Appendix

Proteome-scale mapping of binding sites in the unstructured regions of the human proteome

Caroline Benz[1,*], Muhammad Ali[1,*], Izabella Krystkowiak[2,*] , Leandro Simonetti[1], Ahmed Sayadi[1], Filip Mihalic[3], Johanna Kliche[1], Eva Andersson[3], Per Jemth[3], Norman E. Davey[2,#], Ylva Ivarsson[1, #]

Affiliations:

1. Department of Chemistry - BMC, Uppsala University, Box 576, Husargatan 3, 751 23 Uppsala, Sweden
2. Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.
3. Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, Husargatan 3, 751 23 Uppsala Sweden

Table of content

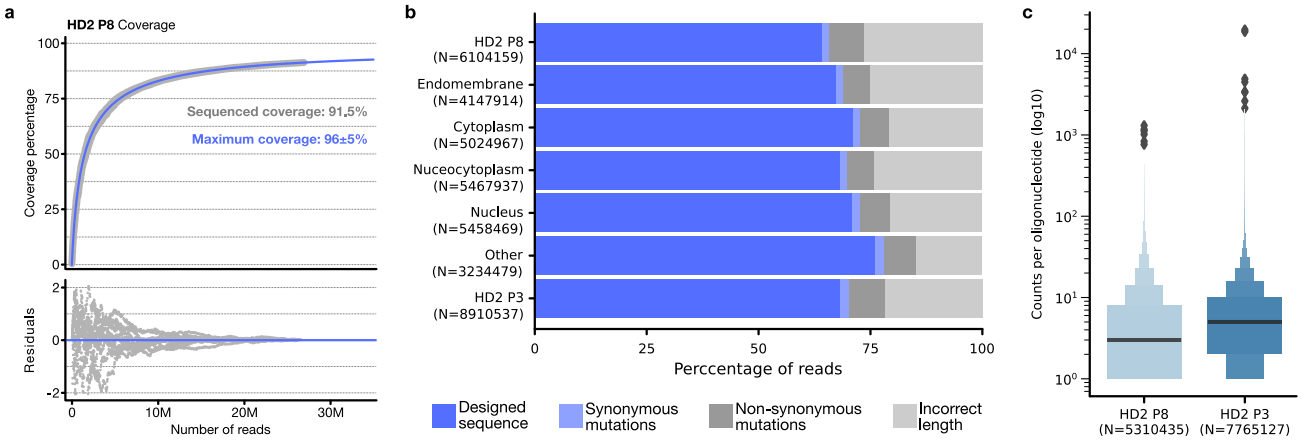


Figure S1. Non-design components and count distribution of the libraries. (a) Coverage of the HD2 P8 library as a function of sequencing depth, fitting of data to a double hyperbolic model (in blue) and residuals are shown. The contribution of reads of all sub-libraries were considered for calculating the coverage of the HD2 P8 library, as it is the result of the balanced combination of all compartment-specific sub-libraries. (b) Percentages of reads obtained by NGS of the libraries generated in this work associated with sequences that matched the designed sequences, those that presented mutations either silent or non-silent, and those that due to frame-shifting, insertion or deletion mutations did not match the expected oligonucleotide length. (c) Distribution of NGS counts for the designed oligonucleotides in the HD2 P8 and HD2 P3 libraries.

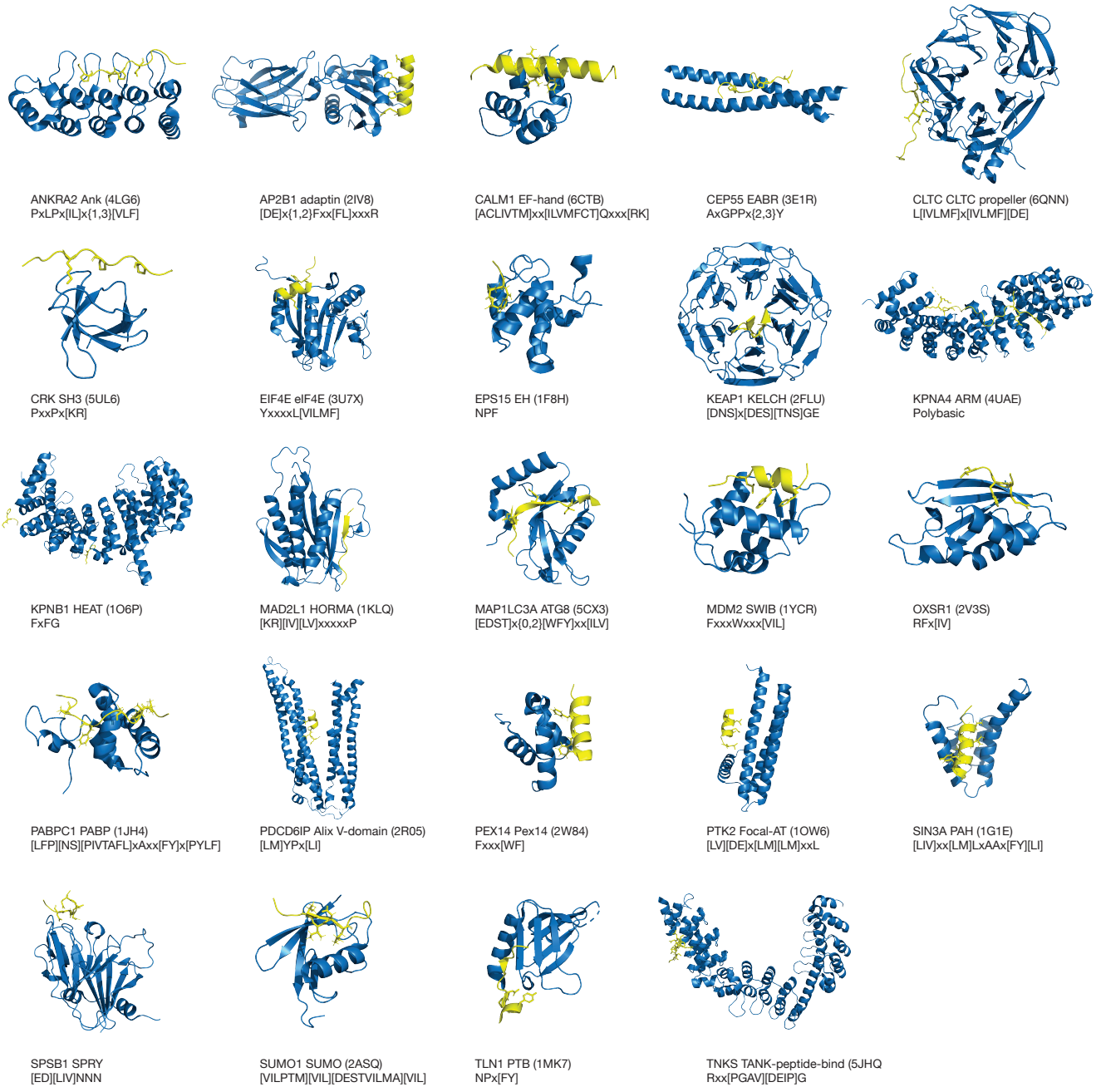


Figure S2. Representative structures of the bait proteins domains (blue) with bound ligands (yellow) and indicated binding motifs. PDB codes are indicated in the figure.

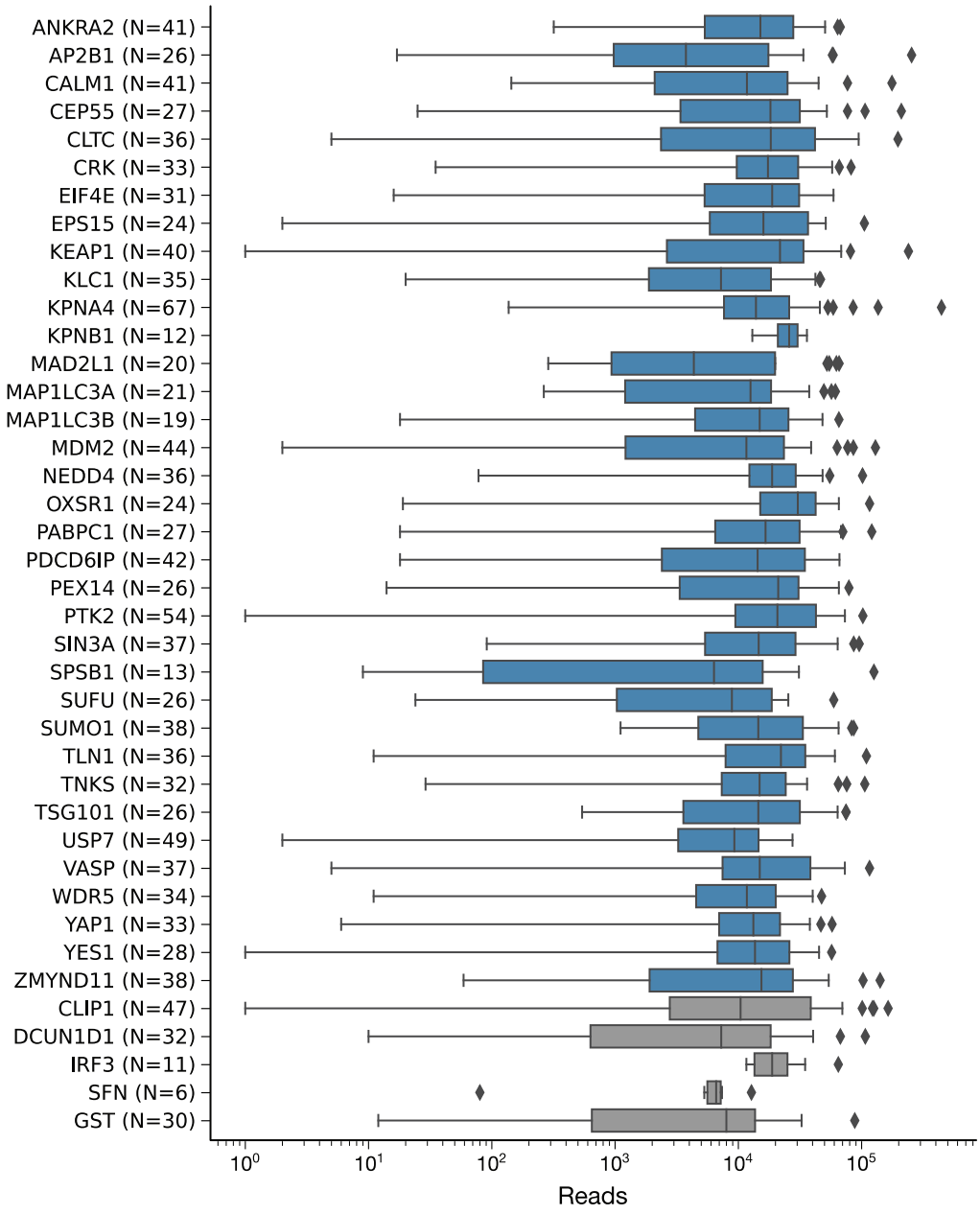


Figure S3. Distribution of NGS reads for binding enriched phage pools. Distribution of the NGS reads that matched the peptides encoded by the library are shown grouped by bait. Control baits are shown in gray. N = number of barcoded phage pools included in the analysis per bait.

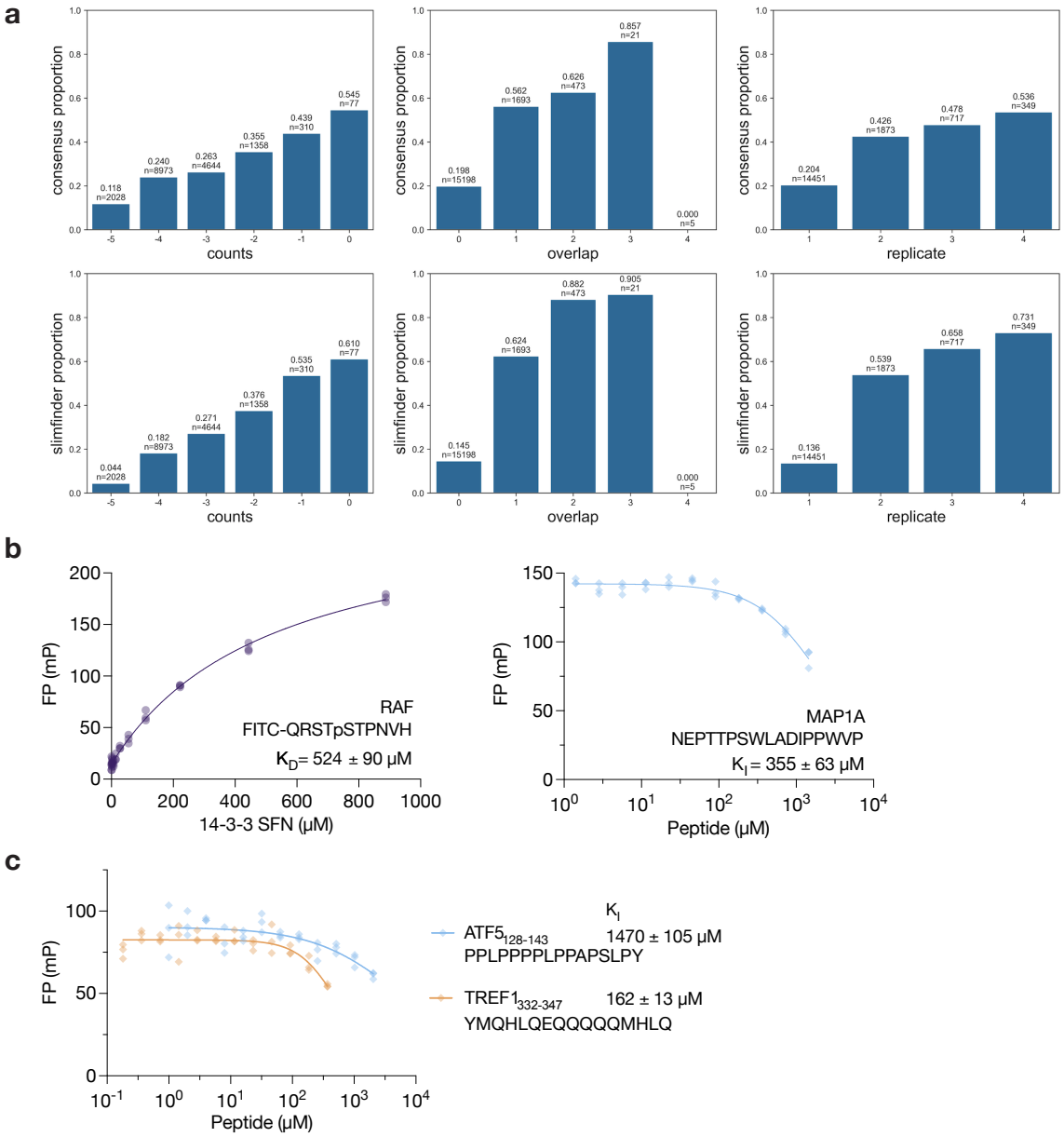


Figure S4. Analysis of the proportion of motif-containing peptides (a) and validation of low affinity interactions of peptides lacking consensus motifs (b-c).
(a) Proportion of peptides returned for a bait containing the expected ELM consensus for the bait (consensus proportion) or the SLiMFinder consensus enriched in the peptides in the screen of that bait (slimfinder proportion). The plots show the data grouped by three peptide metric statistics: count - $\log^{10}$ of the normalized counts, overlap - the number of overlapping peptides returned in the screen for that bait, replicates - the number of replicates that the peptide was returned. (b) FP affinity measurements of the interactions between 14-3-3 SFN and the probe peptide phospho-RAF1255-264 (FITC-QRSTpSTPNVH; left) and the competition with unlabeled $MAP1A_{1836-1851}$ (right). (c) FP competition experiment for the Keap1 KELCH domain between the FITC-labeled probe peptide $NFE1L1_{226-243}$ and the unlabeled peptides $ATF_{128-143}$ and $TREF_{223-347}$ that lack Keap1 KELCH binding motif.

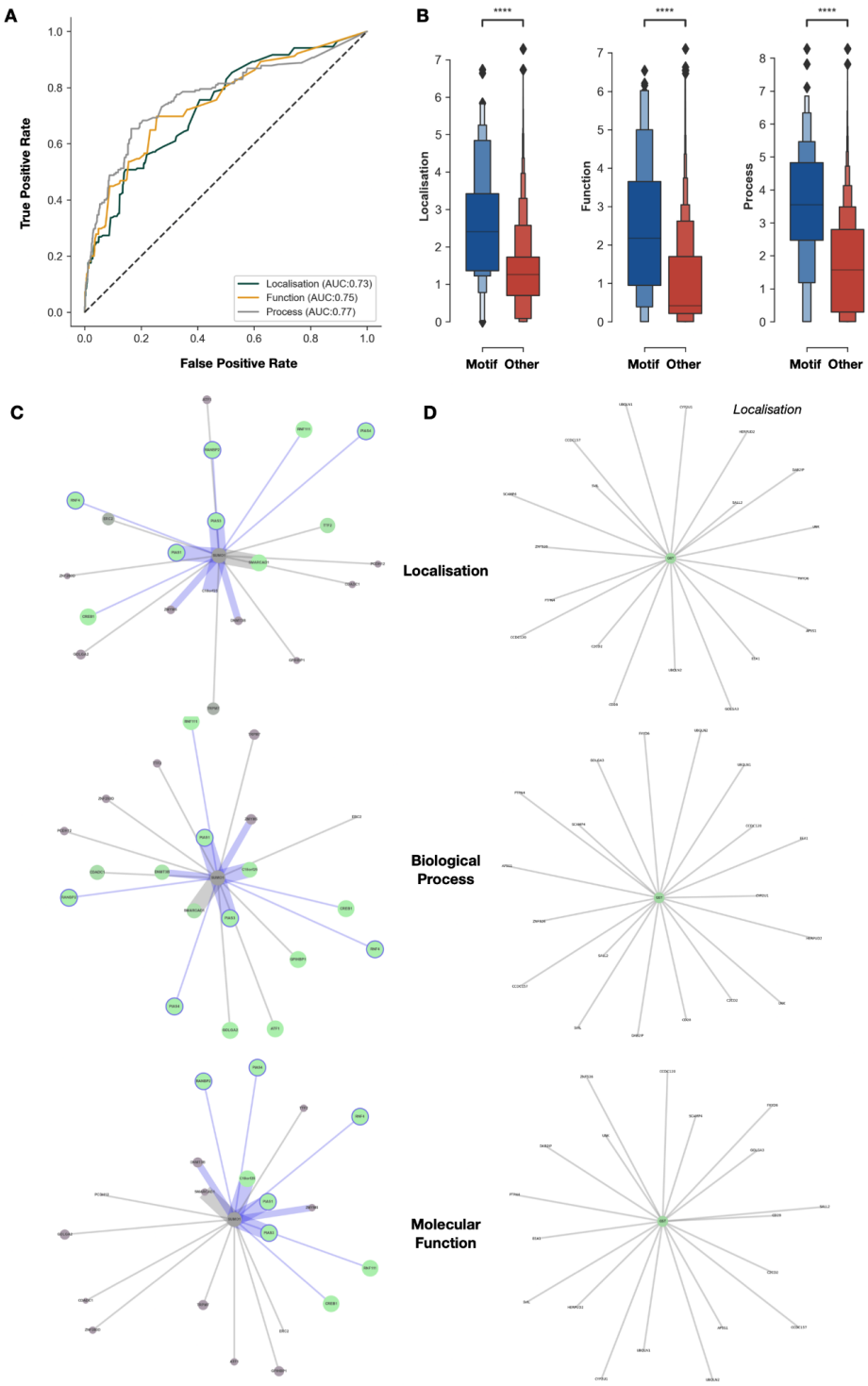


Figure S5. Overview of GO term enrichment analysis. (A) ROC curve showing the discriminator power of the shared GO terms analysis metrics for validated motifs (B) Box plots of Localisation, Biological Process and Molecular Function GO term enrichment in validated motif containing peptides and other returned peptides. (C-D) Network of representation of the SUMO (C) and GST (D, control) bait interactions with high/medium confidence peptide-containing proteins. Network is overlaid with data from a shared GO terms analysis. Three networks are displayed for each bait: Localisation, Biological Process and Molecular Function. Node colour and size is the p-value of the best shared ontology term (greener and larger nodes are more enriched). Edge weights and lengths represent the best peptide confidence for the protein (more confident peptides are closer to the bait with thicker edges). The edges of known interactors are shown in blue. Nodes representing validated motif-containing proteins are ringed in blue.

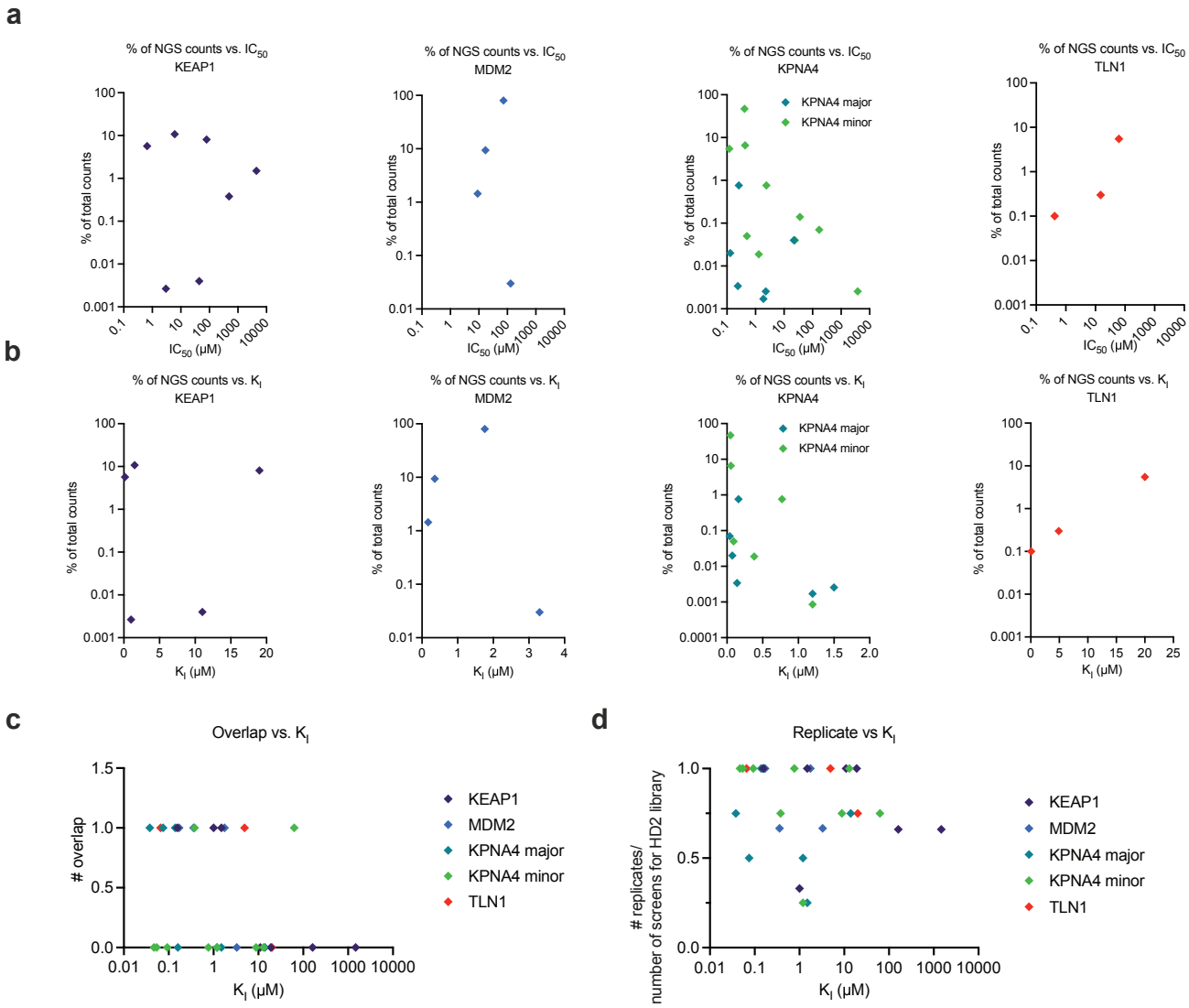


Fig S6. Analysis of correlation between affinity and ProP-PD results. The % NGS counts per were plotted against (a) IC_{50} values, and (b) K_I values. The affinities (K_I values) were further plotted against (c) the observed overlaps of the peptides in selections, and (d) the occurrence of peptides in replicate selections. A KPNA4 outlier was removed for better visualization.

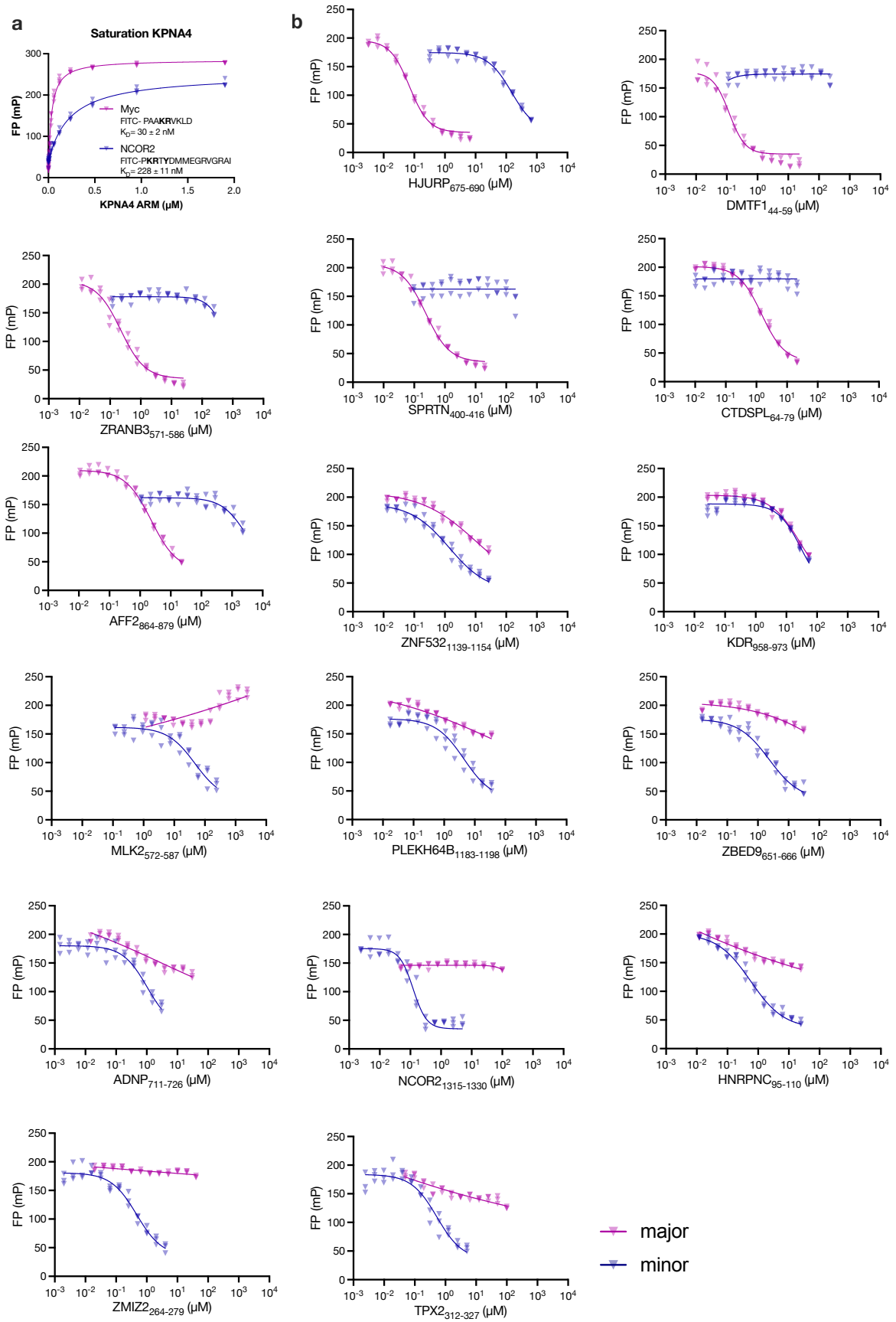


Fig S7. FP affinity measurements of KPNA4. (a) Saturation curves with probes for the for the major groove (Myc320-328 FITC-PAAKRVKLD; purple) and minor groove (NCOR21307-1322; FITC-PKRTYDMMEGRVGRAI (blue)). (b) FP competition binding experiment of 18 peptides against the major groove probe FITC-Myc320-328 (purple) and the minor groove FITC-NCOR21307-1322 (blue). Unlabeled peptides used for competition: HJURP675-690, DMTF144-59, ZRANB3571-586, SPRTN400-416, CTDSPL64-79, AFF2864-879, ZNF5321139-1154, KDR958-973, MLK2572-587, PLEKH64B1183-1198 ZBED9651-666, ADNP711-726, NCOR21307-1322, HNRPNC95-110, ZMIZ2264-279, TPX2312-327.

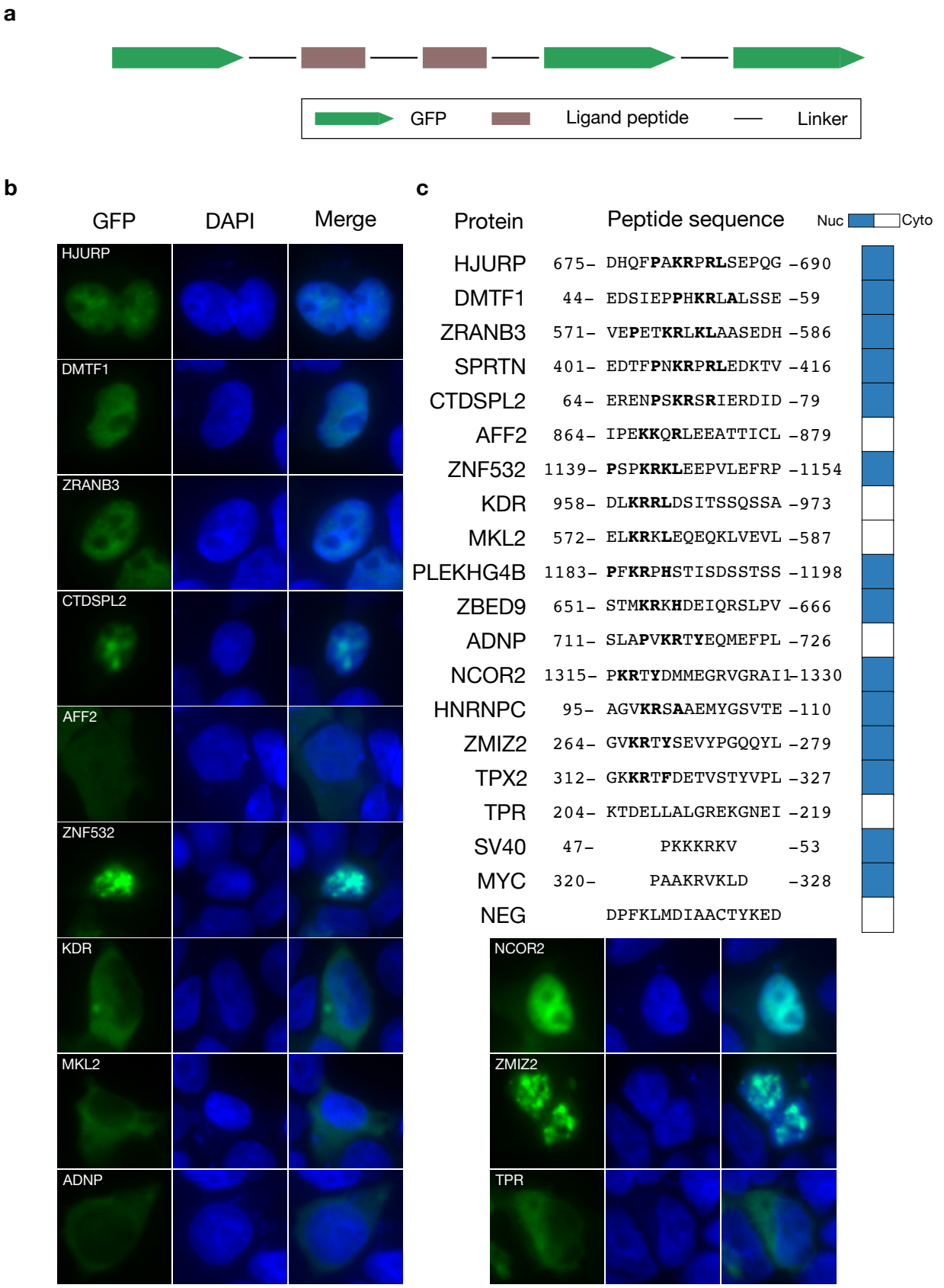
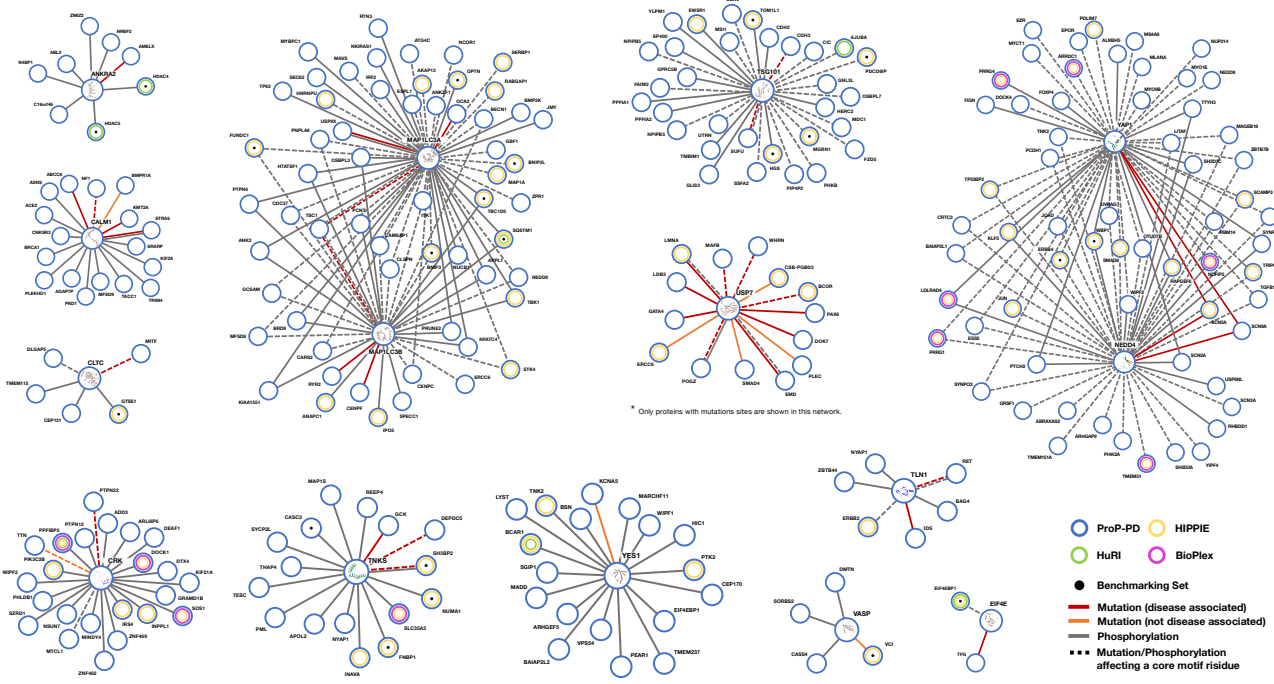


Fig S8. Validation of KPNA4 binding sequences as NLSs. (a) Schematic of the construct used as NLS sensor. (b) Cellular localization of NLS sensor with different peptides inserted. HEK293 cells were transiently transfected with the NLS sensor and fixed 36 h after transfection, and imaged using epifluorescence microscopy. The nucleus was stained with DAPI. (c) Overview of the peptide sequences fused to the NLS sensor, with the resulting cellular localization indicated.

a



b

ANKRA2

AMELX	169-PMLPDLTLEAWPSTDK-184
HDAC4	344-PLYTSPSLPNIHGLGLP-359
HDAC5	364-YTSPSLPNISLGLQAT-379
C16orf46	371-VFPPPVLPSLVPSVI-386
N4BP1	824-LPQQPHFPVLPALPSLG-839
ABL2	561-SSSSVVPPLPRLPILP-576
ZMIZ2	390-NQEVHSPFLPDLKPNL-405
NRBF2	255-DIPIPNLPPLDFTSPE-270

CALM1

ABCC6	208-AAFPSKATFWWVSGLV-223
ASNS	532-WLSHYWMPKWINADDP-547
ACE2	1-MSSSSWLLLSLVAVTA-16
NF1	779-DTHAKWEQATKLILNY-794
BMPR1A	438-WEMARRCITGGIVEEY-453
KMT2A	3689-AWKSLTDKVQEARSNA-3704
STRA6	639-WGLAIPLLHNPTLQVF-654
SRARP	148-STWGLVKDSLKALSSC-163
MFSD9	101-CWSSVVGRRSSLLACI-116
PKD1	3418-LSWPDLLSDPSIVGSN-3433
AGAP7P	320-KWHSLALACTPISSS-335
KIF2A	603-VSPQLPTPHEAVSQMV-618
BRCA1	372-WIPEAVSDIQKVNEWFS-387
CNKSR3	538-TEPSLLVSNFTRLKLL-553
PLEKHO1	366-WSDAKRVLQEVRELRD-381
TACC1	4-SSWQLLSPVQWAKWTW-19
TRIM4	334-SASSWPVFSSAWNYFA-349

CLTC

CEP131	92-PTEPTDFLMLFESPS-107
DLGAP5	831-PGGNLIDPSPLQPGEF-846
MITF	252-ANTLPVSGNLIDLYGN-267
GTSE1	648-LPLIDFCTPEAHVAV-663
TMEM115	336-APESSLITFEAAPPTL-351

CRK

ADD3	12-TPPPPSMPHKERYFDR-27
ARL6IP6	80-SPVLPDKRNGIFPAAA-95
DEAF1	425-ALAVPPTPTKAAPPA-440
DOCK1	1794-DVADVPPLPLKSVA-1809
DTX4	311-SGVTPVPKNLNGSSP-326
GRAMD1B	269-TKPDASQLPKKSTTN-284
INPPL1	139-PPLPPPSSTSISAPT-154
IRS4	823-NDNSETVMPLPGKFLG-838
KIF21A	1209-KELSPPPGLPSKIGSI-1224
MINDY4	286-TASSPPHLPSKRLPPW-301
MTCL1	1395-LMDISPFLPSKGLPST-1410
NSUN7	580-LSETVTKPPLPQKNTA-595
PHLDB1	943-PHSSPPPLPAKASRQL-958
PIK3C2B	152-KKLSPPPLPPRASIND-167
PPFIBP2	415-PFLAEHKPFLPGKLS-430
PTPN12	672-LSPPPLPEPFESFVL-687
PTPN22	612-EIPPPLPVRTPESFIV-627
SOS1	1218-SPLLPPLPEPEVRTPD-1225
SZRD1	76-NSTSPPHLPVKSLAQR-91
TTN	10960-PEVPKKPVPEKKVPVP-10975
WIPF2	268-PTNESAPELPQRHNSL-283
ZNF462	283-TNLPDVPNFSAPSPTS-298
ZNF469	1511-APSLPGKSGCSVALM-1526

EIF4E

TFG	97-LESSQVKYLRRELIEL-112
EIF4EBP1	52-IIYDRKFLMECRNSPV-67

MAP1LC3A/B

MAP1LC3A binders

AKAP13	597-ISDGLEVYLLAAGIG-612
ANKZF1	526-LLSAPLPSSGGPLLLHA-541
BMP2K	1126-VPTTELVVSSITPHQS-1141
BNIP3L	32-LNSSWVELPMNSSNGN-47
ESPL1	1519-TPAPGPEAASGEWELL-1534
GBF1	1334-RSATDADVVNSSGWLVV-1349
MAP1A	1784-EEDKLTPSPFEIISPP-1799
USP9X	1731-YECEESFTTLNVDIRN-1746
OCA2	665-ADIHDFEIILHRVEWA-680
ZPR1	189-ASPFTLIIDDPSGNSF-204
HNRNPU	466-YFPIPEEYTSIQNVPL-481
NKIRAS1	158-IEPFTLLASKLSQPQS-173
SEC62	359-NGNDFEMIIKEELEQQ-374
TP63	51-SPEVEPQHIWDFLEQP-66
IRF2	148-SNGVSDLSPEYAVLFS-163
MAVS	430-SPPSSGCFEDLAISAST-445
OPTN	168-LNSSGSSSDSFVEIRM-183
RABGAP1	16-SVSTLNSSEDFVLVSRQ-31
JMY	942-VLRESHLLPDTDPLT-957
MYBPC1	28-AGTTPAKDWTLVSPP-43
SERBP1	16-RFDQLFDDSDSPFEVL-31
NCOR1	2423-EPAPLLSAQVHLHSDS-2438
ATG4C	443-EKKQLKRFSTEEFVLL-458
RTN3	236-DDVIEKLSPESSFEVI-251

MAP1LC3A / MAP1LC3B binders

FUNDC1	8-PQDYESDDSYEVLDL-23
BRD9	577-DVTHDIYEFLKSPEPA-592
BECN1	92-ESANSPTTLIGEASDGG-107
MFSD8	16-DTPGSWEWDLLETEEH-31
NFATC4	242-GPEDSWLLSAPGPTP-257
STK4	428-PQDGIWSFLKSVTVED-443
CARS2	545-STTSIWELLDQRTKDQ-560
TBK1	163-ELEDDEQFVSHIGTEE-178
KIAA1551	1706-SPSRDGFEMLQNPVKD-1721
CENPC	528-SDWNVVKSEESPVYSN-543
CDC37	1-MVDLSWWDHIEVSDDE-16
ANK2	1588-VEEEWVLVSEEIEEA-1603
SQSTM1	334-CDDDWDLLSSKEVDPS-349
HTATSF1	120-AESSWFWVEEDRNTNV-135
BNIP3	80-QGSWVELHFSNNGNGG-95
ITK	195-NEEVQLLDSSEIHWWR-210
CAMSAP1	725-SHDSSEPWTLLRQDSDS-740
PCNT	644-SQELPWVHLQGVQDGD-659
TSC1	400-LCHSDDFVHHSLPQAT-415
TBC1D5	778-DDDSSKDSGFTIVSPL-793
APPL1	676-PNQASSEGQFVVLESS-691
PNPLA6	1277-FFSSGFDLAEIVSRI-1292
PTPN6	245-WEEFSSLQQKEVKNLH-260
NEDD9	619-DGSERSWNDIYDVHL-634
NUCB1	328-DTGEGWEIVEMHPAYT-343
CLSPN	764-SLLTKESHSNSSFELI-779
PRUNE2	2722-EWEMLSPQPVQKNMIP-2737
GCSAM	141-AHSPEDEYLLLMPHRI-156
OSBPL3	32-QDSWEVVEGLRGEMNY-47

MAP1LC3B binders

CENPF	1382-APLDESNSYEHLTLSD-1397
RYR2	1340-DSDFEVLMKTAHGHLV-1355
SPECC1	326-SPDASDFEHITAETPS-340
ANAPC1	702-GSDDWEYLLNSDYHQ-717
ERCC6	1278-DWVLVEAEANRVAQDA-1293
IPO5	649-ENMSDDGWEFVNLGD-664

TLN1

RET	1051-ALPSTWIENKLGMSD-1066
ERBB2	1238-KGTPTAENPEYLGLDV-1253
IDS	144-NHTDDSPYSWSFPPYH-159
NYAP1	539-DPTVGPLTPLWTVPATI-554
ZBTB44	222-VDSSLAFPWTFPFGID-237
BAG4	181-TPVSRWIYPQQDCQTE-196

TNKS

APOL2	246-IGRISAEGGEQVERVV-261
CASC3	144-GERQGDGQESTEPVE-159
DEPDC5	671-SNSNQFPGDGMSFLNFS-686
FNBP1	513-QDRESPDGSTTEQSQ-528
GCK	58-LPTYVRSTPEGSEVGD-73
INAVA	640-LREISPDGAPVQVFVPE-655
MAP1S	642-ESHNSPAEGSERLSLS-657
NUMA1	1742-PRDQPDGTSVPGEPAS-1757
NYAP1	348-RTEPGDGVSKLPVLCHS-363
PML	524-GPPSFSEVIGSEVFL-539
REEP4	170-PLYLEDQVSRREPIG-185
SH3BP2	408-PQLPHLQKEPPDQSF-423
SLC35A5	391-RDLSGNLMESGDGE-406
SYCP2L	679-TEERELPEGISSSLE-694
TESC	155-SGIADGAMMEAASVCM-170
THAP4	163-SQEQAQQALERPGDG-178

TSG101

CBX6	295-LLPETVDSPSWREP-310
CDH2	842-AADNPTAPFDSLLV-857
CDH3	765-ANTDPTAPPYDTLLVF-780
CIC	892-VPSAPPPKAQSVSPVQ-907
FZD5	1-MARPDPSAPFSLLLL-16
GNL3L	40-VPSAPHFVHPNDHANR-55
GPRC5B	384-VLNGFLPTAPPSHTG-399
HERC2	3481-PSAVDPSAPSASARPF-3496
MDC1	1156-NTPEPVVTAPPELQPS-1171
MGRN1	401-VSPAIPSAPLPEITY-416
MSI1	249-ERTPLPSAPVLPELTA-264
PIP4P2	12-LLSASHSGNVLTAPP-27
SSFA2	1151-ASVALLPTASRTGSV-1166
SUFU	11-GPTAPPAPGPTAPPAF-26
TMBIM1	1-MNPSAPPYEDRNPL-16
UTRN	867-SQPSAPDFVQRGFDSF-882
YLPM1	1523-GSIVRPSAPPRSSVP-1538
AJUBA	109-FRLEPPAPALSPRSSF-124
EP400	46-ASPSAPQSPSYQIQQL-61
FAIM2	92-PSAPPSYEEATSGEGM-47
HGS	344-KSPSPSAPVPITPEAA-359
NPIPB5	556-QLTPLPPSAPSGADDN-571
OSBPL7	225-PSAPVPITPHQASVTTE-240
TOM1L1	309-EPSAPSQDLLDLSPSP-324
GLIS3	610-ERFAPSAPSPHISPR-625
PHKB	699-QSTTPSAPELGQQPDV-714
PDCD6IP	707-KDLQQSIARESPAPSEI-722
EWSR1	76-GYTDPTAPQAYSQPVQ-91
PPFIA1	942-MSLTSPSAPPTSRTTL-957
PPFIA2	956-LAIQEMVSLTSPSAPP-971

USP7 *

BCOR	1422-LQQLLPASQSTQLPCS-1437
CSB-PGBD3	216-DAEPGPSSSLGSMLMPV-231
DOK7	351-SLSSYASGSSLDVWRAT-366
EMD	52-SSPAPSSYSFSDLNST-67
ERCC6	216-DAEPGPSSSLGSMLMPV-231
GATA4	349-NSSNSATTSSSSEMRPI-364
LDB3	342-HTAISSASSTTAPASSP-357
LMNA	607-VGGPISSKESASSVTV-622
MAFB	56-SSTPLSTPCSSVPSSP-71
PAX6	279-SNTPSHIPISSSFSTS-294
PLEC	980-SSSGSEAAVPSVCFLV-995
POGZ	252-TPSTSTPTATQPTSL-267
SMAD4	212-ANFPNIPVASTSQPAS-227
WHRN	608-QPFSSMPSCSGTVFSA-623

VASP

VCL	841-DPPPPPPDLEQLRLTD-856
DMTN	184-PAKIETDVNPCPPSLA-199
SORBS2	1005-EDSPDPPIPHSYSSDR-1020
CASS4	127-TAQVDEFPDPPTSARI-142

YAP1/NEDD4

YAP1 binders

MS4A8	235-PVTSPPSVSSEIQANK-250
MYCT1	219-TPPPPAVSGIIKAFPD-234
MLANA	99-NAPPAVEKLSAEQSPP-114
DOCK4	1896-ESKHPPPVSVYERTLR-1911
FOXP4	620-SSSPPRISPPQYSHQV-635
ARRDC1	400-ILPPEVSSWGPYEAP-415
MYO9B	2142-DPPLYCLPPASGQTNG-2157
MYO1E	955-APPPGGWHQNGVIRNQ-970
ALKBH5	300-SVLPSSYASDRLSGNN-315
NEDD9	107-QVPPSYQNQGIYQVPT-122
PRRG4	193-VALTRKHSVSPPPYP-208
EPOR	493-LIPAAEPLPSWVACS-508
PDLIM7	96-APAADPPRYTFAPSVS-111
EZR	466-VMTAPPPPPPPWPEPV-481
FIGN	392-SSQSSRALPPPSYSTA-407
NUP214	492-SATVTGEPPSWSSSD-507

YAP1 / NEDD4 binders

PRRG1	193-PPPSWSDIVNSNSASAI-208
BAIAP2L1	431-SVVIPPPTWLECLSMG-446
CRTC3	436-TEAQAQNSPPPWPAP-451
ERBB4	1045-RSEIGHSPPPAWTPMS-1060
ESS2	96-RPPPPWWTATFETP-111
JCAD	328-LCLSDPGLEPPVWVPP-343
JUN	160-SASLHSEPPVWANLSN-175
KLF5	316-AEMLQNLPPPSYAAT-331
LDLRAD4	249-EGPPPPLYSEVMGHHPG-264
NDFIP2	172-PVPPPWSVATSLPTYD-187
OTUD7B	763-ALLPPPWKVADSYSNG-778
RAPGEF6	1521-EGPHTHLKPPDYSVAV-1536
RBM14	570-NSTPPPWERTRLSPPR-585
SCAMP3	40-VYNPFETREPPPAWEP-55
SCN5A	1971-TSFPPSYDSVTRATSD-1986
SCN9A	1947-DATSSTTSPPSYDSVT-1962
SH2D3C	56-DMVTVPKSPPAYARSS-71
SMAD4	269-LCGPESPPPPWSRLSP-284
SYNPO	560-APPTYTETLSTAPLA-575
TGFB1I1	32-LDPPPEYSHQPQTGSG-47
TP53BP2	864-EEYPPLPPPPWPSGEP-879
TRIP6	124-EPPPPPPAPWSSLKPN-139
UVRAG	512-KRLQYKPPPPWSNSL-527
WBP1	128-EDVVHRFPPPWTVT-143
WIPF2	244-SGQSLAPPPPWAQPP-259
ZBTB7B	466-DAPWPPPSTAAASP-481
NDFIP1	57-DESGFKPPEYVVATT-72
PTCH2	10-LPSWVPPARTAAPQI-25
SCN2A	1964-DMTPSTTSPPSYDSVT-1979
LITAF	52-DGKGMNPPSYTTQPAP-67
MAGEB18	255-EYQQVPNSDPPWPEFL-270
PCDH1	1025-NPPKMPSKQVGQPFQL-1040
SYNPO2	614-ADFPAPPPYSAVTPPP-629
TNK2	623-DWDARPLPPPAWDDV-638
TTYH3	487-ENTPLIGRESPPPSYT-502

NEDD4 binders

ABRAXAS2	301-ASDPPPPYSDFHPNNQ-316
GRSF1	125-YLEDLPPPPPWSLAPS-140
ARHGAP9	181-PQPLMSEPPVWSNLVD-196
PI4K2A	12-RAQPPDWTFPSGSGAH-27
SH2D3A	224-ASERPPPYSELVPRVP-239
RHBDD1	289-RNSPPPYGFHLSPEEM-304
SCN3A	1960-GSSSTTSPPSYDSVTK-1975
TMEM151A	433-DTEPLESPPSWSDALY-448
TMEM51	229-SIEPLTPPPQWDEVQE-244
YIPF4	1-MQPPGPPPAWPATPGD-16
USP6NL	625-GLAHPPSYSSNPPVYH-640

YES1

KCNA5	73-PGVRPLPPLPELLPRP-88
WIPF1	388-NGSTSRALPAPQLPS-403
PTK2	362-PQKEGERALPSIPKLA-377
BSN	2893-VRKIPSPPPEEAHLP-2908
EIF4EBP1	68-TKTPPRDLPTPSPVTS-83
PEAR1	928-IRDLPSLPGPPRESSY-943
VPS54	24-VDPSKHIRPVPSLPDV-39
ARHGEF5	828-SAPGSSRIPRPLPPLP-843
MADD	731-AAVGVSKPLPSVPSHI-746
BCAR1	627-DKTSSIQSRPLPSPPK-642
March11	46-PRQLPPLPASPETPER-61
HIC1	255-PERPLAERELPPRPLS-270
CEP170	643-RDLPQLPNEEKSLESH-658
BAIAP2L2	284-ASQLEPDRRSLPPPLPS-299
TMEM237	16-PPRALPPVPVSQDDIPL-31
LYST	2103-SSPAPPHPSSLLTQS-2118
SGIP1	245-TGTPPPLPFKNVPNPP-260
DLGAP2	611-NSPPPVPPRTTSKPLI-626
TNK2	627-RPLPPPPANDDVAQDE-642

Fig S9. Additional PPI networks based on peptides with reported disease-associated mutations or phosphosites. (**a**) Networks showing interactions of reproducibly selected high/medium confidence peptides with overlapping disease-associated mutations or phosphorylation. Peptides with mutations or phosphorylation residues occurring in the core motif or in the flanking regions (+/- 2 residues) are shown. Mutations associated with diseases are colored in red (orange if they are not disease associated). Phosphorylated sites are colored in gray. Dashed-edge lines represent core motif residues holding a mutation or a phosphorylation site. (**b**) Sequences of high/medium confidence peptides with disease-associated mutations in the binding motif. Motif-containing regions are highlighted with blue background, core motifs are indicated in bold letters, phosphorylation sites within or in vicinity of the motifs are indicated by a box, and disease associated mutations are indicated in red bold letters (mutation of core residue).