This package comprises the predicted msdRs/msdPs, and the source data and scripts for msdR/msdP prediction (updated on September 12, 2024).

Supplementary Data:

File Name	Description
DataS1-unique_mirbase_and_mirgenedb_human_mirnas.7z	Unique unique miRBase human miRNAs (n=2,632) and unique mirGeneDB human mature miRNAs and corresponding miRNA passenger strands (n=503) used in this study.
DataS2-unique_msdRs_10-15nt_match.7z	Unique msdRs predicted using 10–15 nt matches between miRNA and template RNAs.
DataS3-merged_unique_msdPs_10-15nt_match.7z	Unique msdPs translated from msdRs.
DataS4-msRNA_artifacts_for_false_discovery_rate_estimation.7z	Simulated msRNA artifacts for estimating the false discovery rates for msdR prediction.
DataS5-number_of_predicted_msdRs.xlsx	Summary of the number of msdRs predicted using either database-documented miRNAs or control msRNA artifacts.
DataS6-46_msdPs_with_homologs_in_nr_15nt_match.xlsx	Summary of 46 msdPs with homologs in the NCBI nr protein database.
DataS7-six_msdPs_with_identical_hits_in_nr_15nt_match.xlsx	Summary of six msdPs with identical hit sequences in the NCBI nr protein database.
DataS8-1193_new_msdPs_15nt_match.xlsx	Summary of 1,193 new msdPs stated in the paper.
DataS9-941_new_msdPs_matching_with_internal_mass_spectra_data.xlsx	Summary of 941 msdPs matching with mass spectra from cancer cell lines.
DataS10-555_msdPs_identified_via_pepquery2_search.xlsx	Summary of 555 msdPs with confident peptide-spectrum matches via PepQuery2 search.
DataS11-DNA_sequencing_for_msdR0112_msdR0188.7z	Sanger sequencing source data for RT-PCR amplified RNA fragments of msdRs0122 and msdRs0188.
DataS12-Mass_spectrometry_report_for_msdP0188.pdf	Complete mass spectrometry report for msdP0188.
DataS13-human_proteins_with_DxDGD_and_over_500_AAs.xlsx	Summary of 358 DxDGD motif-containing human proteins with >500 amino acids.
DataS14-human_proteins_in_PDB_with_TM-score_greater_than_0.5.xlsx	Summary of 33 human proteins in the PDB with TM-score > 0.5 to the palm subdomain of three viral RdRPs.
DataS15-human_proteins_in_AFDB_with_TM-score_greater_than_0.5.xlsx	Summary of 84 human proteins in the AlphaFold Protein Structure Database with TM-score > 0.5 to the palm subdomain of three viral RdRPs.

Other Data:

File Name	Description
All_IRES.fa	Summary of all experimentally validated IRES sequences from the IRESbase database (n=1,328).
All_IRES_reformat.tsv	Reformatted IRES sequences ranked by sequence length from long to short (n=1,328).
four_longer_peptides_validated_by_MS.txt	Four longer consecutive peptides confirmed by mass spectrometry (MS) analysis (n=4).
list_hsa	List of all human miRNA identifiers (n=2,656).
list_hsa_nr_sort	List of nonredundant human miRNA identifiers (n=2,632) sorted by miRNA name in ASCII order.
list_ires	List of identifiers for all IRES sequences (n=1,328).
list_ires_hsa	List of identifiers of human IRES sequences (n=691).
list_mirgenedb_comm	List of common identifiers for mature and passenger miRNAs in the mirGeneDB database (n=503)
list_pepquery_database	List of names of databases indexed in PepQueryDB (n=48).
list_uniq_msdPs	List of names of unique msdPs (n=1,193).
mature.fa	Mature miRNA sequences from the miRBase database (n= 48,885).
mirbase_human_mirna_U2T.fa	Mature human miRNA sequences with U replaced by T (n=2,656).
scripts_for_msdRP_prediction_and_analysis.7z	Scripts used for msdR/miP prediction and analysis (n=21).
summary_of_pepquery2_confident_PSMs.tsv	Summary of confident peptide-spectrum matches by PepQuery2 search in the .tsv format.

Notes: The sizes of raw data (results) yielded from simulated msRNA artifacts and PepQuery2 search are extremely large and cannot be uploaded to Zenodo due to limited space (i.e., 50G). These data can be provided upon request to the corresponding author (Xiaoqiang Huang, xiaoqiah@umich.edu).