Supplementary figures

Figure S1. Overview of the mRNA-display process, Related to Figure 1. (A) The structure of the universal puromycin-containing linker oligo, which contains puromycin (“P” in purple circle), biotin (“B” in blue circle), and two inosine bases (“I” in pink). The underscored sequence in this linker can hybridize with a “linker hybridization sequence” at the 3’ end of a SMART-display generated mRNA that lacks a stop codon. (B) This hybridization facilitates the ligation of the 3’ end of the SMART-display generated mRNA with the 5’ end (5’Phos-) of this universal puromycin-containing linker sequence. (C) At the end of the translation process, puromycin enters the A-site of ribosome and forms covalent link with the translated peptide. (D) The fusion product is released from the ribosome.

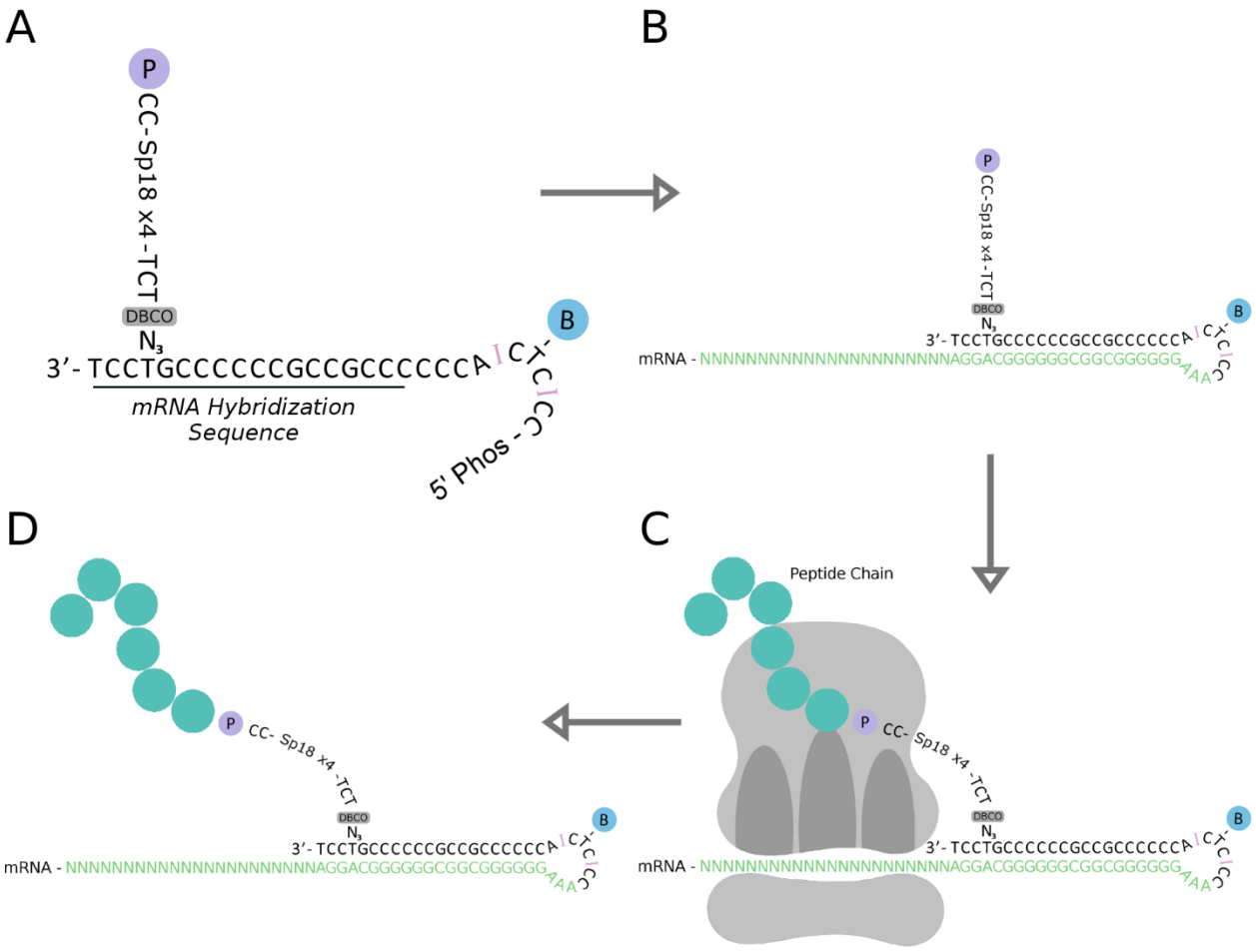


Figure S2. Test of association between two genes, Related to Figure 1 and Figure 4. (A) A continency table for the read pairs mapped to gene A (rows) and gene B (columns). Every mapped read pair is assigned to one and only one cell in this contingency table. The null hypothesis is that the mapping of a read pair to one gene (gene A) is independent to the mapping of this read pair to the other gene (gene B), where a read pair is considered mapped to a gene when either end of this read pair is mapped to that gene. (B) Flowchart of PROPERseqTools for processing PROPER-seq data. Linker sequence and adaptor sequences were trimmed (Adaptor trimming). Low quality reads and reads that were too short were removed (Quality filtering). The resulting read pairs were mapped to Refseq genes (Mapping), and those with the two ends mapped to two difference genes were obtained (Identification of chimeric read pairs). Non-redundant chimeric read pairs were used as the input for test of association (Statistical test).

A

Contingency table

		Mapped to gene B		
		Yes	No	
Mapped to gene A	Yes	x_{11}	x_{12}	
	No	x_{21}	x_{22}	
				Total number of uniquely mapped non-duplicate chimeric read pairs

B

PROPERseqTools flowchart

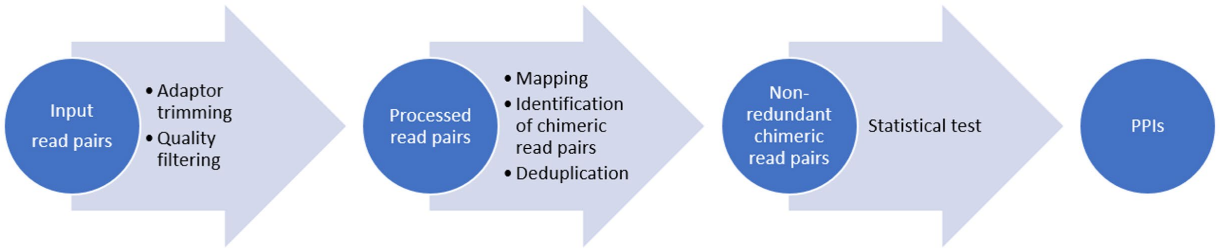


Figure S3. Testing antibody specificity to displayed fusion products, Related to Figure 2. (A) Size difference between unligated mRNA and puromycin-containing linker ligated mRNA. Bioanalyzer RNA Pico traces for the mRNA transcribed from a FLAG tag containing GFP gene before (grey) and after ligation to the puromycin linker sequence (blue). Migration time (x axis) reflects fragment size. The increase in fragment size between the unligated and the ligated sequences, based on the difference in migration time, is about 100 bases. (B) Western blot of the display products. The translation outputs of the puromycin-containing linker ligated mRNA were purified with either MyOne Streptavidin T1 beads (T1 column) or with NanoLink streptavidin beads (NL column) that reacted with the biotin on the puromycin-containing linker. The released materials from the beads were blotted with an anti-FLAG antibody (T1, NL columns). The supernatants of the bead selections were blotted as controls (T1 supernatant, NL supernatant). The Streptavidin T1 beads were used in the PROPER-seq protocol. The expected size of GFP protein with a FLAG tag is approximately 27 kDa. The expected size of the display complex (GFP protein, puromycin-containing linker, and mRNA) is approximately 350 kDa. (C) Specificity of antigen-antibody interaction. The selectivity of anti-GFP antibody was measured by the ratio of qPCR quantifications of each mRNA (column) in mixed bead purified mRNA-protein fusion products after vs. before pulling down with the anti-GFP antibody (y axis). The ratio for MAPKAPK2 was 0 because MAPKAPK2 was not detected post-selection. Error bar: standard error. (D) Venn diagram of the RNAs generated by the SMART-display process (Display 1) (Step G, Figure 2) and the original RNAs (Origin) (Step B, Figure 2). (E) Overlap of displayed genes between two repeated experiments (Display 1, Display 2). (F) Bioanalyzer traces of cDNA libraries generated from SMART-display generated fusion products (Intact protocol, green curve) and two control display libraries (blue and grey curves). One control library was generated by the same SMART-display process without ligating the puromycin-containing linker to the RNA (No-puromycin). The other control library was generated by digesting the SMART-display output library with proteinase K and removing all released contents (Protein digestion).

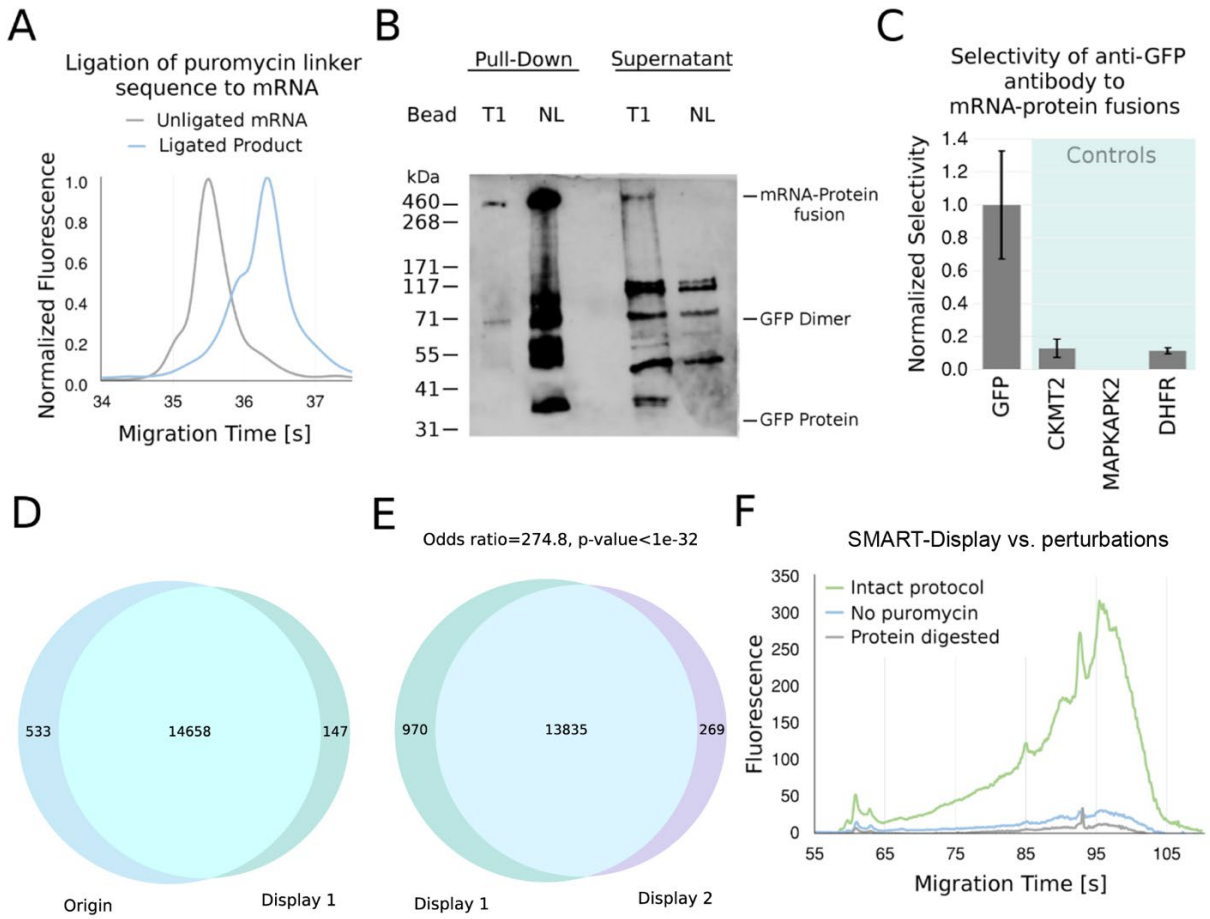


Figure S4. Comparison of the standard INLISE procedure with two variations, Related to Figure 3. (A) Flowchart of the standard protocol (PROPER-seq column) and the two variations (No-linker column and Proteinase column). (B) The ratio (y axis) of the quantities of DNA after vs. before the second last step (Streptavidin T1 selection) in the standard INLISE procedure (first column) as well as in the two variations (2nd and 3rd columns). All the ratios of a biological replicate (HEK1 or HEK2) were normalized to the ratio of the standard INLISE procedure of the same biological replicate. (C-D) Bioanalyzer traces of the sequencing library generated by the standard INLISE procedure (blue curve) and the two variations (green curves) in HEK1 (C) and HEK2 (D). The fluorescence signals are made comparable by normalizing to the concentration of the input sample (relative fluorescence, y axis).

A

	PROPER-seq All steps performed	**No-linker** No interaction linker ligated to prey library	**Proteinase** Bait sample digested with Proteinase K
SMART-Display	TSO reaction		
	PCR amplification		
	Transcription		
	Linker ligation		
	In vitro translation		
	Incubation in high salt condition		
INLISE	Streptavidin T1 pull-down		
	TSO based conversion to DNA		
	Restriction digestion		
	Ligation of interaction linker	Ligation reaction with no linker	Proteinase K digestion of bait library
	Interaction and crosslinking		
	Proximity ligation		
	Fragmentation and library preparation		
	Streptavidin T1 selection		
	Library amplification and sequencing		

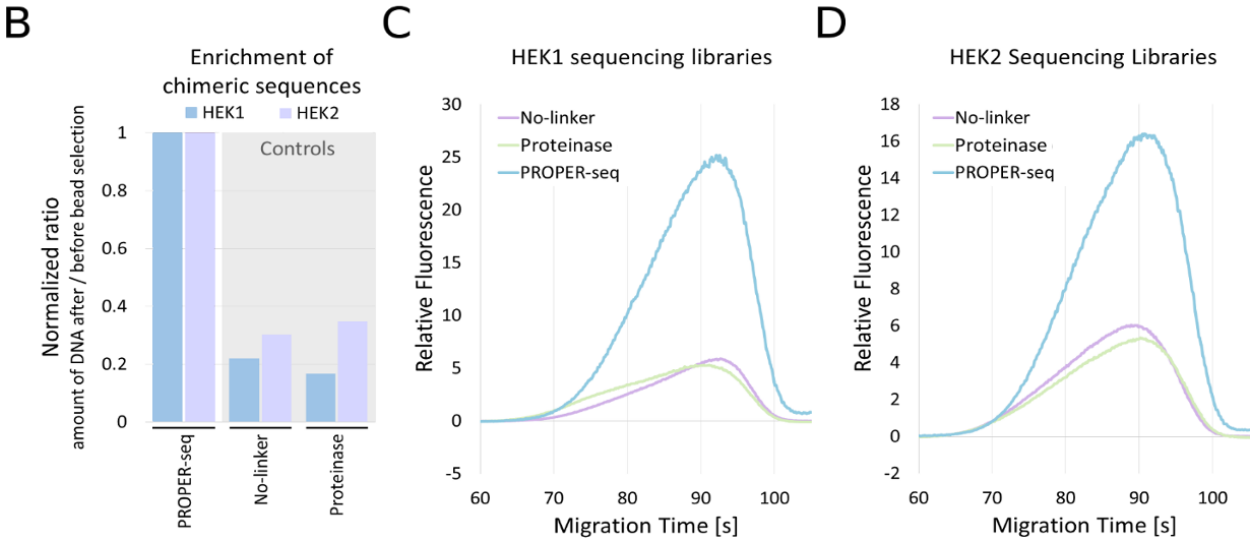


Figure S5. Reproducibility between biological replicates, Related to Figure 4. (A) A Venn diagram of the identified PPIs from each of the two HEK293T replicates (HEK1, HEK2). (B) The number of identified PPIs (y axis) from each biological replicate (HEK1, HEK2) with respect to the criteria of calling PPIs. The criteria were BH-corrected p-value < 0.05 and # read pairs > nX, where n was changed from 4 (default, dotted vertical line) to 40 (x axis). (C) The odds ratio of the two sets of PPIs identified from the two replicates (y axis) with respect to nX. For reference, the odds ratio between HuRI and HI-II-14 is marked as a horizontal line. (D-F) The same plots as (A-C) for the two Jurkat replicates (JKT1, JKT2). (G-I) The same plots as (A-C) for the two HUVEC replicates (HUVEC1, HUVEC2).

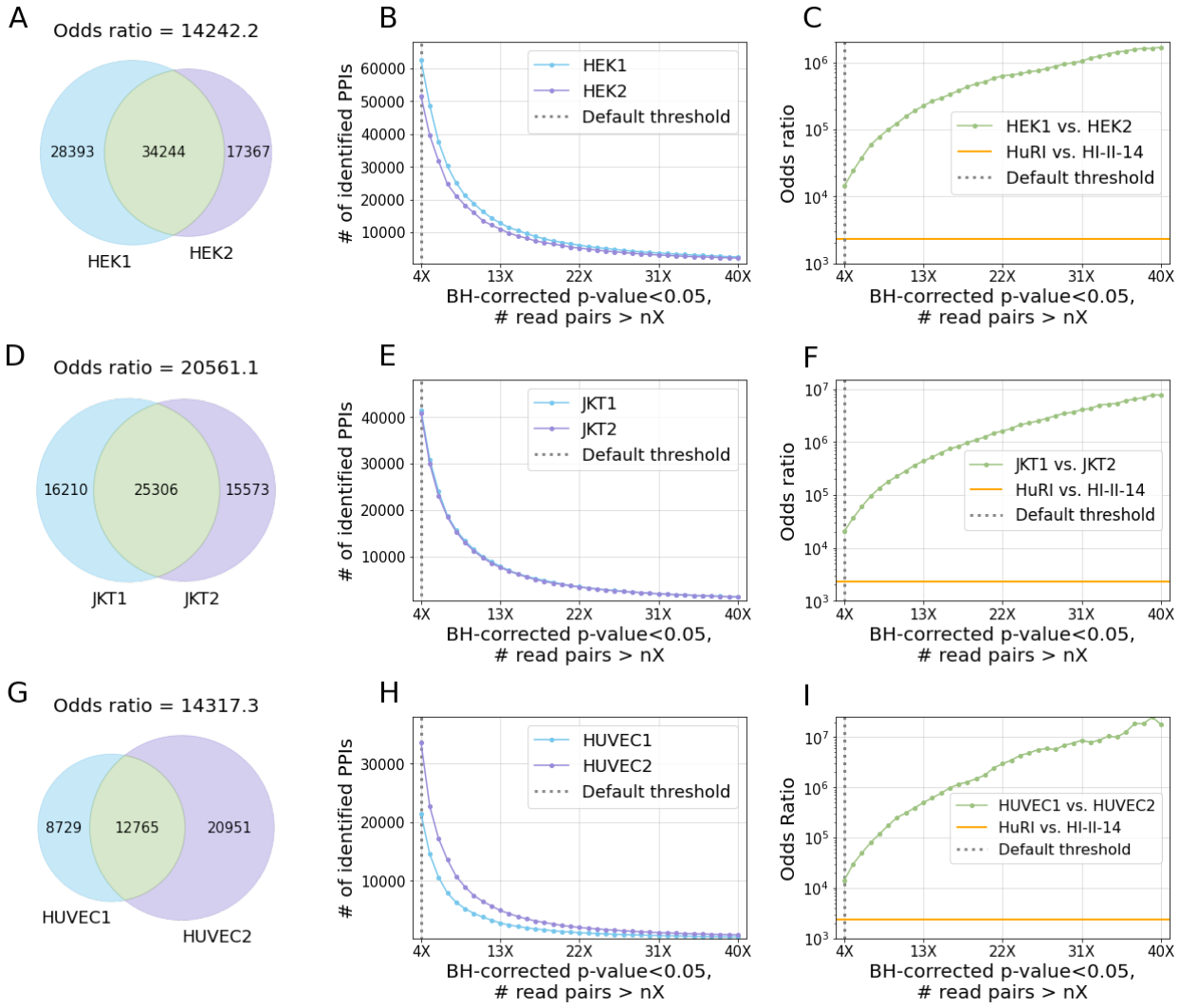


Figure S6. Precisions and recalls, Related to Figure 4 and Figure 6. PROPER-seq derived PPIs from HEK (A-C), Jurkat (D-F), and HUVEC (G-I) were compared to three types of known PPIs that were retrieved from APID, including all the PPIs that were identified by affinity purification-mass spectrometry (AP-MS), co-immunoprecipitation (co-IP), and liquid chromatography-mass spectrometry (LC-MS) derived PPIs (columns). The precisions of recalls of the PPIs identified from PROPER-seq's permutation dataset are marked in grey dots. The permutations were based on only the genes involved in PROPER-seq detected PPIs.

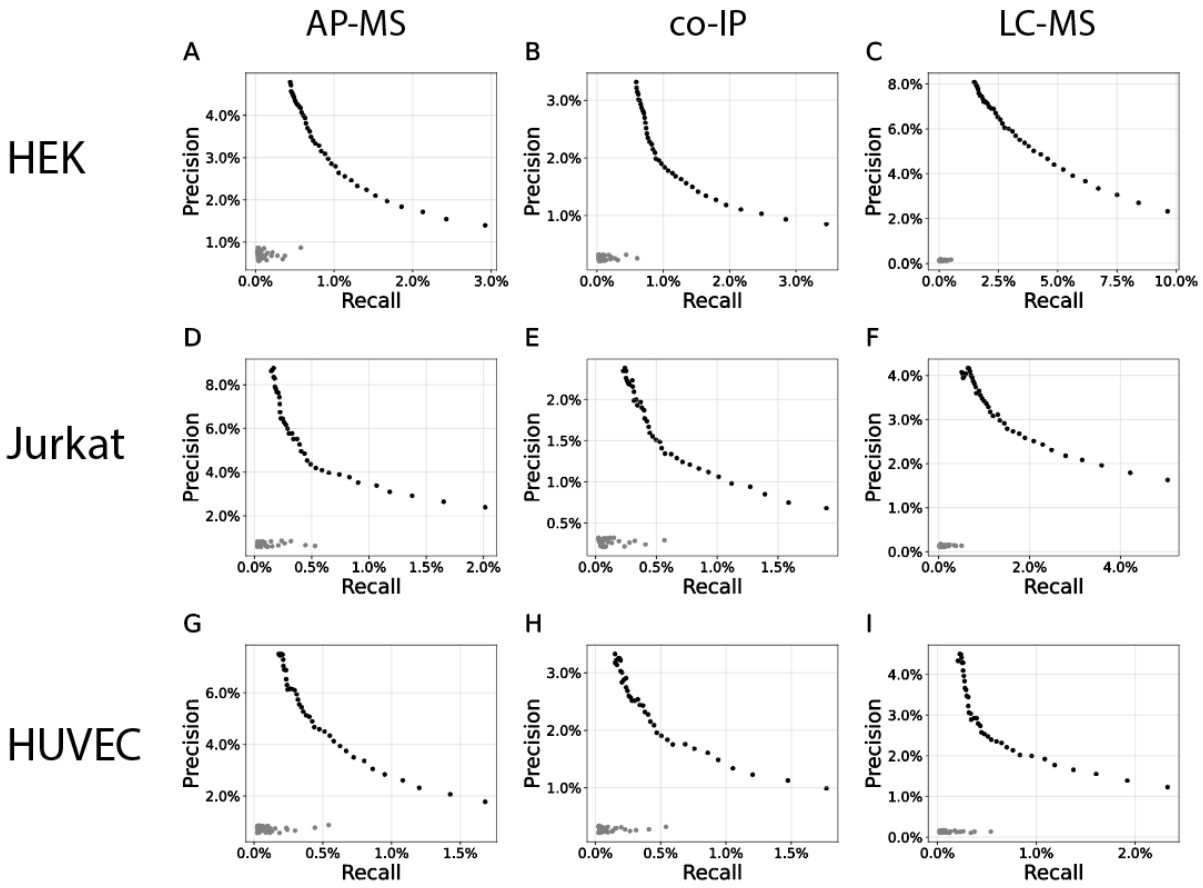


Figure S7. Precisions and recalls of each replicate, Related to Figure 4. Precision-recall curves of PROPER-seq derived PPIs from two biological replicates of HEK293T (blue and purple dots, A-C), Jurkat (blue and purple dots, D-F), and HUVEC (blue and purple dots, G-I) compared to three types of PPIs that are derived from other experimental methods, including all the APID PPIs that are detected by affinity purification-mass spectrometry (AP-MS), co-immunoprecipitation (co-IP), and liquid chromatography-mass spectrometry (LC-MS) derived PPIs (columns). The precisions and recalls calculated from permutation data (black dots) are included for reference. The permutations were based on only the genes involved in PROPER-seq detected PPIs.

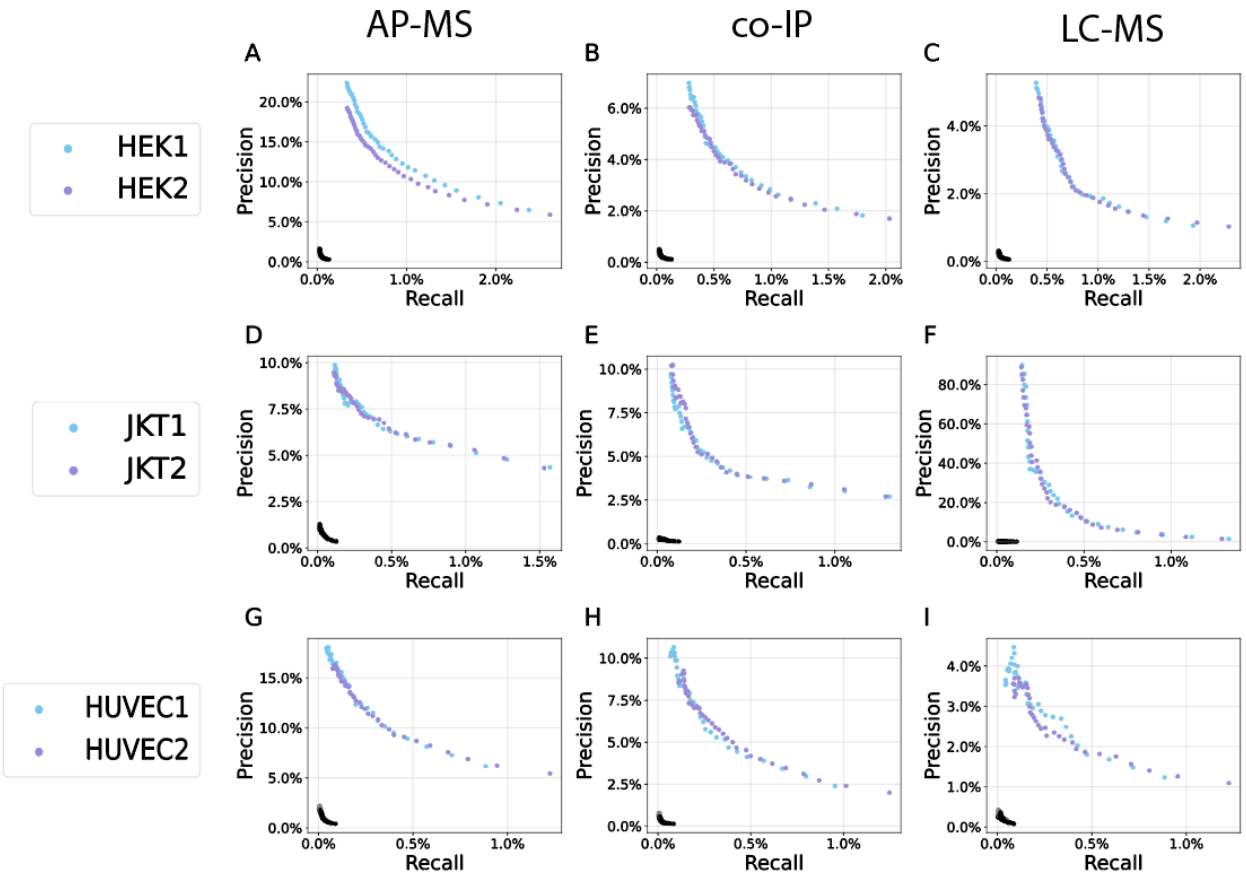


Figure S8. Log-log plots of clustering coefficient vs. degree, Related to Figure 4. Scatterplots (log-log plots) of clustering coefficient C(k) vs. degree (k), based on (A) Binary PPIs curated by Kovacs et al. (DOI: 10.1038/s41467-019-09177-y) (Lit-BM-13), (B) Non-binary PPIs curated by Kovacs et al. (Lit-NB-13), (C) The subset of predicted binary PPIs using Lit-BM-13 as the input data by the L3 algorithm with L3 scores > 50% quantile (L3-BM), (D) The subset of predicted non-binary PPIs using Lit-NB-13 as the input data by the L3 algorithm with L3 scores > 50% quantile (L3-NB), (E) The subset of PPIs predicted using HI-II-14 as the input data by the L3 algorithm with L3 scores > 25% quantile (L3-HI-II-14-lg), (F) The subset of PPIs predicted using HI-II-14 as the input data by the L3 algorithm with L3 scores > 75% quantile (L3-HI-II-14-sm), (G) the entire prePPI, (H) the subset of prePPI with structure score > 1, and (I) PROPER v1.0.

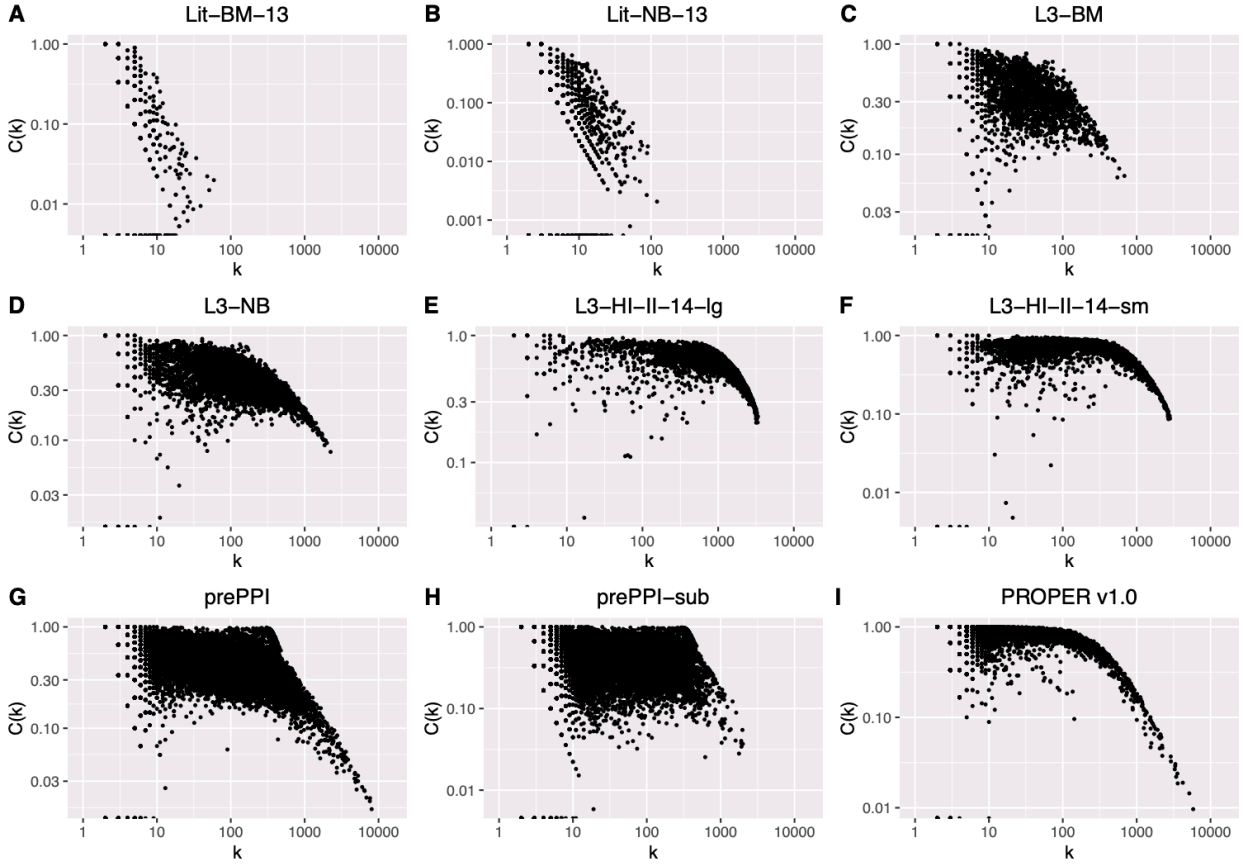


Figure S9. Q-Q plots, Related to Figure 4. Q-Q plots of AP-MS, Co-IP, LC-MS, PROPER v1.0 confirmed prePPIs (y axis) vs. the entire prePPI (x axis), based on structure score that reflects domain-domain interactions (A-D) and protein-peptide score that reflects domain-peptide interactions (E-H). See Table S1 for the descriptions for AP-MS, Co-IP, LC-MS datasets.

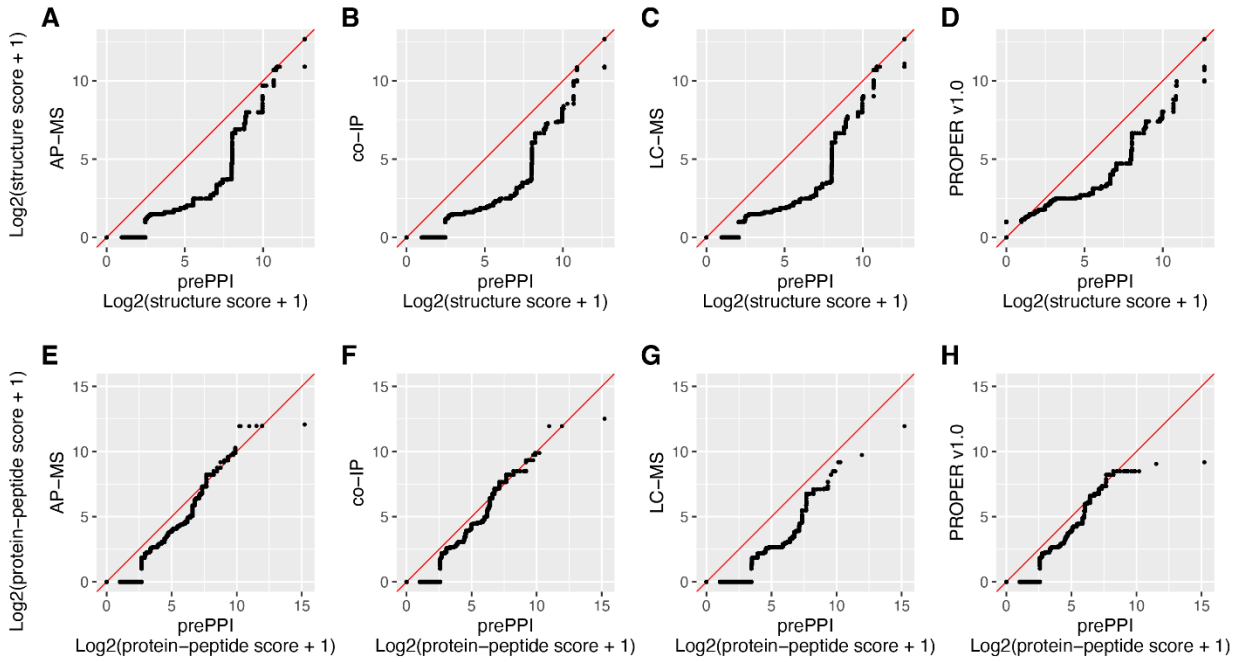


Figure S10. Immunoprecipitation of LEO1, Related to Figure 5. HEK293T lysates were immunoprecipitated with rabbit anti-human LEO1 antibody (anti-LEO1) or anti-rabbit IgG as an isotype control (anti-IgG). Both the precipitate and the supernatant were blotted with LEO1 antibody. Ladder: pre-stained protein ladder. Input: 5% of precleared cell lysates. The precipitates were used as input in PARP1 blots (Figure 5O).

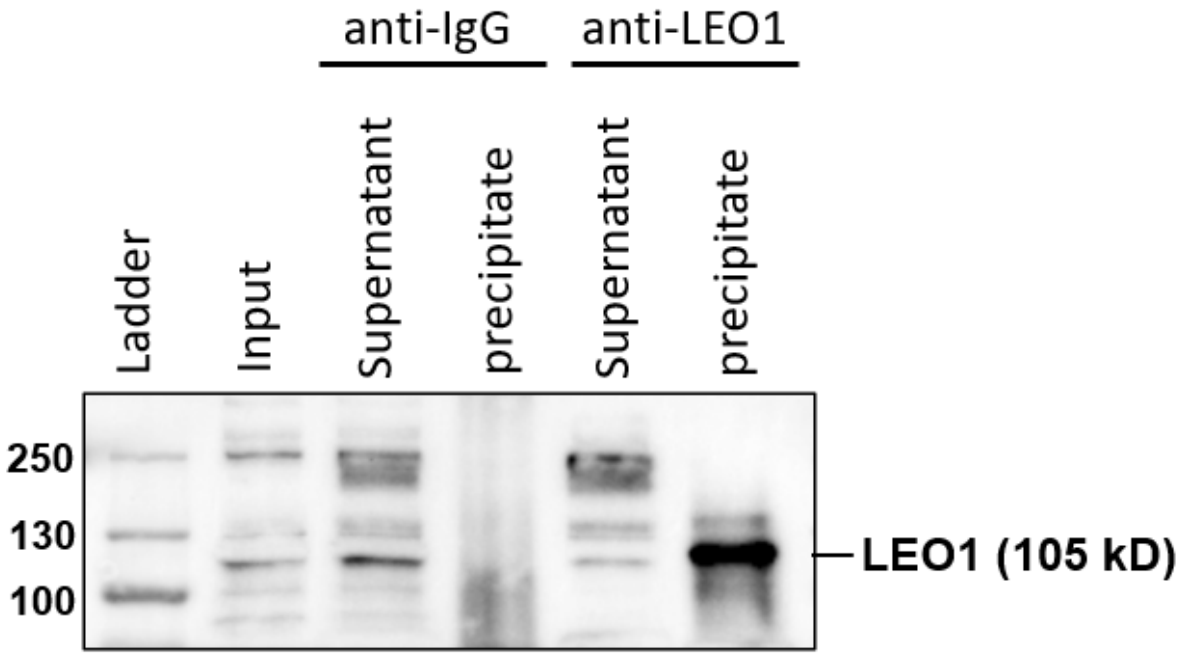
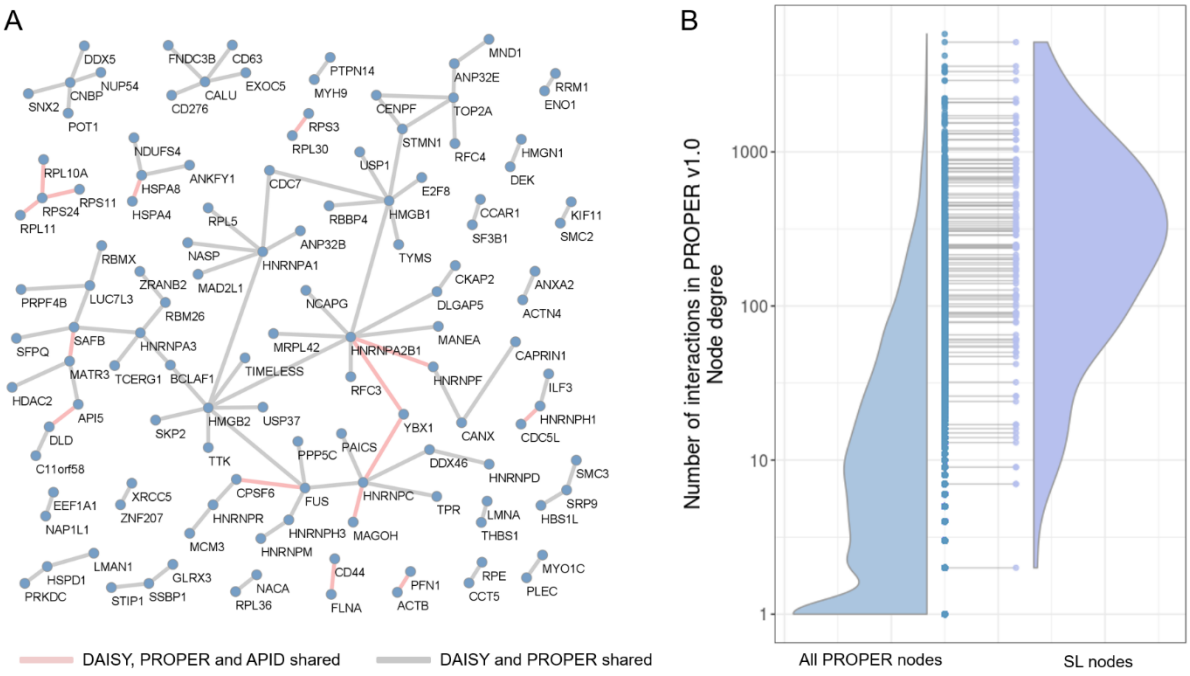


Figure S11. The overlap between DAISY SL gene pairs and PROPER v1.0, Related to Figure 4 and Figure 6. (A) The overlap between DAISY SL gene pairs and PROPER PPIs. Grey edge: DAISY SL gene pair that is also a PROPER PPI. Pink edge: DAISY SL gene pair that is a PROPER PPI and an APID documented PPI. (B) The degree distribution (half violin plot in blue) of all the PROPER v1.0 nodes (blue dots) vs. the degree distribution (half violin plot in purple) of all the SL nodes (purple dots). All degrees are based on the PROPER v1.0 network. The nodes with the same degrees are indicated by horizontal lines.



Supplementary tables

Table S1. Summary of PROPER-seq and perturbation libraries, Related to Figure 4. The libraries generated at the same time were given the same experiment ID (Exp ID). The total number of read pairs, the number of non-duplicate read pairs mapped to protein coding genes and the number of non-duplicate chimeric read pairs that were mapped to two different protein coding genes were listed in the last three columns.

Library ID	Expt ID	Cell line	Number of read pairs	# of non-duplicate read pairs mapped to coding genes	Non-duplicate uniquely mapped chimeric read pairs
HEK1	1	HEK293T	343,861,373	205,881,483	12,581,208
HEK2	2	HEK293T	248,657,713	173,300,648	7,747,982
JKT1	3	Jurkat	444,413,111	262,211,890	9,988,056
JKT2	4	Jurkat	390,643,931	236,283,970	9,385,745
HUVEC1	5	HUVEC	359,807,741	194,690,153	6,404,274
HUVEC2	6	HUVEC	483,597,124	283,434,465	9,705,398

Table S2. Datasets used in this work, Related to Figure 4 and Figure 6.

Name	Description	# PPIs	# proteins
APID	All the experimentally-derived human PPIs in APID, downloaded from http://cicblade.dep.usal.es:8080/APID/init.action	322,260	16,965
AP-MS	Affinity purification-mass spec detected PPIs that are included in APID	131,224	13,650
Co-IP	Co-IP detected PPIs that are included in APID	50,290	9,088
LC-MS	Liquid chromatography–mass spec detected PPIs that are included in APID	33,195	4,548
APID-binary	The experimentally derived binary PPIs curated into the APID database (level 2)	63,954	12,572
APID-non-binary	The PPIs derived from non-binary methods in the APID database	258,306	15,847
Lit-BM-13	Binary PPIs curated by Kovacs *et al. Nat Commun* 10, 1240 (2019).	4,386	3,249
Lit-NB-13	Non-binary PPIs curated by Kovacs *et al. Nat Commun* 10, 1240 (2019).	10,152	5,382
prePPI	Zhang, Q. C. *et al.* Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556-560, doi:10.1038/nature11503 (2012)	1,279,381	16,903
prePPI-sub	Subset of prePPI with structure scores > 10.	619,619	13,222
L3-BM	The subset of predicted binary PPIs using Lit-BM-13 as the input data by the L3 algorithm (Kovacs *et al. Nat Commun* 10, 1240) with L3 scores > 50% quantile.	56,890	2,726
L3-NB	The subset of predicted non-binary PPIs using Lit-NB-13 as the input data by the L3 algorithm with L3 scores > 50% quantile.	387,971	4,694
PROPER v1.0	The PPIs derived from the merged PROPER-seq libraries of HEK1, HEK2, JKT1, JKT2, HUVEC1 and HUVEC2	210,518	8,635
HEK	The PPIs derived from merged PROPER-seq libraries of HEK1 and HEK2	109,539	7,292
Jurkat	The PPIs derived from merged PROPER-seq libraries of JKT1 and JKT2	72,409	5,136
HUVEC	The PPIs derived from merged PROPER-seq libraries of HUVEC1 and HUVEC2	51,125	4,266
		# gene pairs	**# genes**
DAISY	Jerby-Arnon, L. *et al.* Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell* 158, 1199-1209 (2014).	2,802	2,077

Table S3. The estimated screening completeness, sampling sensitivity, assay sensitivity, overall sensitivity, precision for PROPER v1.0, and the estimated protein interactome size, Related to Figure 4.

	PROPER v1.0
Screening completeness	64.7%
Sampling sensitivity	35.4%
Assay sensitivity	43.38%
Overall sensitivity	15.36%
Precision	5.77%
Human protein interactome size	8.5×10^5