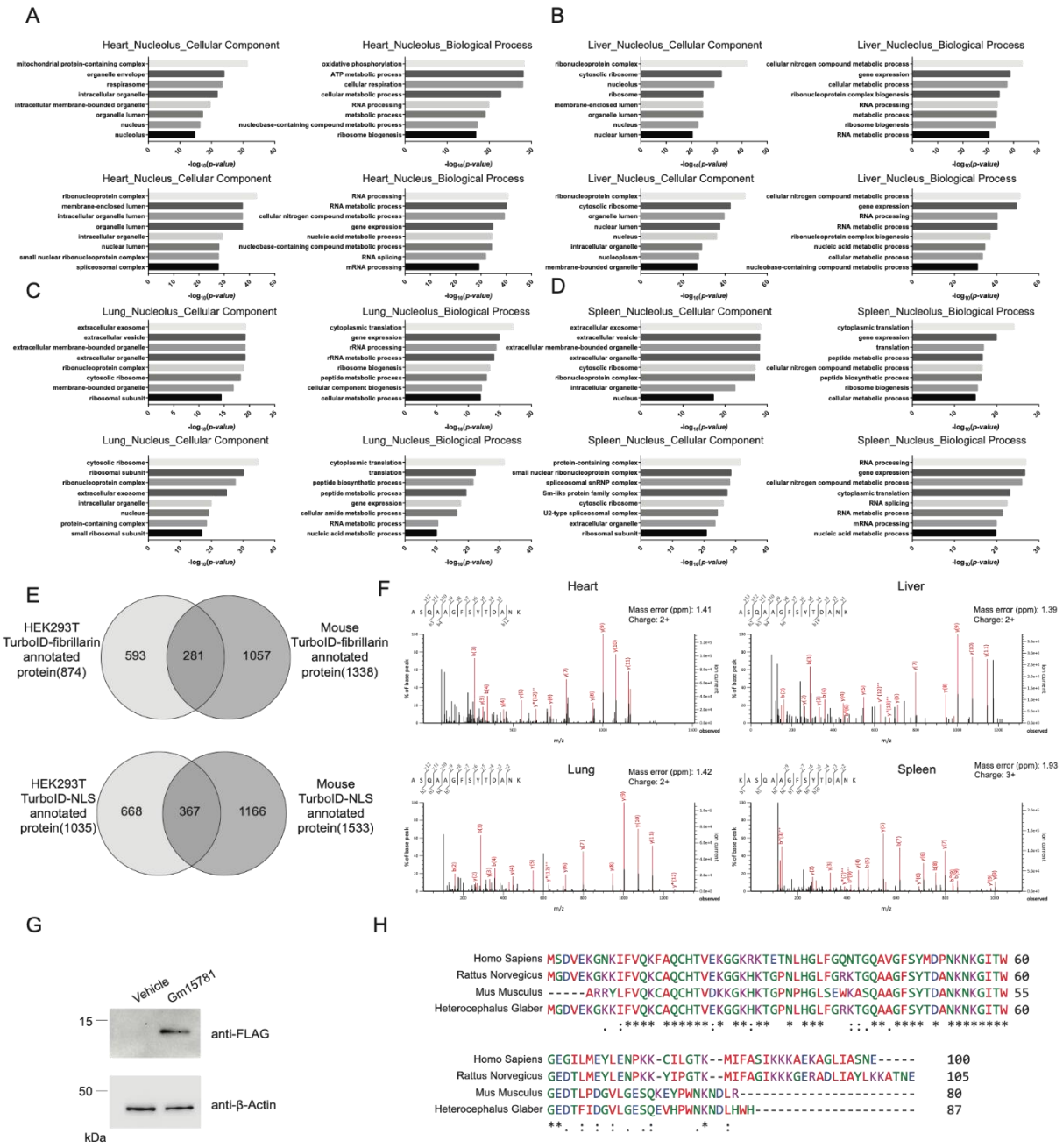

Figure S5. **Data analysis of proximity biotinylation of canonical proteins, and unannotated microproteins and alt-proteins, *in vivo* by MicroID (related to Figure 5)**. (A-D) GO (cellular component and biological processes) analysis of genes enriched (fold change ⩾ 30) in small proteins identified using nucleolus- and nucleus-targeted MicroID in murine (A) heart; (B) liver; (C) lung; (D) spleen. (E) Venn diagrams compile the comparison of the annotated protein between HEK 293T TurboID-fibrillarin and mouse TurboID-fibrillarin, HEK 293T TurboID-3xNLS and mouse TurboID-3xNLS, respectively. (F) MS/MS spectra of the tryptic peptide from microprotein Gm15781 detected by MicroID peptidomics in different tissues (heart, liver, lung and spleen). (G) Over-expression of a construct containing the full 5'UTR and Gm15781 coding sequence with a FLAG tag appended to the C-terminus of Gm15781 in

Hepa1-6 cells, followed by lysis and Western blotting. (H) Protein-level conservation of *in silico* translated Gm15781 transcript homologs from *Homo sapiens* (hCG1644511), *Rattus norvegicus* (XM_003751799.4), *Mus musculus* (AC132412.2) and *Heterocephalus glaber* (XM_004864407.3).