

Supplemental information

**Rapid UPF1 depletion illuminates the temporal
dynamics of the NMD-regulated human transcriptome**

Volker Boehm, Damaris Wallmeroth, Paul O. Wulf, Oliver Popp, Luiz Gustavo Teixeira Alves, Lucie Reinecke, Maximilian Riedel, Emanuel Wyler, Marek Franitza, Kerstin Becker, Karina Polkovnychenko, Simone Del Giudice, Nouhad Benlasfer, Philipp Mertins, Markus Landthaler, and Niels H. Gehring

Figure S1

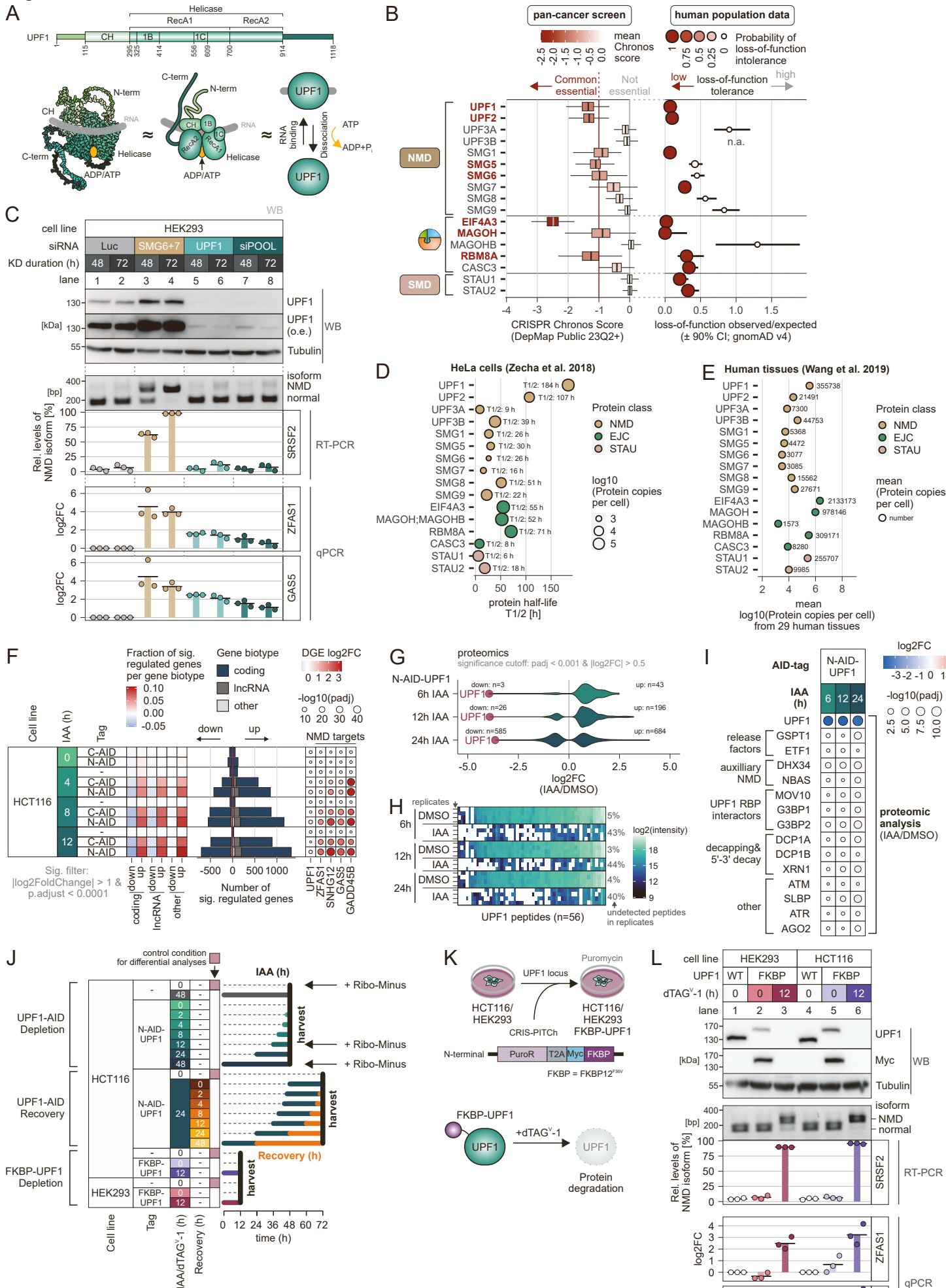


Figure S1. Characterization of UPF1 downregulation via conditional degron tags, related to Figure 1.

- (A) (Top) Schematic domain representation of UPF1 (based on UniProt ID Q92900-2). (Bottom) Scheme of UPF1 RNA-binding function and folded protein domains based on AlphaFold prediction (AF-Q92900-F1-model_v4) and UPF1-RNA-ADP:AlF₄ crystal structure (PDB ID 2xzo).
- (B) Gene essentiality scores from a pan-cancer screen (DepMap) or loss-of-function observed/expected ratios (\pm 90% confidence interval; CI) from human population data (gnomAD v4) are shown for individual NMD, EJC and SMD factors. The Chronos score is depicted as boxplot, the box extends to the 25th and 75th percentile with the median in bold line, outliers are not shown.
- (C) (Top) Western blot showing levels of UPF1 protein after knockdown in HEK293 cells with different siRNAs for 48 or 72h, Tubulin serves as loading control, o.e. = overexposure. (Middle) Detection of the NMD target SRSF2 by end-point RT-PCR. The relative mRNA levels of SRSF2 isoforms were quantified. (Bottom) Probe-based qPCR of NMD targets ZFAS1 and GAS5, shown as log₂ fold change (log₂FC). Individual data points and mean (bars) are shown (n=3).
- (D) Protein half-lives and copy numbers from HeLa cells determined by mass spectrometry (Zecha et al. 2018) for NMD, EJC and STAU factors are shown.
- (E) Mean protein copies for NMD, EJC and STAU factors from 29 human tissues determined by mass spectrometry (Wang et al. 2019).
- (F) (Left) Comparison of RNA-Seq data from untagged, C- or N-terminal AID-tagged UPF1 HCT116 cell lines, treated with 500 μ M IAA for the indicated time regarding the fraction of significantly differentially expressed genes (separated by up-/downregulation) versus all detected genes per GENCODE biotype. (Middle) Absolute number of significantly regulated genes stratified by GENCODE biotype. (Right) Expression changes of UPF1 or individual NMD target genes.
- (G) Violin plot depicting the proteomic analysis of significantly differentially expressed proteins in the indicated conditions and timepoints, comparing IAA-treated with control cells. The number of significant up- or downregulated proteins is indicated as n; UPF1 is shown as dark magenta points.
- (H) Intensities of UPF1 peptides detected by mass spectrometry in the indicated conditions depicted for each replicate.
- (I) Differential protein expression levels of additional UPF1 interacting factors in IAA-treated versus DMSO control conditions, determined by mass spectrometry.
- (J) Overview of experimental timeline, highlighting the timing of IAA/ dTAG^V-1 addition, if applicable recovery time and the point of sample harvesting. Control conditions per dataset are indicated.
- (K) Schematic representation of cell line generation allowing the dTAG^V-1-inducible depletion of UPF1. For details see Methods.
- (L) Western blot, RT-PCR and qPCR analyses of N-terminal FKBP-tagged UPF1 HEK293 or HCT116 cell lines treated with 0.25 μ M dTAG^V-1 for the indicated time, visualized as described in panel (C). Myc-FKBP-tagged UPF1 was also detected via Myc antibodies. Tubulin serves as loading control.

A

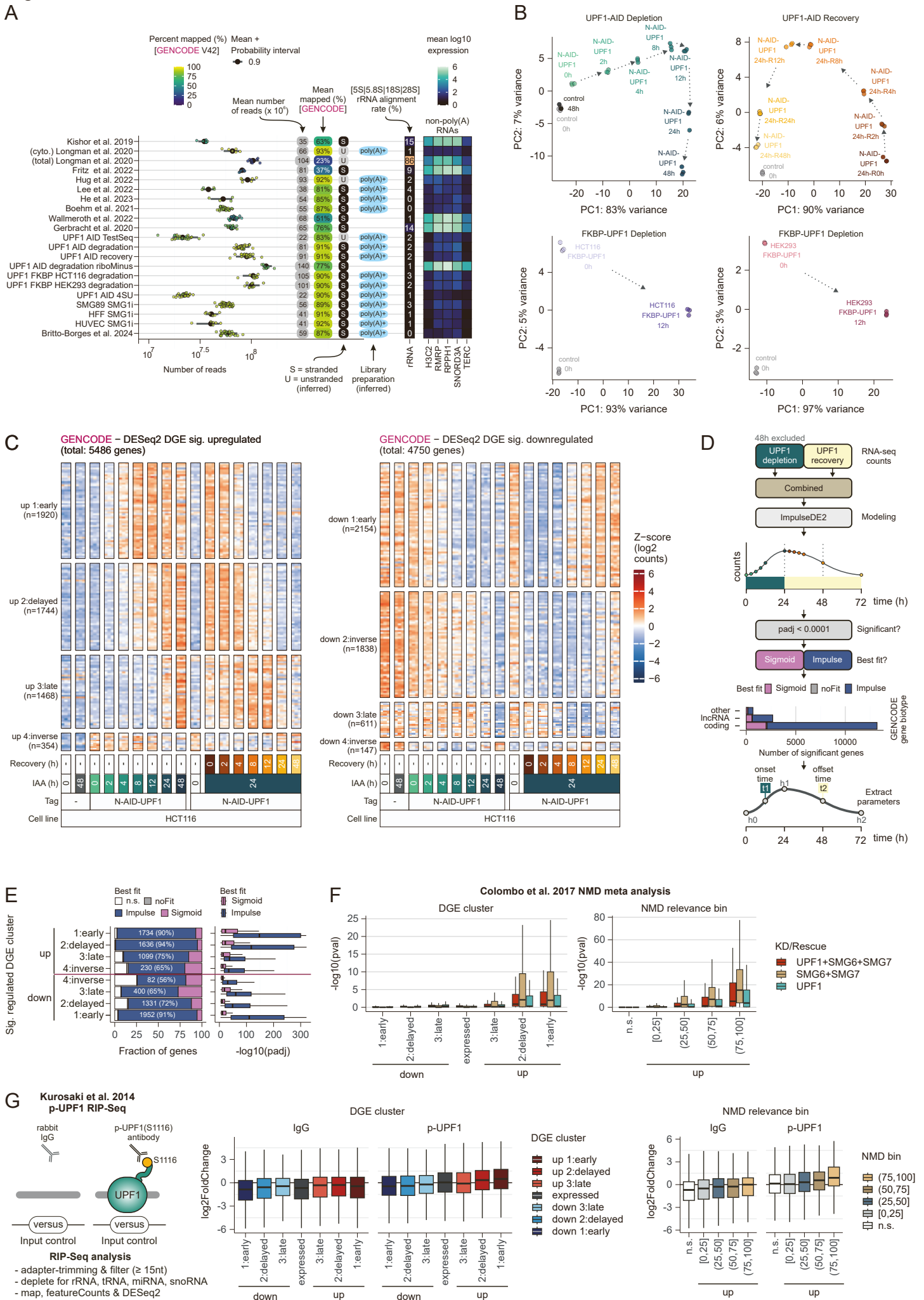


Figure S2. Identification of distinct classes of UPF1-regulated transcripts, related to Figure 2.

- (A) Technical characteristics of all short-read RNA-Seq datasets analyzed in this study, including published as well as newly generated datasets. Number of reads per replicate are shown as points, mean reads per dataset are indicated in gray boxes and percentage of mapped reads (GENCODE V42 annotation) by Salmon is indicated by fill color. Library strandness was inferred by Salmon and RNA selection method was inferred from mean log10 expression of non-poly(A) RNAs. The percentage of reads mapping to rRNA are shown.
- (B) Principal component analysis of gene-level counts from the indicated RNA-Seq datasets, arrows were added to visualize the time course.
- (C) Heatmap of Z-scores of log2-transformed gene-level counts from (left) upregulated and (right) downregulated genes (rows) and individual replicates of the indicated conditions (columns), clustered by hierarchical clustering with $k = 4$. The total number of genes that were differentially regulated at least once compared to the 0h IAA control is indicated.
- (D) Schematic overview of the ImpulseDE2 workflow. RNA-Seq counts of UPF1 depletion and recovery were combined to fit models (sigmoid or impulse model). For genes with significant time-dependent expression changes the best model was selected and the model parameters were extracted. The number of genes with significant gene expression changes per GENCODE gene biotype and according to their best model fit are shown on the bottom.
- (E) (Left) Fraction of genes per DGE cluster according to the best ImpulseDE2 fit. (Right) Boxplot of statistical significance of the best fit of DGE cluster genes, outliers are not shown.
- (F) Boxplot of statistical significance of (left) DGE cluster genes and (right) NMD relevance bins from published NMD meta analyses on the indicated knockdown (KD)/Rescue conditions. Outliers are not shown.
- (G) (Left) Schematic overview about the analysis of published phospho-UPF1 (p-UPF1) RIP-Seq data. Boxplot of log2FC of (Middle) DGE cluster or (Right) NMD relevance bin genes of IgG (control) or p-UPF1 RIP-Seq counts versus input. Outliers are not shown.

Figure S3

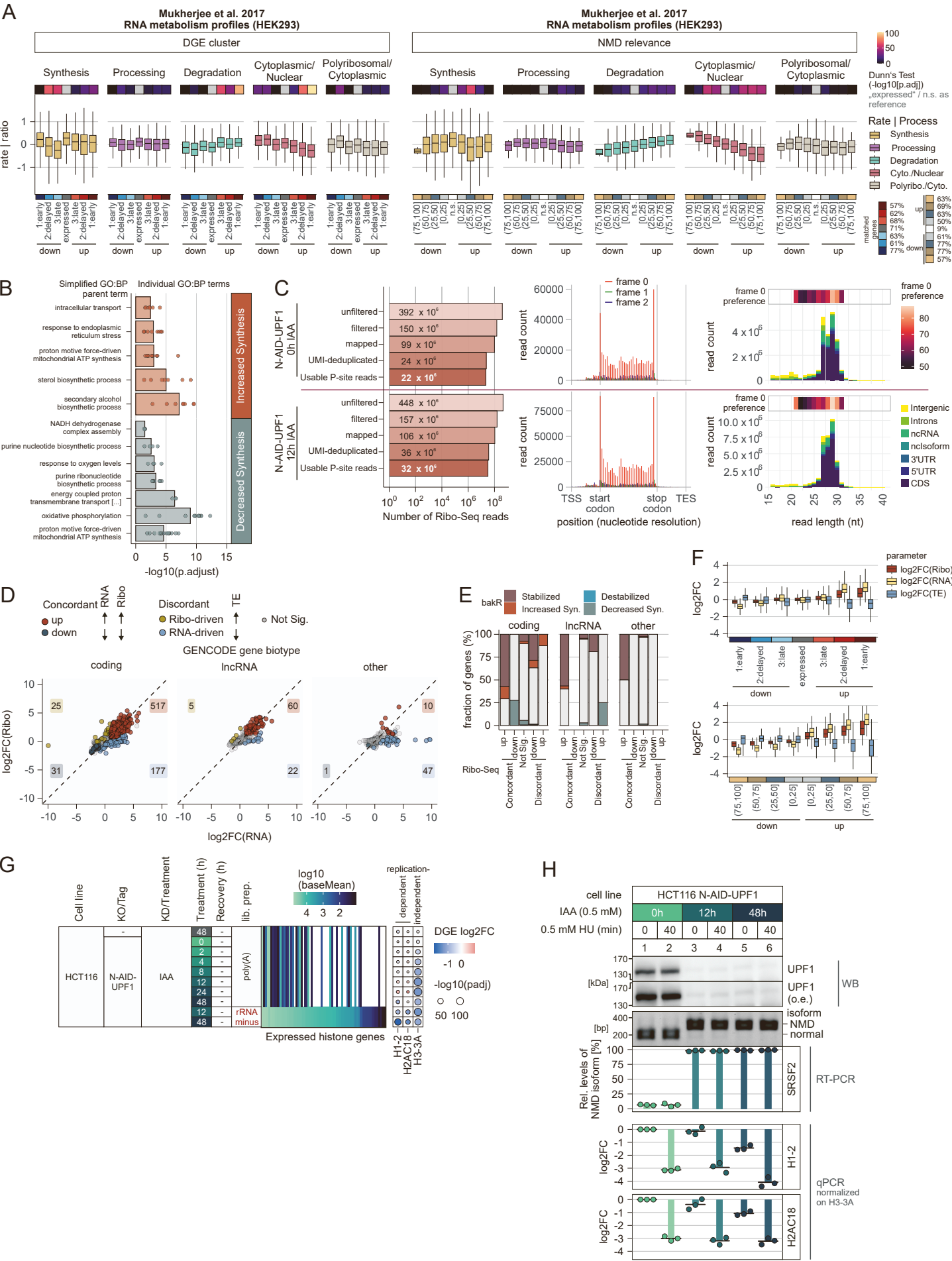


Figure S3. Exploring kinetic and translation characteristics of UPF1-regulated genes, related to Figure 3.

- (A) Boxplot of published RNA metabolism profiles plotting centered and scaled values per (Left) DGE cluster genes and (Right) NMD relevance bins. Statistical significance was determined using two-sided Dunn's test of multiple comparisons with the expressed genes or not significant (n.s.) bin as reference.
- (B) Genes exhibiting significantly increased or decreased synthesis (Syn.) in the differential kinetic analysis of 12h IAA N-AID-UPF1 compared to 0h IAA control were ordered by statistical significance of the synthesis rate ($p.adjust(ksyn)$) and subjected to functional enrichment analysis via g:profiler using the gene ontology biological process (GO:BP). The background gene set were all expressed genes. Significant terms were simplified using rrvgo and parent term mean adjusted p-value are plotted as bars. The individual enriched GO:BP terms are shown with their gSCS-corrected p-value as points.
- (C) (Left) Number of aggregated Ribo-Seq reads from three replicates for each condition at each analysis step. (Middle) Metagene plot of P-site read coverage aggregating signal over all covered transcripts. (Right) Read count distribution per read length in nucleotides and biotype/region. Frame 0 preference per read length is depicted.
- (D) Combined analysis of RNA expression changes ($\log_2FC(RNA)$) and ribosome occupancy changes ($\log_2FC(Ribo)$) of 12h versus 0h IAA condition, stratified by GENCODE biotype. Details for classification of genes in concordant or discordant classes are defined in the Methods section. The numbers of genes for each class are indicated. TE= Translation efficiency.
- (E) Fraction of genes per GENCODE biotype, metabolic analysis (from bakR) and translation analysis.
- (F) Boxplot of \log_2FC of RNA expression, ribosome occupancy and translation efficiency per DGE clusters (top) or NMD relevance bins (bottom), stratified by DGE up-/downregulation. Outliers of boxplot are not displayed.
- (G) Expression levels of histone genes in the respective conditions. Differential expression of individual genes used for further qPCR analysis is shown.
- (H) (Top) Western blot showing levels of UPF1 protein from N-terminal AID-tagged UPF1 HCT116 cell lines, treated with 500 μM IAA and 500 μM hydroxyurea (HU) for the indicated time, o.e. = overexposure. (Middle) Detection of the NMD target SRSF2 by end-point RT-PCR. The relative mRNA levels of SRSF2 isoforms were quantified. (Bottom) Probe-based qPCR of replication-dependent histone mRNAs H1-2 and H2AC18, normalized to replication-independent histone mRNA H3-3A, shown as \log_2 fold change (\log_2FC). Individual data points and mean (bars) are shown (n=3).

A

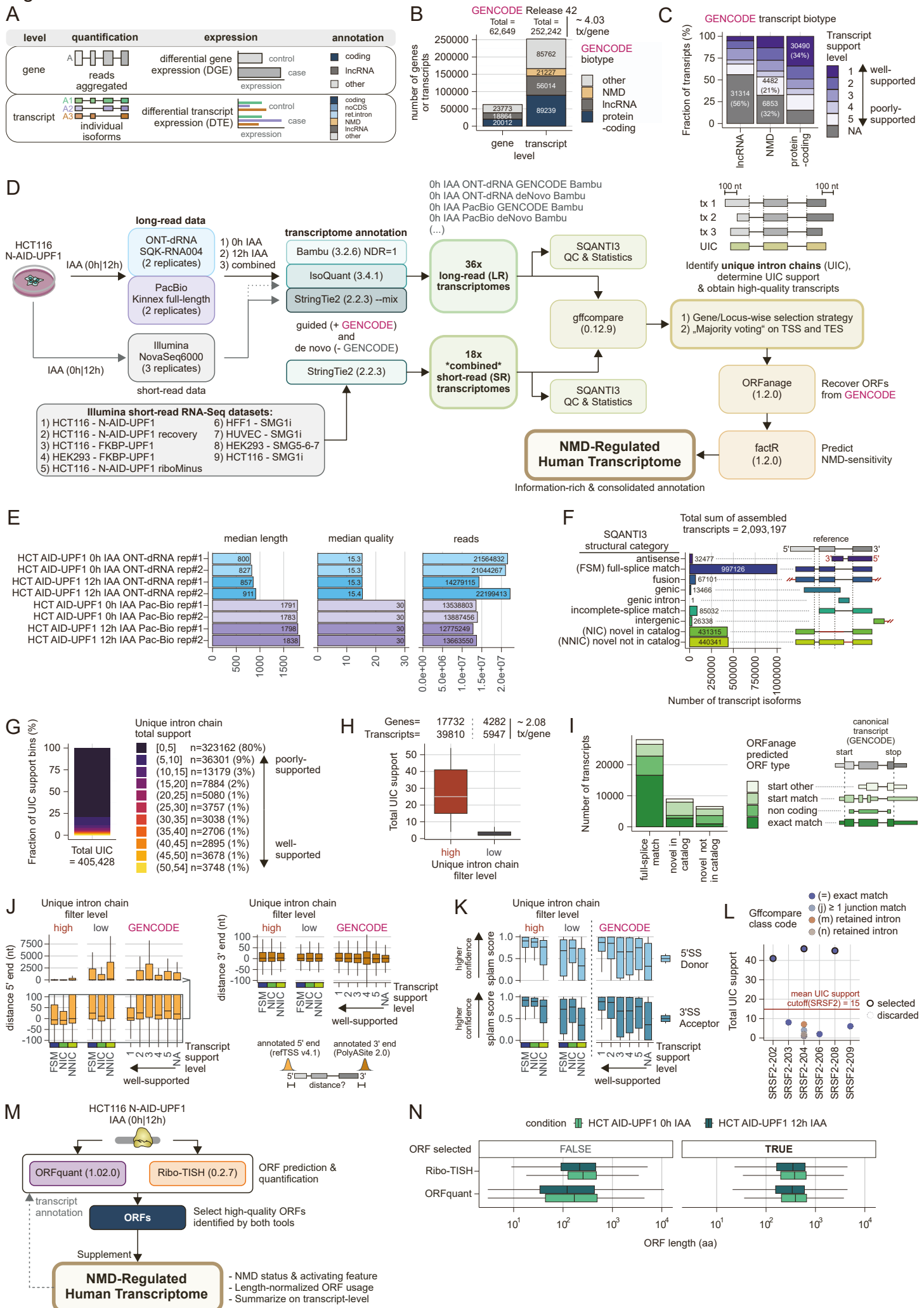


Figure S4. NMD-regulated human transcriptome characteristics, related to Figure 4.

- (A) Overview about gene- and transcript-level high-throughput analyses. For gene-level, RNA-Seq reads are typically summarized from all transcript isoforms expressed from the particular gene. Differential expression analysis is then performed between control and case conditions based on transcript- or gene-level counts. The GENCODE annotation offers different levels of granularity for transcripts or genes, as for example the NMD-annotation is only available on the transcript-level. GENCODE abbreviations: Coding = protein_coding, noCDS = protein_coding_CDS_not_defined, ret. intron = retained_intron, NMD = nonsense_mediated_decay.
- (B) GENCODE release 42 annotated number of genes and transcripts per biotype.
- (C) Fraction of GENCODE transcripts per transcript support level.
- (D) Overview of NMD-regulated human transcriptome (NMDRHT) definition. For details, see Methods or respective scripts.
- (E) Quality control and statistics of long-read RNA-Seq data.
- (F) Total assembled transcripts from 54 individual transcriptomes were analyzed regarding their structural category using SQANTI3 with GENCODE v42 as reference.
- (G) Fraction of unique intron chains (UIC) support, which is the number of individual transcriptomes (out of 54) in which the respective transcript with identical intron chain was detected.
- (H) Boxplot of total UIC support after two-phase filtering (high and low filter level). Outliers are not shown.
- (I) Number of transcripts with ORFanage-predicted ORFs stratified by ORF type with respect to start and stop codon positions in relation to the GENCODE canonical reference.
- (J) Boxplot of distance of (left) 5' end and (right) 3' end of transcripts to reference data (refTSS and PolyASite), per structural category or GENCODE transcript support level. Outliers are not shown.
- (K) Boxplot of aggregated transcript-wise splam score for splice donor or acceptor site confidence per structural category or GENCODE transcript support level. Outliers are not shown.
- (L) Example of UIC filtering process for SRSF2 transcripts. Low-support transcripts are filtered out.
- (M) Schematic overview about ORF prediction to supplement NMDRHT annotation.
- (N) Boxplot of predicted ORF length from ORFquant and Ribo-TISH, stratified by ORF selection (based on overlap). Outliers are not shown.

Figure S5

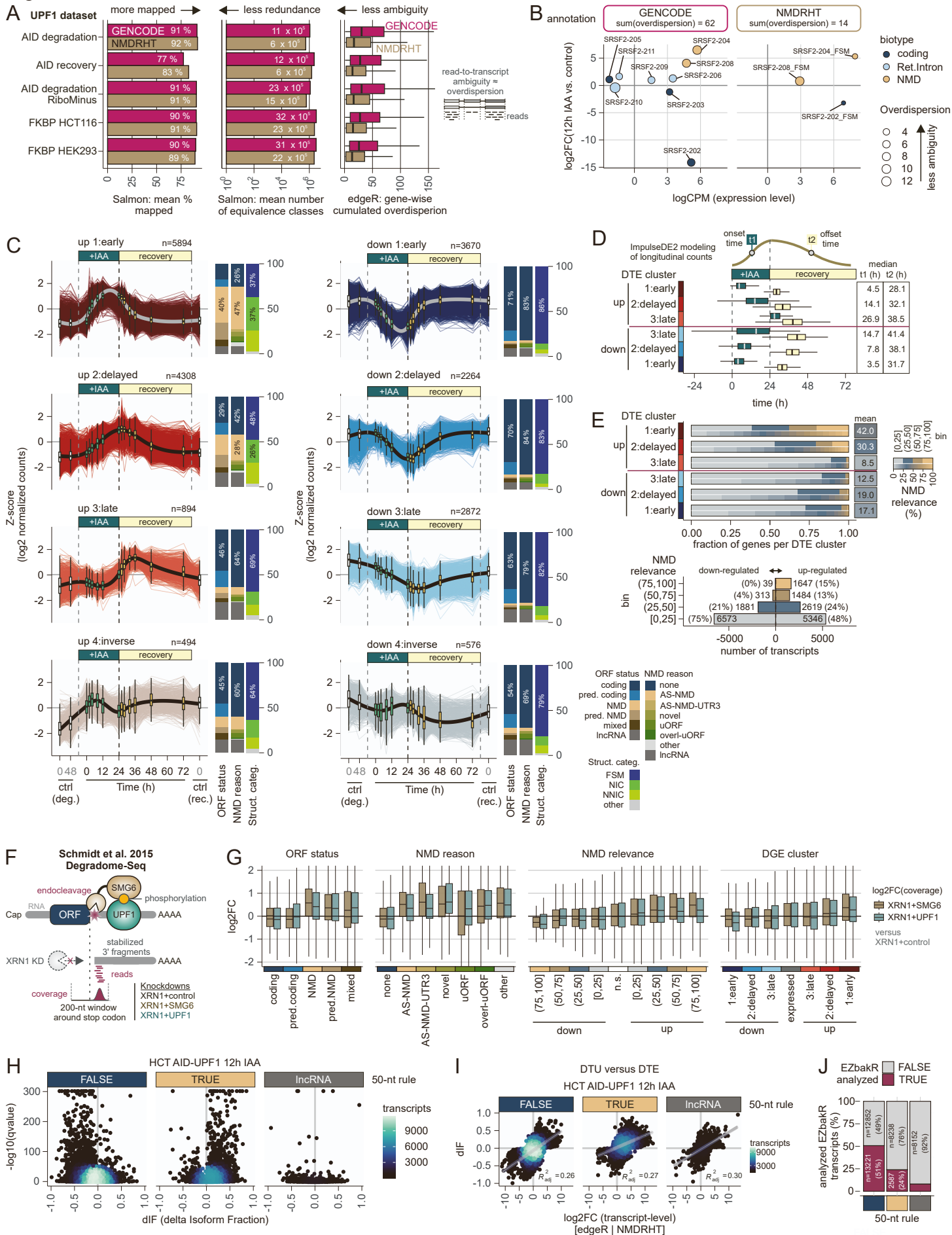


Figure S5. Quantification and characterization of NMDRHT-annotated transcripts, related to Figure 5.

- (A) Comparison of transcript quantification based on GENCODE or NMDRHT annotation for various UPF1-depletion datasets. (Left) Fraction of reads mapped by Salmon. (Middle) Total number of equivalence classes per condition. (Right) Cumulated overdispersion determined by edgeR.
- (B) SRSF2 transcript expression changes (12h IAA vs control) versus expression levels, quantified using GENCODE or NMDRHT annotation. The estimated overdispersion (quantification ambiguity) per transcript and cumulated for the gene is displayed. RI = retained intron.
- (C) Z-scaled log2-transformed transcript-level counts of combined degradation and recovery N-AID-UPF1 RNA-Seq data over IAA treatment/recovery time in hours. Control samples are shown at the terminal ends of the x-axis. Upregulated (left column) or downregulated (right column) clusters after hierarchical clustering ($k = 4$) are depicted. NMDRHT was used as annotation. The fraction of transcripts stratified by ORF status, NMD reason and structural category per differential transcript expression (DTE) cluster are plotted.
- (D) Boxplot of ImpulseDE2-derived modelled onset (t_1) and offset (t_2) time parameters for the indicated DTE cluster transcripts (outliers are not displayed).
- (E) (Top) NMD relevance as fraction of transcripts per DTE cluster, depicted as binned (upper bars) or absolute percentage (lower bars). (Bottom) absolute number of transcripts per binned NMD relevance, separated by up-/downregulation. Inverse clusters were excluded.
- (F) Scheme of the analysis of published Degradome-Seq data.
- (G) Boxplot of log2FC of Degradome-Seq data per DTE clusters (left) or NMD relevance bins (right), stratified by DTE up-/downregulation. Outliers of boxplot are not displayed.
- (H) Volcano plot of differential transcript usage (DTU) analysis.
- (I) Correlation between differential transcript usage (DTU, IsoformSwitchAnalyzeR; NMDRHT-based) and differential transcript expression (DTE, edgeR; NMDRHT-based) in 12h UPF1-depleted cells, stratified by the 50-nucleotide (50-nt) rule. The adjusted coefficient of determination from the linear model fit is indicated.
- (J) Fraction of NMDRHT-annotated transcripts per 50-nucleotide rule that were detected and analyzed by EZbakR on the transcript-level.

Figure S6

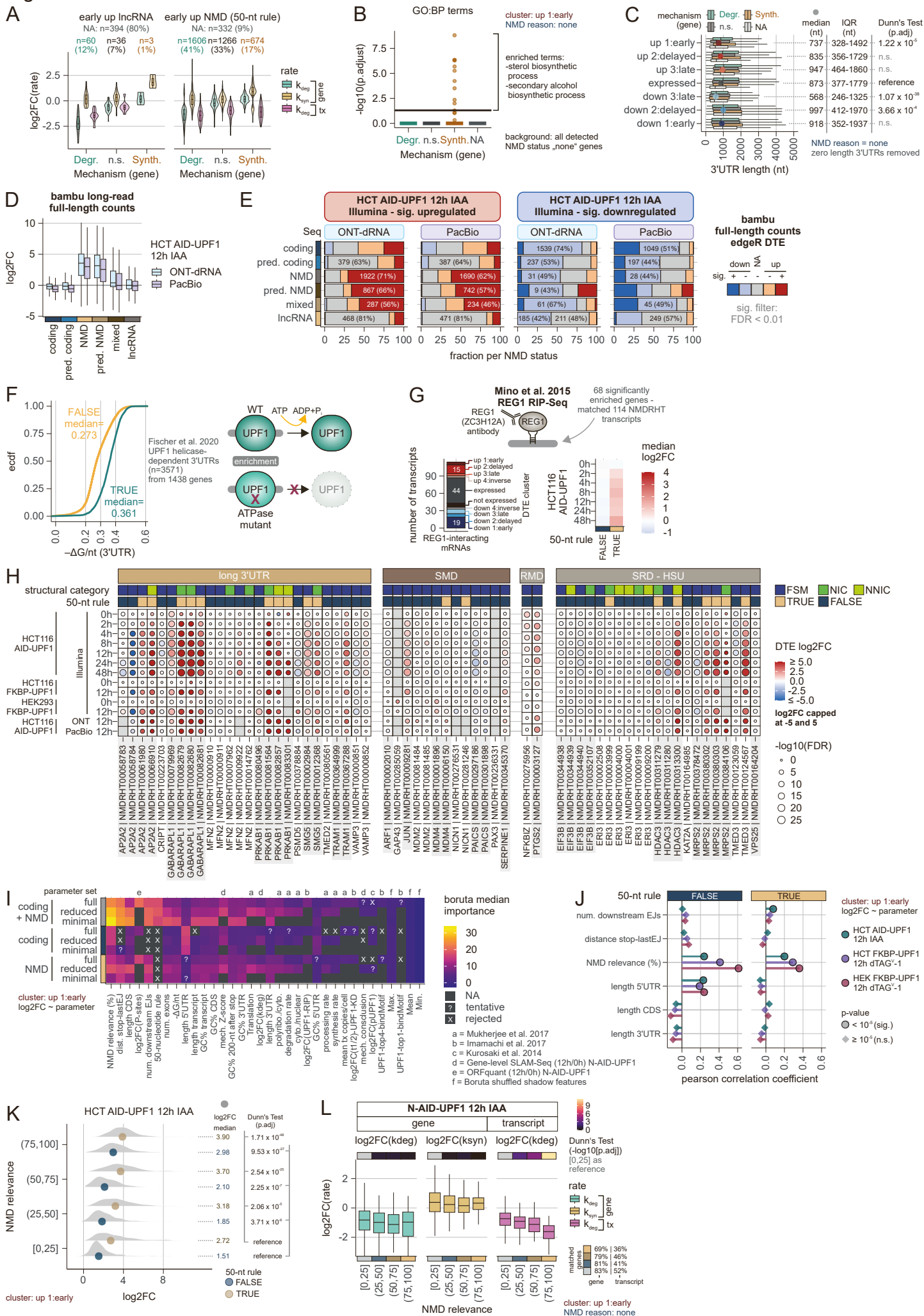


Figure S6. Characteristics of non-canonical UPF1-regulated transcripts, related to Figure 6.

- (A) Violin-Boxplot of log2FC of differential degradation (kdeg; gene- or transcript-level) or synthesis (ksyn; gene) rate constants of early upregulated cluster transcripts (left: lncRNAs, right: NMD-annotated) per gene-level bakR mechanistic conclusion.
- (B) Functional enrichment analysis of non-NMD-annotated, upregulated cluster transcripts stratified by gene-level bakR mechanistic conclusion via g:profiler, focused on gene ontology biological process (GO:BP). The background gene set were all expressed genes. The individual enriched GO:BP terms are shown with their gSCS-corrected p-value as points. Significant terms were simplified using rrvgo and parent terms are indicated.
- (C) Boxplot of non-NMD-annotated transcript 3'UTR length per DTE cluster and gene-level bakR mechanistic conclusion. Zero length 3'UTRs were removed, outliers of boxplot are not shown. Statistical significance was determined using two-sided Dunn's test of multiple comparisons with the expressed transcripts as reference.
- (D) Boxplot of DTE log2FC determined from bambu-quantified full-length long-read counts, stratified by ORF status. Outliers of boxplot are not shown.
- (E) Fraction of significantly (left) upregulated and (right) downregulated NMDRHT-annotated transcripts after 12h of N-AID-UPF1 depletion in HCT116 cells, stratified by ORF status and whether the transcript was concordantly and/or significantly regulated in long-read full-length bambu-quantified conditions.
- (F) (Left) Empirical cumulative distribution function (ecdf) of length-normalized, minimum thermodynamic free energy ($-\Delta G/\text{nt}$) of the 3'UTR of previously identified UPF1 helicase-dependent mRNAs (TRUE) or all other transcripts (FALSE). (Right) Scheme of helicase-dependent target determination.
- (G) (Top) Scheme of REG1 RIP-Seq data analysis. (Left) Number of REG1-interacting NMDRHT-annotated transcripts. (Right) Heatmap of median differential transcript expression (DTE) of REG1-interacting NMDRHT-annotated transcripts, stratified by UPF1 depletion time and 50-nt rule.
- (H) Differential transcript expression (DTE) of individual NMDRHT-annotated transcripts from genes previously implicated in long 3'UTR NMD, SMD, RMD and highly-structured SRD in the indicated conditions. The NMDRHT-annotated structural category and 50-nucleotide rule information is depicted.
- (I) Boruta-derived median importance of selected transcript features for explaining differential transcript expression from early upregulated transcripts per indicated parameter set.
- (J) Pearson correlation coefficient between differential transcript expression and the indicated transcript features for different 12h-depleted UPF1 conditions and stratified by 50-nucleotide rule.
- (K) Distribution and median log2FC of early upregulated cluster transcripts, stratified by 50-nucleotide rule and NMD relevance bin for 12h N-AID-UPF1 depletion in HCT116 cells. Statistical significance was determined using two-sided Dunn's test of multiple comparisons with the lowest NMD relevance bin transcripts set as reference.
- (L) Boxplot of transcript- or gene-level log2FC of differential degradation (kdeg) or synthesis (ksyn) rate constants of early upregulated, non-NMD-annotated cluster transcripts per NMD relevance bins. Statistical significance was determined using two-sided Dunn's test of multiple comparisons with the lowest NMD relevance bin transcripts set as reference.

Figure S7

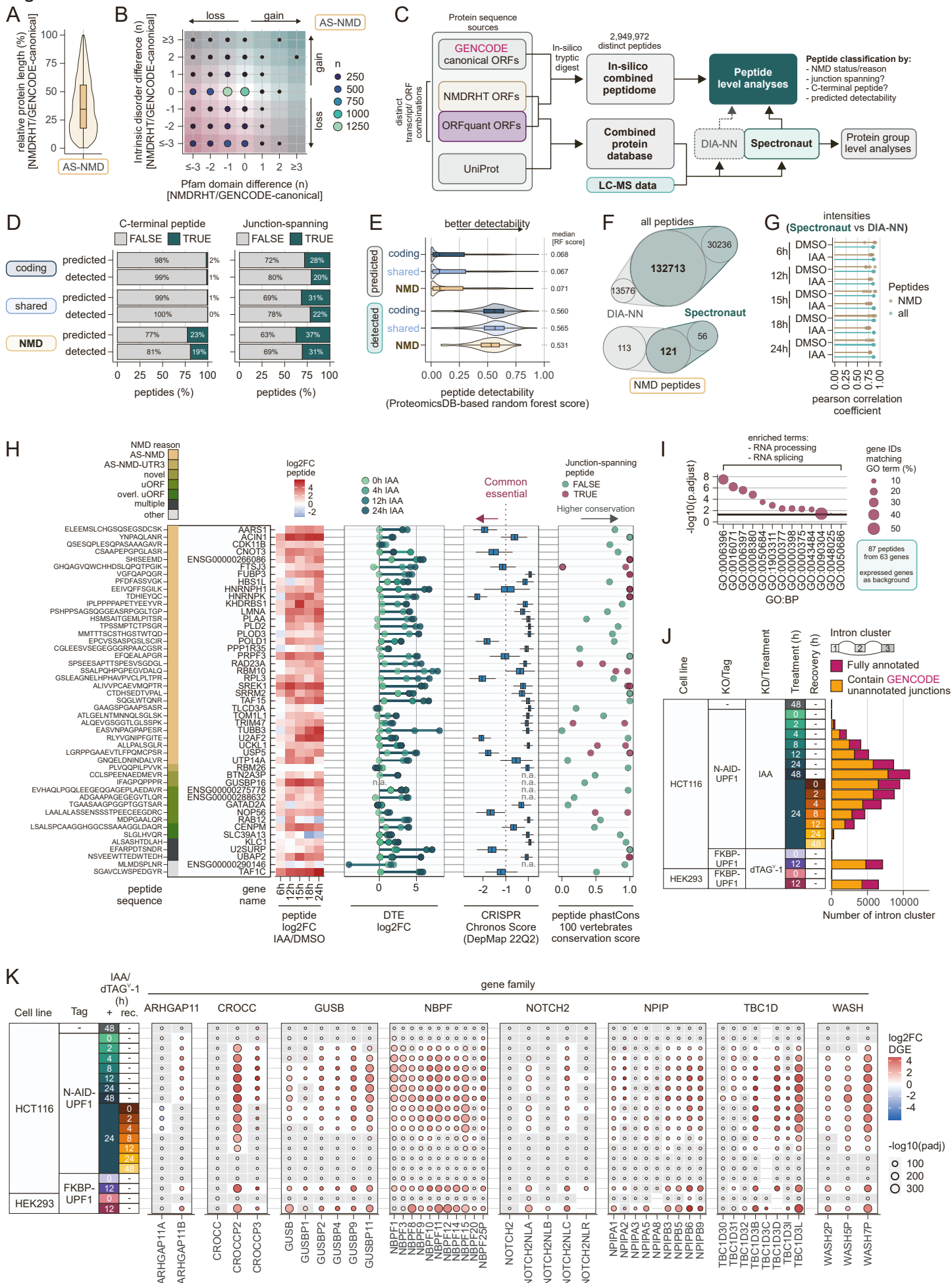


Figure S7. Proteome-level analyses of UPF1-depletion, related to Figure 7.

- (A) Relative encoded protein length of NMDRHT-annotated AS-NMD CDS in relation to their GENCODE-annotated canonical reference.
- (B) Difference in the number of Pfam protein domains and intrinsic disordered regions between AS-NMD-encoded proteins and their GENCODE canonical reference, as determined by InterProScan analysis.
- (C) Schematic overview of the proteomics analysis pipeline.
- (D) Fraction of (left) C-terminal peptides and (right) junction-spanning peptides stratified by peptide class and whether the peptide was predicted or detected.
- (E) Peptide detectability of predicted or detected peptides in different peptide classes was determined by PeptideRanger, based on ProteomicsDB-based random forest score.
- (F) Overlap of Spectronaut- and DIA-NN-detected peptides either for (top) all or (bottom) only NMD-informative peptides.
- (G) Pearson correlation coefficient between peptide intensities determined by Spectronaut and DIA-NN, for all or only NMD-informative peptides.
- (H) (Left) Differential expression of high-quality NMD-informative peptides (detected by both Spectronaut and DIA-NN, 1 peptide per gene). (Middle left) Mean differential transcript expression of corresponding transcripts encoding for the respective proteins/peptides. (Middle right) Essentiality of the gene based on DepMap 22Q2 Chronos Score. (Right) Aggregated phastCons 100 vertebrate conservation score of sequence encoding for the respective peptide, stratified by whether the peptide is splice junction-spanning.
- (I) Functional enrichment analysis of NMD-informative peptides with enough valid values via g:profiler, focused on gene ontology biological process (GO:BP). The background gene set were all expressed genes. The individual enriched GO:BP terms are shown with their gSCS-corrected p-value as points. Significant terms were simplified using rrvgo and enriched parent terms are indicated.
- (J) Analysis of alternative splicing on the intron cluster level, determined by LeafCutter.
- (K) Gene expression changes of gene family members implicated in human brain development.