Article

A reference map of the human binary protein interactome

https://doi.org/10.1038/s41586-020-2188-x

Received: 19 April 2019

Accepted: 14 February 2020

Published online: 08 April 2020

Check for updates

Katja Luck[1,2,3,33], Dae-Kyum Kim[1,4,5,6,33], Luke Lambourne[1,2,3,33], Kerstin Spirohn[1,2,3,33], Bridget E. Begg[1,2,3], Wenting Bian[1,2,3], Ruth Brignall[1,2,3], Tiziana Cafarelli[1,2,3], Francisco J. Campos-Laborie[7,8], Benoit Charloteaux[1,2,3], Dongsic Choi[9], Atina G. Coté[1,4,5,6], Meaghan Daley[1,2,3], Steven Deimling[10], Alice Desbuleux[1,2,3,11], Amélie Dricot[1,2,3], Marinella Gebbia[1,4,5,6], Madeleine F. Hardy[1,2,3], Nishka Kishore[1,4,5,6], Jennifer J. Knapp[1,4,5,6], István A. Kovács[1,12,13], Irma Lemmens[14,15], Miles W. Mee[4,5,16], Joseph C. Mellor[1,4,5,6,17], Carl Pollis[1,2,3], Carles Pons[18], Aaron D. Richardson[1,2,3], Sadie Schlabach[1,2,3], Bridget Teeking[1,2,3], Anupama Yadav[1,2,3], Mariana Babor[1,4,5,6], Dawit Balcha[1,2,3], Omer Basha[19,20], Christian Bowman-Colin[2,3], Suet-Feung Chin[21], Soon Gang Choi[1,2,3], Claudia Colabella[22,23], Georges Coppin[1,2,3,11], Cassandra D'Amata[10], David De Ridder[1,2,3], Steffi De Rouck[14,15], Miquel Duran-Frigola[18], Hanane Ennajdaoui[1,4,5,6], Florian Goebels[4,5,16], Liana Goehring[2,3], Anjali Gopal[1,4,5,6], Ghazal Haddad[1,4,5,6], Elodie Hatchi[2,3], Mohamed Helmy[4,5,16], Yves Jacob[24,25], Yoseph Kassa[1,2,3], Serena Landini[2,3], Roujia Li[1,4,5,6], Natascha van Lieshout[1,4,5,6], Andrew MacWilliams[1,2,3], Dylan Markey[1,2,3], Joseph N. Paulson[26,27,28], Sudharshan Rangarajan[1,2,3], John Rasla[1,2,3], Ashyad Rayhan[1,4,5,6], Thomas Rolland[1,2,3], Adriana San-Miguel[1,2,3], Yun Shen[1,2,3], Dayag Sheykhkarimli[1,4,5,6], Gloria M. Sheynkman[1,2,3], Eyal Simonovsky[19,20], Murat Taşan[1,4,5,6,16], Alexander Tejeda[1,2,3], Vincent Tropepe[10], Jean-Claude Twizere[11], Yang Wang[1,2,3], Robert J. Weatheritt[4], Jochen Weile[1,4,5,6,16], Yu Xia[1,29], Xinping Yang[1,2,3], Esti Yeger-Lotem[19,20], Quan Zhong[1,2,3,30], Patrick Aloy[18,31], Gary D. Bader[4,5,16], Javier De Las Rivas[7,8], Suzanne Gaudet[1,2,3], Tong Hao[1,2,3], Janusz Rak[9], Jan Tavernier[14,15], David E. Hill[1,2,3]✉, Marc Vidal[1,2]✉, Frederick P. Roth[1,4,5,6,16,32]✉ & Michael A. Calderwood[1,2,3]✉

Global insights into cellular organization and genome function require comprehensive understanding of the interactome networks that mediate genotype–phenotype relationships[1,2]. Here we present a human 'all-by-all' reference interactome map of human binary protein interactions, or 'HuRI'. With approximately 53,000 protein–protein interactions, HuRI has approximately four times as many such interactions as there are high-quality curated interactions from small-scale studies. The integration of HuRI with genome[3], transcriptome[4] and proteome[5] data enables cellular function to be studied within most physiological or pathological cellular contexts. We demonstrate the utility of HuRI in identifying the specific subcellular roles of protein–protein interactions. Inferred tissue-specific networks reveal general principles for the formation of cellular context-specific functions and elucidate potential molecular mechanisms that might underlie tissue-specific phenotypes of Mendelian diseases. HuRI is a systematic proteome-wide reference that links genomic variation to phenotypic outcomes.

The reference human genome sequence has enabled systematic study of genetic[6] and expression[4] variability at the organism[6], tissue[4], cell type[7] and single-cell level[8]. Despite advances in sequencing genomes, transcriptomes and proteomes, we still understand little about the cellular mechanisms that mediate phenotypic and tissue or cell type variability. A mechanistic understanding of cellular function and organization emerges from studying how genes and their products, primarily proteins, interact with each other, forming a dynamic interactome that drives biological function. Analogous to the reference human genome sequence[9,10], a reference map of the human protein interactome, generated systematically and comprehensively, is needed to provide a scaffold for the unbiased proteome-wide study of biological mechanisms, generally and within specific cellular contexts.

It remains infeasible to assemble a reference interactome map by systematically identifying endogenous protein–protein interactions (PPIs) in thousands of physiological and pathological cellular contexts[11,12]. However, systematic affinity purification of exogenously expressed bait proteins in immortalized-cell contexts[13] as well as binary PPI detection assays in cell models[2,14] have generated biophysical human protein interactome maps of demonstrated high functional relevance. Specifically, yeast two-hybrid (Y2H) represents the only binary PPI assay that can be operated at sufficient throughput to systematically screen the

A list of affiliations appears at the end of the paper.

human proteome for binary PPIs. Using Y2H followed by validation in orthogonal assays, we previously generated HI-II-14 consisting of approximately 14,000 PPIs involving 4,000 proteins from screening around 40% of the genome-by-genome search space[2]. In contrast to curated interactions from hypothesis-driven small-scale studies and protein complex interactome maps that favour highly expressed proteins, HI-II-14 covers the proteome more uniformly and is relatively free from ascertainment and expression biases.

To increase interactome coverage and generate a reference map of human binary PPIs, we expanded the ORFeome collection to encompass approximately 90% of the protein-coding genome. We screened this search space a total of nine times with a panel of three Y2H assay versions (Fig. 1a, b). The resulting PPI map quadruples the number of identified PPIs overall, and upon integration with genome, transcriptome and proteome resources enables biological discovery across most cellular contexts, offering a reference map of the human binary protein interactome.

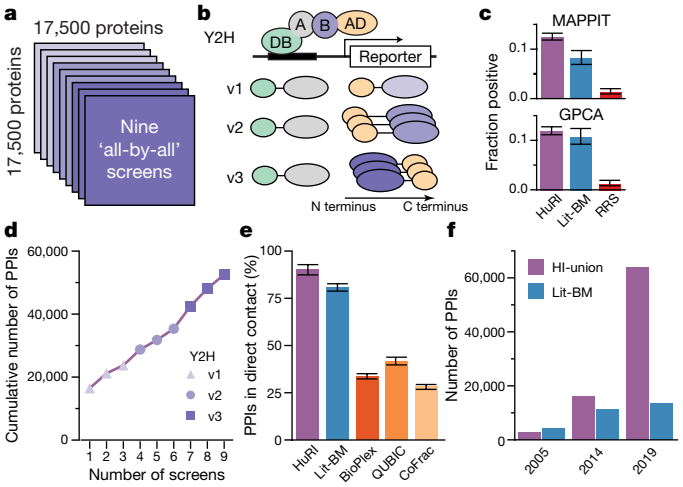


Fig. 1 | Generation of a reference interactome map using a panel of binary assays. a, Overview of HuRI generation. **b**, Schematic of the Y2H assay versions. **c**, Experimental validation. Lit-BM, literature-curated binary PPIs with multiple evidence; RRS, random protein pairs. Error bars are 68.3% Bayesian confidence interval. MAPPIT: $n = 2,281, 383$ and 475; GPCA: $n = 1,639, 382$ and 465 (from left to right). **d**, Number of PPIs detected with each additional screen. **e**, Fraction of direct contact pairs among five PPI networks. Error bar is standard error of proportion. $n = 121, 410, 1,169, 584$ and 1,211 PPIs (from left to right). **f**, Number of PPIs identified over time from screening at the Center for Cancer Systems Biology (CCSB) and Lit-BM.

Generation and characterization of HuRI

The newly established human ORFeome v9.1 covers 17,408 protein-coding genes (Extended Data Fig. 1a, b, Supplementary Tables 1, 2), forming a search space (Space III) that encompasses over 150 million pairwise combinations (Fig. 1a). This search space more than doubles the space screened to generate HI-II-14 and represents the most comprehensive search space systematically screened for human PPIs. Limitations in sensitivity of the PPI assay can be overcome by using different assays[15,16] or different versions of the same assay[17,18]. To maximize sensitivity we used three Y2H assay versions (Fig. 1b, Supplementary Tables 3, 4) that showed good sensitivity and low false-positive rates when benchmarked against gold-standard positive and random reference sets (PRSv1 and RRSv1, respectively)[15], while detecting complementary PPI sets (Extended Data Fig. 1c, d, Supplementary Table 5). We assessed assay performance using a test space of approximately 2,000-by-2,000 human genes[2] (Extended Data Fig. 1e, Supplementary Table 6). The Y2H versions were complementary, in that three screens for each of the three versions doubled the number of detected PPIs and proteins relative to nine screens using a single version (Extended Data Fig. 1f, Supplementary Table 7).

To map the reference interactome, we performed nine screens of Space III, followed by pairwise verification by quadruplicate retesting and sequence confirmation. PPIs verified by two orthogonal binary PPI assays, MAPPIT[19] and GPCA[20], were recovered at rates on par with high-confidence binary PPIs from the literature (each having at least two pieces of experimental evidence, with at least one from a binary assay type; Lit-BM) over a large range of score thresholds (Fig. 1c, Extended Data Fig. 1g, h, Supplementary Table 8). Each additional screen identified new PPIs and proteins, with the largest gains obtained by switching assay versions (Fig. 1d, Extended Data Fig. 1i). The dataset, versioned HI-III-20 (Human Interactome obtained from screening Space III, published in 2020), contains 52,569 verified PPIs involving 8,275 proteins (Supplementary Table 9). Although our knowledge of the interactome remains incomplete, we refer to HI-III-20 as a reference map of the human binary protein interactome (HuRI) given its systematic nature, extensive coverage, and scale.

Molecular mechanisms can be more readily inferred from direct than indirect PPIs, yet the fraction of PPIs reported in various human protein interactome maps that are direct remains unknown. Structures from protein complexes with at least three subunits[21] show that, within those complexes, more PPIs in HuRI correspond to direct biophysical contacts than do PPIs from Lit-BM (90% versus 81%, $P = 0.019$, two-sided Fisher's exact test, $n = 121$ and 410 for HuRI and Lit-BM, respectively) or from protein complex interactome maps (<50%, $P < 0.001$ for all tested maps, two-sided Fisher's exact test, $n = 584–1,211$) (Fig. 1e, Supplementary Table 10). Combining HuRI with all previously published systematic screening efforts at CCSB yields 64,006 binary PPIs involving 9,094 proteins (HI-union) (Supplementary Table 11), which is approximately fivefold more PPIs than the entirety of high-quality binary PPIs curated from the literature (Fig. 1f, Extended Data Fig. 2, Supplementary Tables 12–14). The union of Lit-BM and HI-union represents the most complete collection of high-quality direct PPI data available to date (http://interactome-atlas.org).

Complementarity of the three Y2H versions might stem from steric constraints that differ between protein fusions used in the assays. Integrating HuRI with structures of PPIs[21], we observed reduced sensitivity of Y2H assays where the interaction interface was close (<20 Å) to whichever terminus was fused to the Gal4 activation domain (Extended Data Fig. 3a, b, Supplementary Table 15). PPIs found in several screens had larger interaction interfaces ($P = 0.03$, two-sided permutation test, $n = 234$) and corresponded more often to direct PPIs within, rather than between, protein complexes ($P = 3 \times 10^{-18}$, two-sided Fisher's exact test, $n = 1,817$); however, HuRI PPIs found in a single screen were observed to have precision as high as those found in multiple screens (Extended Data Figs. 3c–g, 4, Supplementary Tables 16–18, Supplementary Note 1). These results reinforce previous observations[12,22] that the protein interactome might be dominated by weaker and more transient PPIs that are difficult to detect, as indicated by the fact that most PPIs in HuRI were found in only one screen (Extended Data Fig. 3h, i). This is reflected in an increased estimate of the binary protein interactome size (Extended Data Fig. 3j), so that HuRI is estimated to represent 2–11% of the binary protein interactome[23] (Supplementary Note 2).

Functional relationships in HuRI

On the basis of the observation that HuRI is enriched in direct PPIs, we also hypothesize that proteins in HuRI with similar interaction interfaces tend to share interaction partners. For example, retinoic acid receptors RXRG and RXRB (Extended Data Fig. 5a, left) share previously reported interaction partners involving binding to retinoic acid receptor RAR types[24] and oxysterol receptors NR1 group H types[25]. We derived a profile similarity network (PSN) from HuRI (Supplementary Table 19), and found that the number of pairs of proteins in HuRI with

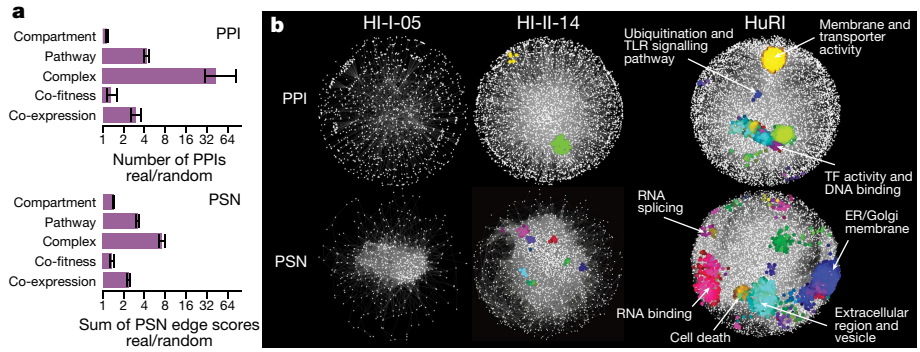


Fig. 2 | Complementary functional relationships in HuRI between genes. **a**, Enrichment of HuRI and its profile similarity network (PSN) for protein pairs with shared functional annotation, showing mean and 95% interval of 100 random networks. **b**, Functional modules in HuRI and its PSN and in previously published interactome maps from CCSB.

similar interaction profiles is higher than random ($P < 0.01$, one-sided empirical test) (Extended Data Fig. 5b) and proteins of overall higher sequence identity tend to exhibit higher interaction profile similarities ($P < 0.01$, one-sided empirical test) (Extended Data Fig. 5c). We also found that proteins with a tendency to share interaction partners often have interaction interfaces that are similar, as opposed to complementary, and therefore tend not to interact with one another, except where proteins originate from a common ancestor that could self-interact[26] (Extended Data Fig. 5d). Indeed, for only 5% of the protein pairs found to interact in HuRI did the pair share more than 10% of their interaction partners. Both HuRI PPIs and the PSN are enriched for links between proteins of similar function ($P < 0.01$, one-sided empirical test) (Fig. 2a, Supplementary Table 20, Extended Data Fig. 5e, f) and both contain more functional modules[27] than our previously published interactome maps[2,14] (Fig. 2b, Extended Data Fig. 5g). Thus, HuRI and the PSN are complementary maps of functional relationships between proteins.

As shown above, global sequence identity between two proteins can be indicative of shared interaction interfaces; however, it probably fails to identify pairs of proteins whose shared interaction interface is small. Indeed, 50% (502) of all protein pairs in HuRI with interaction profile similarities ≥0.5 exhibit ≤20% sequence identity, so that functional relationships identified by the PSN are not necessarily identifiable by sequence identity. One such protein pair is the endoplasmic reticulum (ER) transmembrane protein TMEM258 and the uncharacterized protein C19ORF18, which have only 10% sequence identity but share 80% of their interactors (Extended Data Fig. 5a, right). TMEM258 catalyses the first step in *N*-glycosylation of proteins in the ER and might have a role in protein translocation across the ER[28]. Roles in protein transport and ER function have also been ascribed to two of the four shared interaction partners, ARL6IP1[29] and IER3IP1[30], which suggests that C19ORF18 and potentially the other two shared interaction partners (MFSD6 and AC012254.2) contribute to ER-related functions of protein maturation and transport.

Uncharted disease-related interactions

Unlike Lit-BM, HuRI was generated by systematically testing protein pairs for interaction. While Lit-BM is highly biased towards the most studied genes[2], HuRI covers the genome-by-genome space more uniformly and at increased depth compared to Lit-BM and our previous screening efforts (Fig. 3a, Extended Data Fig. 5h). Notably, we find that the agreement between Lit-BM and HuRI is highest among the best-studied genes, in which Lit-BM is most complete and where approximately 40% of the PPIs in HuRI have been previously identified (Fig. 3b). HuRI substantially expands the number of biomedically interesting genes for which high-quality direct PPI data are available (Fig. 3c, Extended Data Fig. 5i), and finds new interaction partners for these genes in previously uncharted regions of the protein interactome (Fig. 3d, Extended Data Fig. 5j).

Essential proteins were often found to have significantly more interaction partners[31]. However, correlation between two variables does not necessarily indicate causality, especially where there are other confounding variables (Extended Data Fig. 5k). We find that protein popularity (measured by publication count) and endogenous expression level strongly correlate with: (i) each other; (ii) the number of interaction partners ('degree') in protein complex and literature-curated protein interaction networks but not in HuRI; as well as (iii) gene properties such as essentiality, age[32], fitness effect[33] and loss-of-function intolerance[6] (Extended Data Fig. 5l, Supplementary Table 21). After correcting for popularity and expression level, we found substantially reduced correlations between interaction degree and other gene properties, including gene essentiality (Fig. 3e, Extended Data Fig. 5m, n). In line with previous observations[34], this suggests that correlations between degree and essentiality, age, loss-of-function intolerance, and fitness effects are confounded by underlying expression and study biases, and thus may not reflect causal relationships. These results highlight the value of HuRI as a uniformly-mapped reference for the study of systems properties of genes and networks.

Compartment-specific roles of PPIs

Proteins are localized to specific compartments, carrying out functions that depend both on the subcellular environment and the local PPI network. Despite available proteome-wide datasets on the localization of individual proteins[5], experimental determination of cellular localization-specific PPI networks remains challenging. We find that proteins localizing to a diverse range of subcellular compartments are evenly represented in HuRI (Extended Data Fig. 6a), suggesting that cellular localization-specific PPI networks can be inferred for many different cellular compartments via integration of HuRI with available protein localization data.

Extracellular vesicles have been studied intensively using proteomics approaches[35]. However, our understanding of the molecular mechanisms that lead to protein recruitment into extracellular vesicles and subsequent secretion remains limited. The subnetwork of interactions between extracellular vesicle proteins (Extended Data Fig. 6b) shows significantly higher connectivity in HuRI than in degree-controlled randomized networks ($P < 0.001$, one-sided empirical test) (Extended Data Fig. 6c), enabling prediction of recruiters of extracellular vesicles using the number of interaction partners. Seven of the top twenty-one most connected proteins in this extracellular vesicle network have established roles in extracellular vesicle biogenesis or cargo recruitment[36]. SDCBP (syntenin-1) functions in ESCRT-dependent generation of extracellular vesicles, and its knockout shows reduced extracellular vesicle production[37]. SDCBP has 48 PPIs with other extracellular vesicle

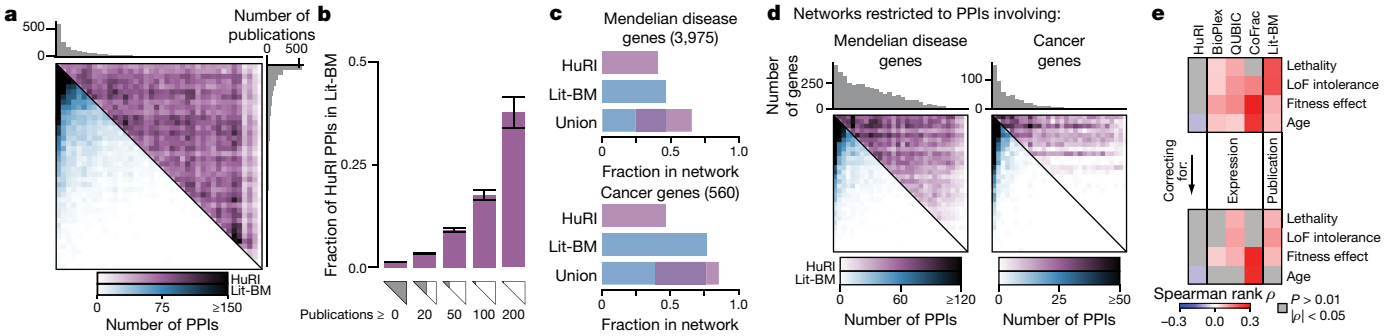


Fig. 3 | Unbiased proteome coverage of HuRI reveals uncharted network neighbourhoods of disease-related genes. a, Heat maps of PPI counts, proteins ordered by number of publications. **b**, Fraction of HuRI PPIs in Lit-BM, for increasing values of the minimum number of publications per protein. Error bar is standard error of proportion, $n = 52{,}569$ to 170 PPIs (from left to right). **c**, Fraction of genes with at least one PPI for biomedically interesting genes. **d**, As in **a**, but restricted to PPIs involving genes from the indicated gene sets. **e**, Correlation between degree and variables of interest, before (top) and after (bottom) correcting for the technical confounding factors. $n = 13{,}441$–$53{,}704$ PPIs per network, P value determined by two-tailed permutation test. LoF, loss of function.

proteins and is frequently detected in extracellular vesicles, which suggests that it regulates recruitment of interacting proteins into extracellular vesicles. To test this hypothesis (Fig. 4a), we knocked out SDCBP in the U373vIII cell line (Extended Data Fig. 6d). Of six SDCBP partners detected in the U373vIII extracellular vesicle proteome, three (CALM1, CEP55 and HPRT1) displayed significantly reduced protein levels in extracellular vesicles in the SDCBP-knockout line ($P < 0.05$, one-sided empirical test, depletion <0.66) (Fig. 4b). By contrast, only 15% of the non-interaction partners of SDCBP were reduced ($P = 0.042$, one-sided empirical test) (Extended Data Fig. 6e). Thus, SDCBP may have a role in the recruitment of proteins into extracellular vesicles, highlighting the potential value of HuRI in studying protein function within specific subcellular contexts.

Proteins often change subcellular localization, for example, as part of their maturation process or in response to external or internal signalling events. These dynamics are difficult to comprehensively capture at proteome-scale, yet existing efforts have already revealed numerous proteins that localize to multiple subcellular compartments with some pairs of compartments exhibiting considerably more shared membership (crosstalk) than others[5]. Despite a tendency for interactions in HuRI to link proteins localized to the same compartment ($P < 0.01$, one-sided empirical test) (Fig. 2a), a considerable number of interactions were identified between proteins never reported to co-localize. To explore whether lack of co-localization of interacting proteins in HuRI might originate from incomplete localization data, we assessed whether these non-co-localized interacting proteins tend to reside in compartments with significant crosstalk and observed a significant positive correlation ($P < 0.001$, one-sided empirical test) (Extended Data Fig. 6f–h). This suggests that HuRI could prove useful in predicting protein localization dynamics.

Principles of tissue-specific function

Despite recent advances in systematic genome-wide identification of tissue-preferentially expressed (TiP) genes[4] (Extended Data Fig. 7, Supplementary Table 22), we lack a concrete understanding of how the surprisingly small set of TiP genes operate together and coordinate their activity with the core 'housekeeping' machinery to mediate tissue-specific functions. Insights can be obtained from investigation of the tissue-specific network context of TiP proteins, inferred from integrating protein interactome data with tissue transcriptomes (Supplementary Table 23). However, we should not expect uniform coverage of TiP proteins with PPIs using experimental methods that demand expression of both partners within a single immortalized cell line or result from screening an incomplete ORFeome. Indeed, contrary to protein complex[11–13] and literature-curated interactome maps[38] as well as our previously published binary PPI datasets[2,14], we find that TiP proteins are well-represented in HuRI (Fig. 5a, Extended Data Fig. 8a).

Restricting HuRI to PPIs between proteins expressed in the same tissue, we observe that TiP proteins engage in as many PPIs and are as central as more uniformly expressed proteins (Extended Data Fig. 8b), in contrast to previous observations using literature-curated PPI networks[39,40]. This result, paired with the fact that PPIs mediated by a TiP protein are effectively also tissue-specific, leads to the finding that the fraction of tissue-specific PPIs in the protein interactome as characterized by HuRI is higher than that of tissue-specific genes in the expressed genome, indicating that substantial information on tissue-specific functions can only be obtained from the interactome. The opposite is observed for Lit-BM, probably owing to its bias against TiP genes (Extended Data Fig. 8c).

To investigate the local network neighbourhoods of TiP proteins within their respective tissue contexts, we used HuRI to derive protein interactome maps for 35 tissues[4,41], each of which contains about 25,000 PPIs (Supplementary Table 24, Extended Data Fig. 8d). Within each tissue PPI network, we focused on the interactions involving at least one TiP protein (Fig. 5b). The TiP PPI networks show extensive interactions between TiP and non-TiP proteins, but with few TiP–TiP PPIs. Despite significant enrichments for HuRI to link proteins that work in the same biological process (Fig. 2a), TiP–TiP PPIs, as highlighted for brain in Fig. 5c, are not enriched, nor is the average shortest path among TiP proteins shorter than in degree-controlled randomized networks ($P > 0.05$, empirical test). Using either metric, TiP proteins were found to be significantly close to each other in only six of 35 tissues. In four of these six tissues, enrichment for network proximity was driven by clusters of specifically expressed keratins or late-cornified envelope proteins (Extended Data Fig. 8e). These results support a model in which

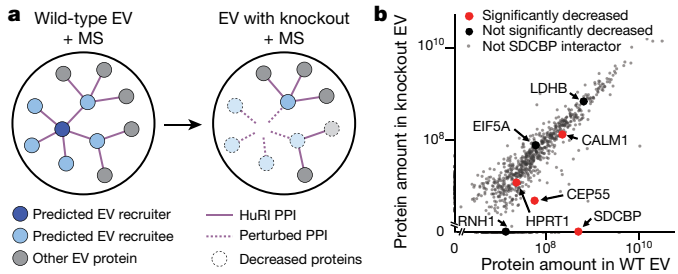


Fig. 4 | Identification of potential recruiters of proteins into extracellular vesicles. a, Schematic of experimental design to test the extracellular vesicle (EV) recruitment function of proteins. MS, mass spectrometry. **b**, Protein abundance from extracellular vesicle for each gene in wild-type (WT) and SDCBP knockout. Data are mean values of $n = 3$ biological replicates.

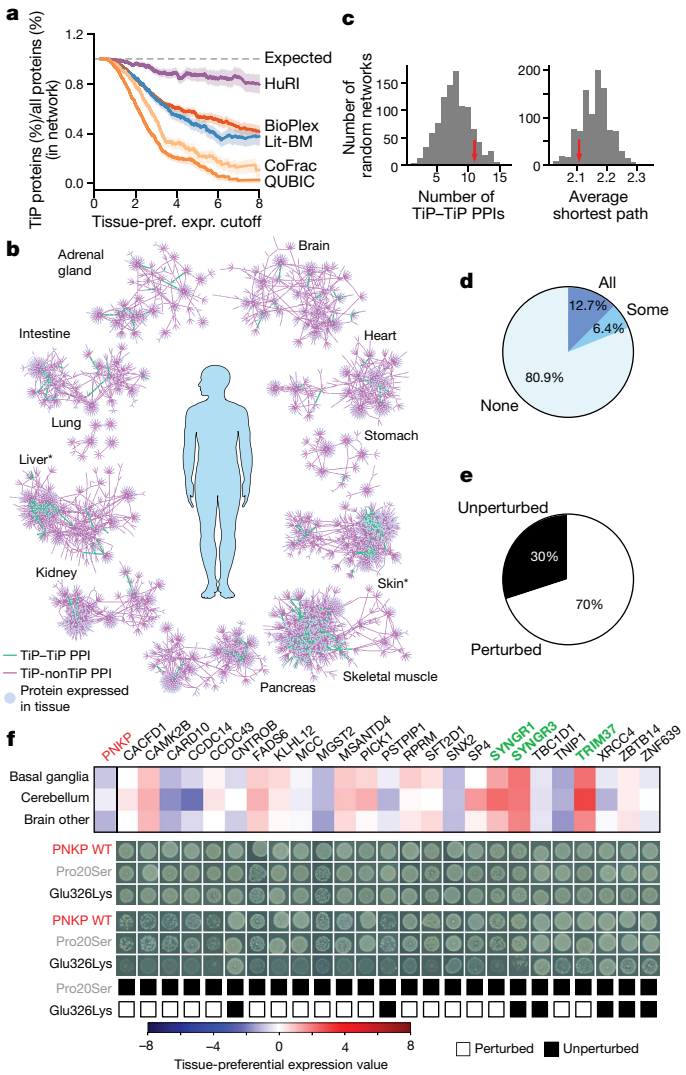


Fig. 5 | Tissue-specific functions are largely mediated by interactions between TiP proteins and uniformly expressed proteins. a, Coverage of TiP proteins by PPI networks for increasing levels of tissue-preferential expression (shaded error bands proportional to standard error of proportion, $n \geq 233$ genes). **b**, Tissue-preferential sub-networks. *$P \leq 0.0014$, one-sided empirical test for TiP proteins being close to each other ($n = 19{,}960–30{,}217$ PPIs per subnetwork, corrected for multiple testing). **c**, Empirical test of closeness of TiP proteins in the brain sub-network (red arrow), 1,000 random networks (grey). **d**, Tissue-specific diseases split by all, some, or none of the causal genes being tissue-preferentially expressed. **e**, Tested causal genes of tissue-specific diseases split by whether PPI to TiP interaction partner was perturbed. **f**, Expression profile of PNKP and interactors in brain tissues and PPI perturbation pattern of disease-causing (Glu326Lys) and benign (Pro20Ser) mutations. Yeast growth phenotypes on SC-Leu-Trp (top) or SC-Leu-Trp-His+3AT (3-amino-1,2,4-triazole) media (bottom). Green protein symbols denote preferentially expressed. Only interactors expressed in the brain are shown.

tissue-specific functions emerge through interactions between TiP proteins and more uniformly expressed members of the basic cellular machinery, presumably modulating and adapting common cellular processes for cellular context-specific needs[42].

One biological process with both cell-type and developmental-stage-specific homeostatic roles is apoptosis. We used HuRI to identify proteins with interaction partners that were enriched for known apoptosis regulators (Supplementary Table 25). Five proteins among the top ten predictions had previously demonstrated roles in apoptosis (Supplementary Note 3). Among three TiP genes predicted to be implicated in apoptosis (Extended Data Fig. 8f, g), we further examined OTUD6A. Abundance of OTUD6A negatively correlated with time-of-death after addition of TRAIL (TNF-related apoptosis-inducing ligand; $P = 0.012$, two-sided, empirical test, $n = 40$ cells), but not after expression of OTUD6A alone (Extended Data Fig. 8h, Supplementary Table 26), suggesting that OTUD6A participates in the apoptosis pathway but is not an inducer of cell death. This and other evidence (Extended Data Fig. 8f, i, Supplementary Note 3) suggests that OTUD6A exerts an apoptosis sensitization function via transcriptional activation in a haematopoietic cellular context.

Mechanisms of tissue-specific diseases

Many Mendelian diseases display tissue-specific phenotypes, rarely explained by tissue-specific expression of genes with disease-associated mutations[43] (Fig. 5d, Extended Data Fig. 9a). Such mutations broadly or specifically affect PPIs involving the mutated protein[44]. Perturbations of PPIs between uniformly expressed disease-associated proteins and TiP proteins in the affected tissues have been suggested to underlie the tissue-specific phenotypes of those diseases[43]. In HuRI-derived tissue PPI networks, we find 130 such PPIs involving 63 distinct non-TiP disease-causal proteins and 94 TiP proteins. Although we see no enrichment for PPIs between causal proteins and TiP proteins (Extended Data Fig. 9b, Supplementary Note 4), this does not rule out the possibility that perturbations of some of these interactions mediate tissue-specific phenotypes of Mendelian diseases.

To explore this hypothesis, we experimentally tested whether pathogenic variants associated with Mendelian diseases were able to perturb these PPIs. Of ten causal proteins tested, seven showed perturbation of PPIs to preferentially expressed interaction partners in the corresponding 'disease tissues' (Fig. 5e, Extended Data Fig. 9c, d, Supplementary Tables 27, 28). One example is PNKP, mutations of which have been associated with microcephaly, seizures and developmental delay. The pathogenic PNKP mutation Glu326Lys does not affect the DNA kinase or DNA phosphatase activity of PNKP, rendering the mechanism of pathogenicity unclear[45]. We observed that Glu326Lys perturbed PPIs with two partners preferentially expressed in the brain, SYNGR1 and TRIM37, whereas a benign control mutation Pro20Ser[46] did not affect any PNKP PPIs (Fig. 5f, Extended Data Fig. 9c, e). TRIM37 facilitates DNA repair[47], suggesting a potential mechanism by which perturbation of this interaction could affect the brain-specific DNA repair function of PNKP. In other examples, HuRI identified CTNNA3 and SUCLA2 to have respective TiP interaction partners TRIM54 and ARL6IP1 (Extended Data Fig. 10), which cause similar diseases with overlapping symptoms[48,49], supporting the relevance of these interactions in the physiopathology. Overall, this study yields hypotheses of molecular mechanisms for otherwise unexplained tissue-specific phenotypes of seven Mendelian diseases (Extended Data Fig. 9d) and demonstrates the utility of HuRI as a reference to study biological mechanisms within specific disease contexts.

Discussion

Here, we present HuRI, a systematically generated human protein interactome map with more than 50,000 PPIs of high biophysical quality. Although HuRI displays highly significant overlap with known functional relationships, the cellular function of most individual PPIs remains to be determined. We show that follow-up studies on the function of proteins and PPIs can be guided by integration of HuRI with contextual genome, transcriptome and proteome data to infer the cellular context in which subnetworks of PPIs operate together to mediate a function. With advances in single-cell transcriptomics[8] as well as systematic determination of subcellular protein localization, inference of functional PPI subnetworks that are specific for a given cellular state will further increase in precision. However, a priori removal of PPIs from HuRI because they are not currently known to function together

in a physiological cellular context could discard data that can help, for example, to understand the functional consequences of dysregulated gene and protein expression causing disease.

Despite our extensive screening, many PPIs remained undetected as Y2H, like all characterized assays[15,16,18], has limited sensitivity, detecting no interactions for half of the tested proteins. This negative result can guide the design of future interactome mapping efforts to target these proteins. Owing to the limitations of Y2H, we expect HuRI to be depleted for PPIs that depend on post-translational processing of human proteins that the yeast cell is unable to catalyse or that require additional partners to stabilize the interaction. Screening only one isoform per gene also misses interaction partners specific to alternative spliceforms[50]. Accurate estimation of the total size of the interactome remains challenging. PPIs display a continuum of binding strength or stability that, along with other factors, could underlie a continuum of detectability, as this study suggests. Furthermore, the results obtained by us and others[12] indicate that very stable and functionally conserved PPIs constitute a minority of the interactome.

Although incomplete, the uniform proteome and interactome coverage of HuRI enable its use as a reference for the study of most aspects of human cellular function. Efforts to further complete this reference will require development of new technologies as well as integration with complementary reference maps of protein complex assemblies[13]. Although multiple challenges remain to be solved for a complete and context-specific map of protein functions, interactions, and higher-level organization, HuRI provides an unbiased genome-scale scaffold with which to coordinate this information as it emerges.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2188-x.

1. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
2. Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
3. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
4. Melé, M. et al. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
5. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
6. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
7. The FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
8. Regev, A. et al. The human cell atlas. *eLife* **6**, e27041 (2017).
9. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
10. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
11. Wan, C. et al. Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).
12. Hein, M. Y. et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
13. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
14. Rual, J.-F. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
15. Braun, P. et al. An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91–97 (2009).
16. Venkatesan, K. et al. An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
17. Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods* **7**, 667–668 (2010).
18. Choi, S. G. et al. Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* **10**, 3907 (2019).
19. Eyckerman, S. et al. Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.* **3**, 1114–1119 (2001).
20. Cassonnet, P. et al. Benchmarking a luciferase complementation assay for detecting protein complexes. *Nat. Methods* **8**, 990–992 (2011).
21. Mosca, R., Céol, A. & Aloy, P. Interactome3D: adding structural details to protein networks. *Nat. Methods* **10**, 47–53 (2013).
22. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Mol. Cell* **55**, 161–169 (2014).
23. Sambourg, L. & Thierry-Mieg, N. New insights into protein-protein interaction data lead to increased estimates of the *S. cerevisiae* interactome size. *BMC Bioinformatics* **11**, 605 (2010).
24. Leid, M. et al. Purification, cloning, and RXR identity of the HeLa cell factor with which RAR or TR heterodimerizes to bind target sequences efficiently. *Cell* **68**, 377–395 (1992).
25. Willy, P. J. et al. LXR, a nuclear receptor that defines a distinct retinoid response pathway. *Genes Dev.* **9**, 1033–1045 (1995).
26. Kovács, I. A. et al. Network-based prediction of protein interactions. *Nat. Commun.* **10**, 1240 (2019).
27. Baryshnikova, A. Systematic functional annotation and visualization of biological networks. *Cell Syst.* **2**, 412–421 (2016).
28. Graham, D. B. et al. TMEM258 is a component of the oligosaccharyltransferase complex controlling ER stress and intestinal inflammation. *Cell Rep.* **17**, 2955–2965 (2016).
29. Yamamoto, Y., Yoshida, A., Miyazaki, N., Iwasaki, K. & Sakisaka, T. Arl6IP1 has the ability to shape the mammalian ER membrane in a reticulon-like fashion. *Biochem. J.* **458**, 69–79 (2014).
30. Abdel-Salam, G. M. H. et al. A homozygous IER3IP1 mutation causes microcephaly with simplified gyral pattern, epilepsy, and permanent neonatal diabetes syndrome (MEDS). *Am. J. Med. Genet. A.* **158A**, 2788–2796 (2012).
31. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
32. Capra, J. A., Williams, A. G. & Pollard, K. S. ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLOS Comput. Biol.* **8**, e1002567 (2012).
33. Pan, J. et al. Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. *Cell Syst.* **6**, 555–568 (2018).
34. Yu, H. et al. High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
35. Kim, D. K. et al. EVpedia: a community web portal for extracellular vesicles research. *Bioinformatics* **31**, 933–939 (2015).
36. Hessvik, N. P. & Llorente, A. Current knowledge on exosome biogenesis and release. *Cell. Mol. Life Sci.* **75**, 193–208 (2018).
37. Imjeti, N. S. et al. Syntenin mediates SRC function in exosomal cell-to-cell communication. *Proc. Natl Acad. Sci. USA* **114**, 12495–12500 (2017).
38. Calderone, A., Castagnoli, L. & Cesareni, G. mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods* **10**, 690–691 (2013).
39. Kiran, M. & Nagarajaram, H. A. Global versus local hubs in human protein-protein interaction network. *J. Proteome Res.* **12**, 5436–5446 (2013).
40. Yang, L. et al. Comparative analysis of housekeeping and tissue-selective genes in human based on network topologies and biological properties. *Mol. Genet. Genomics* **291**, 1227–1241 (2016).
41. Paulson, J. N. et al. Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC Bioinformatics* **18**, 437 (2017).
42. Bossi, A. & Lehner, B. Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.* **5**, 260 (2009).
43. Barshir, R., Shwartz, O., Smoly, I. Y. & Yeger-Lotem, E. Comparative analysis of human tissue interactomes reveals factors leading to tissue-specific manifestation of hereditary diseases. *PLOS Comput. Biol.* **10**, e1003632 (2014).
44. Sahni, N. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).
45. Reynolds, J. J., Walker, A. K., Gilmore, E. C., Walsh, C. A. & Caldecott, K. W. Impact of PNKP mutations associated with microcephaly, seizures and developmental delay on enzyme activity and DNA strand break repair. *Nucleic Acids Res.* **40**, 6608–6619 (2012).
46. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46** (D1), D1062–D1067 (2018).
47. Bhatnagar, S. et al. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116–120 (2014).
48. Olivé, M. et al. New cardiac and skeletal protein aggregate myopathy associated with combined MuRF1 and MuRF3 mutations. *Hum. Mol. Genet.* **24**, 3638–3650 (2015).
49. Novarino, G. et al. Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* **343**, 506–511 (2014).
50. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).

[1]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA. [2]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [3]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [4]The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. [5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [6]Lunenfeld-Tanenbaum Research Institute (LTRI), Sinai Health System, Toronto, Ontario, Canada. [7]Cancer Research Center (CiC-IBMCC, CSIC/USAL), Consejo Superior de Investigaciones Científicas (CSIC) and University of Salamanca (USAL), Salamanca, Spain. [8]Institute for Biomedical Research of Salamanca (IBSAL), Salamanca, Spain. [9]The Research Institute of the McGill University Health Centre (RI-MUHC), Montreal, Quebec, Canada. [10]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada. [11]Molecular Biology of Diseases,

Groupe Interdisciplinaire de Génomique Appliquée (GIGA) and Laboratory of Viral Interactomes, University of Liège, Liège, Belgium. [12]Network Science Institute, Northeastern University, Boston, MA, USA. [13]Wigner Research Centre for Physics, Institute for Solid State Physics and Optics, Budapest, Hungary. [14]Center for Medical Biotechnology, Vlaams Instituut voor Biotechnologie (VIB), Ghent, Belgium. [15]Cytokine Receptor Laboratory (CRL), Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium. [16]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. [17]seqWell, Beverly, MA, USA. [18]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain. [19]Department of Clinical Biochemistry and Pharmacology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [20]National Institute for Biotechnology in the Negev (NIBN), Ben-Gurion University of the Negev, Beer-Sheva, Israel. [21]Cancer Research UK (CRUK) Cambridge Institute, University of Cambridge, Cambridge, UK. [22]Department of Pharmaceutical Sciences, University of Perugia, Perugia, Italy. [23]Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche "Togo Rosati" (IZSUM), Perugia, Italy. [24]Département de Virologie, Unité de Génétique Moléculaire des Virus à ARN (GMVR), Institut Pasteur, UMR3569, Centre National de la Recherche Scientifique (CNRS), Paris, France. [25]Université Paris Diderot, Paris, France. [26]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [27]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA. [28]Department of Biostatistics, Product Development, Genentech Inc., South San Francisco, CA, USA. [29]Department of Bioengineering, McGill University, Montreal, Quebec, Canada. [30]Department of Biological Sciences, Wright State University, Dayton, OH, USA. [31]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. [32]Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario, Canada. [33]These authors contributed equally: Katja Luck, Dae-Kyum Kim, Luke Lambourne, Kerstin Spirohn. ✉e-mail: david_hill@dfci.harvard.edu; marc_vidal@dfci.harvard.edu; fritz.roth@utoronto.ca; michael_calderwood@dfci.harvard.edu

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

HuRI, Lit-BM and all previously published human interactome maps from CCSB are available at http://interactome-atlas.org. The PPI data from this publication are also available through IntAct (https://www.ebi.ac.uk/intact/) with the identifier IM-25472. All HuRI-related networks from this study are available at NDExbio.org (https://tinyurl.com/networks-HuRI-paper). The raw and analysed proteomic data have been deposited in the PRIDE repository (https://www.ebi.ac.uk/pride/) with the accession number PXD012321.

Code availability

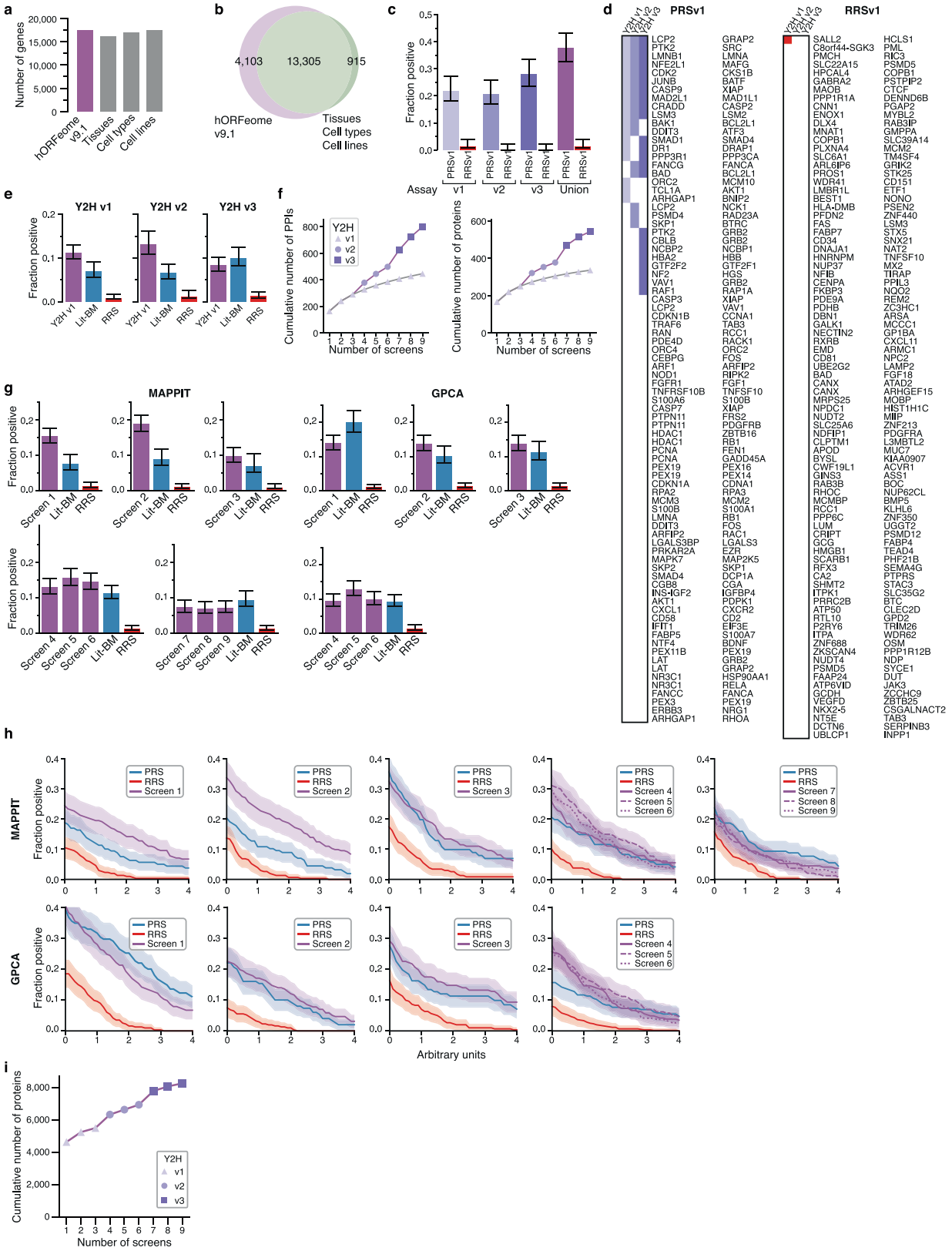
Analysis code is available at github.com/CCSB-DFCI/HuRI_paper.

Acknowledgements The authors gratefully acknowledge, in memoriam, support and insight from D. Allinger. We thank P. Porras Millan and the IntAct team for their help in disseminating our PPI data via IntAct, before and after publication. We thank U. Braunschweig, J. Ellis and B. J. Blencowe for help with data analysis. We also thank Q. Zhu, O. G. Troyanskaya, J. Pan and C. Kadoch for sharing co-expression and co-fitness data, respectively. We thank K. S. Tuttle for help with graphics. This work was primarily supported by the National Institutes of Health (NIH) National Human Genome Research Institute (NHGRI) grant U41HG001715 (M.V., F.P.R., D.E.H., M.A.C., G.D.B. and J.T.) with additional support from NIH grants P50HG004233 (M.V. and F.P.R.), U01HL098166 (M.V.), U01HG007690 (M.V.), R01GM109199 (M.A.C.), Canadian Institute for Health Research (CIHR) Foundation Grants (F.P.R. and J. Rak), the Canada Excellence Research Chairs Program (F.P.R.) and an American Heart Association grant 15CVGPS23430000 (M.V.). D.-K.K. was supported by a Banting Postdoctoral Fellowship through the Natural Sciences and Engineering Research Council (NSERC) of Canada and by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education (2017R1A6A3A03004385). C. Pons was supported by a Ramon Cajal fellowship (RYC-2017-22959). G.M.S. was supported by NIH Training Grant T32CA009361. M.V. is a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS, Wallonia-Brussels Federation, Belgium).

Author contributions The project was conceived and supervised by G.D.B., J.T., D.E.H., M.V., F.P.R. and M.A.C. The Y2H assay versions were developed and benchmarked by K.S., A.D.R. and Q.Z. with help from K.L., B.E.B. and D.B. hORFeome v.9.1 was generated by K.L., D.-K.K., K.S., W.B., M.D., D.B., D.M. and T.H. with help from A.G.C., A. Dricot, A.M., S.R., Y.S., G.M.S., J.-C.T. and X.Y. The preparation of Y2H destination clones by en masse gateway cloning and yeast transformations were performed by D.-K.K., K.S., A.G.C., J.J.K., R.L., D.M. with help from M.G., D.S., S.S., B.T., C.C., G.H., N.v.L., A.R. and J.W. The Y2H screens were performed and the data analysed by K.L., K.S., B.E.B., M.D., A. Dricot, M.F.H., C. Pollis, S.S., B.T., A.T. and T.H. with help from W.B., T.C., B.C., A. Desbuleux, D.B., S.-F.C., A.M., D.M., J. Rasla, A.S.-M., Y.S. and Y.W. The validation experiments were performed and the data analysed by T.C., A. Desbuleux, I.L., S.G.C. and T.H. with help from K.L., L.L., K.S., D.B., S.D.R., Y.J., Y.K., S.R. and J.T. Sequencing and analysis of the sequencing data was performed by W.B., A.G.C., M.G., N.K., J.J.K., J.C.M., Y.S. and T.H. with help from K.L., D.-K.K., M.B., C.C., A.G., R.L., A.R., M.T. and J.W. Integrative downstream analyses were performed by K.L., D.-K.K., L.L., F.J.C.-L., I.A.K. and C. Pons with help from B.C., O.B., G.C., D.D.R., M.D.-F., F.G., G.H., J.N.P., T.R., E.S., E.Y.-L., Y.X., P.A. and J.D.L.R. Follow-up experiments and analyses were performed by K.L., D.-K.K., K.S., R.B., D.C., S.D., A. Desbuleux and A.Y. with help from L.L., T.C., C.B.-C., G.C., C.D.A., H.E., L.G., E.H., S.L. and R.J.W. supervised by S.G., J. Rak and V.T. The web portal was developed by M.W.M. with help from K.L., M.H., T.H., M.A.C. and G.D.B. The paper was written by K.L., D.-K.K., L.L., K.S., D.E.H., M.V., F.P.R. and M.A.C. with help from F.J.C.-L., A.G.C., G.C., S.G., I.A.K., T.H. and A.Y. Authors other than co-first and co-corresponding are listed alphabetically and contributed equally within their group.

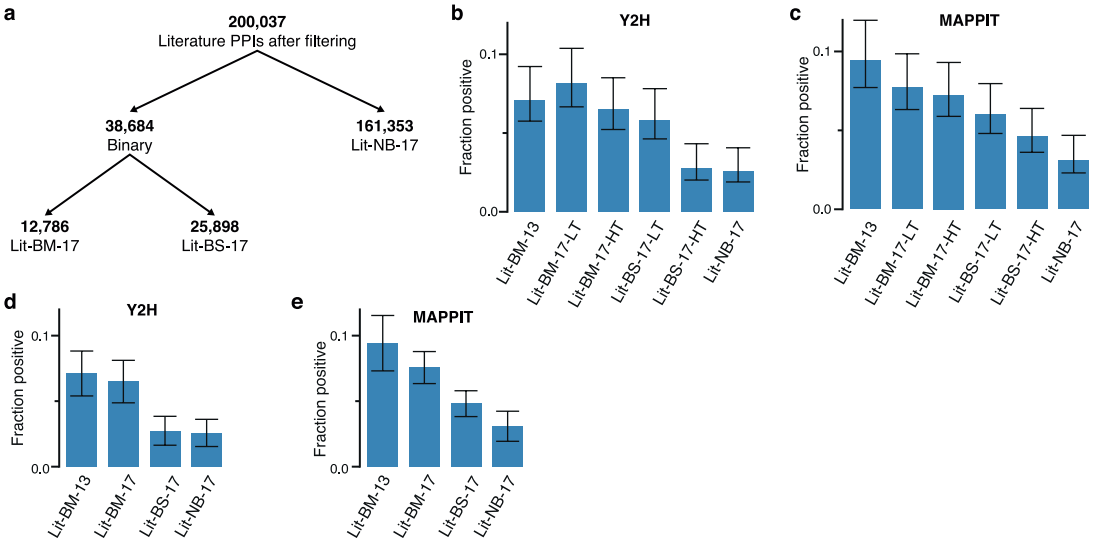
Competing interests J.C.M. is a founder and CEO of seqWell, Inc; F.P.R. and M.V. are shareholders and scientific advisors of seqWell, Inc.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2188-x.
Correspondence and requests for materials should be addressed to D.E.H., M.V., F.P.R. or M.A.C.
Peer review information *Nature* thanks Ulrich Stelzl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

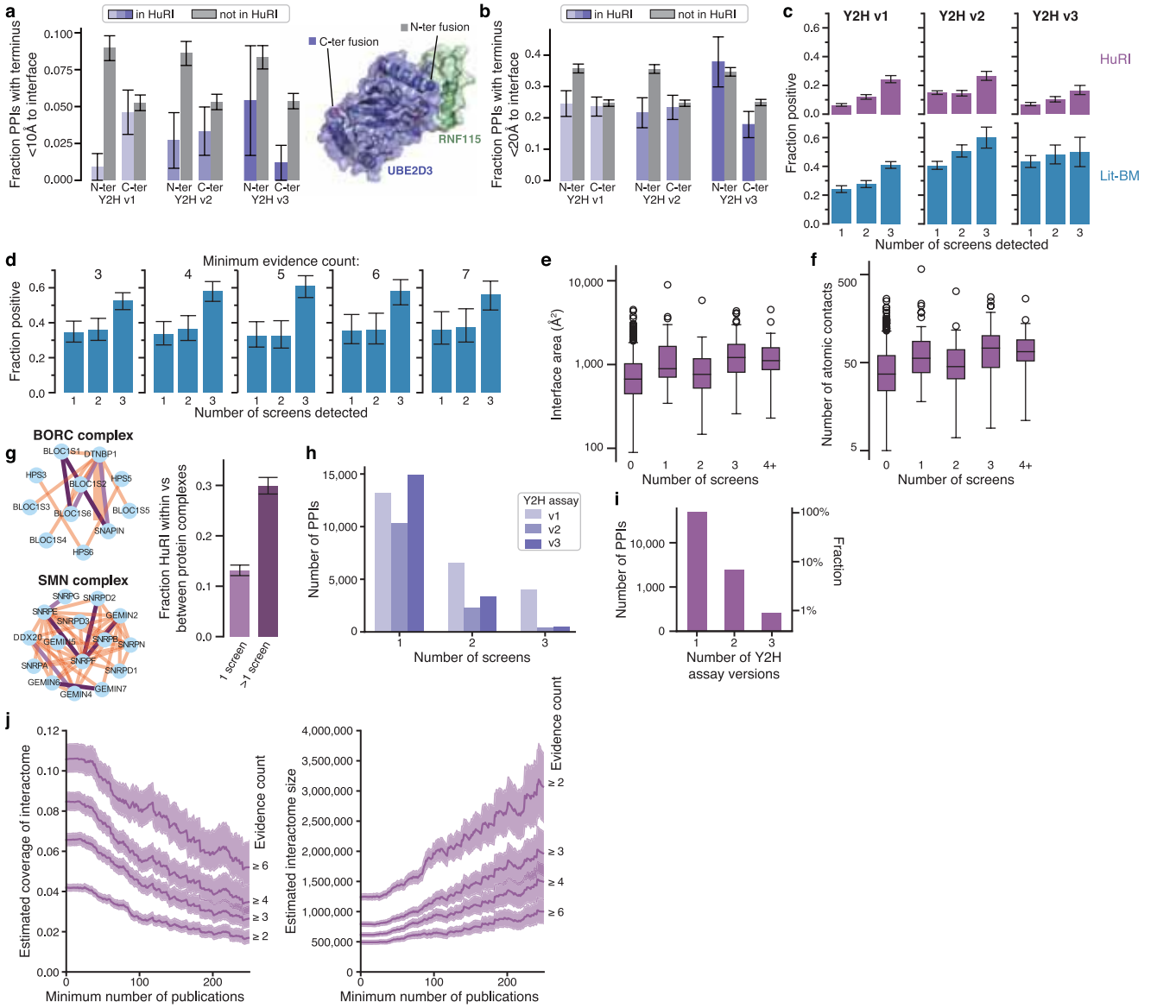


Extended Data Fig. 1 | See next page for caption.

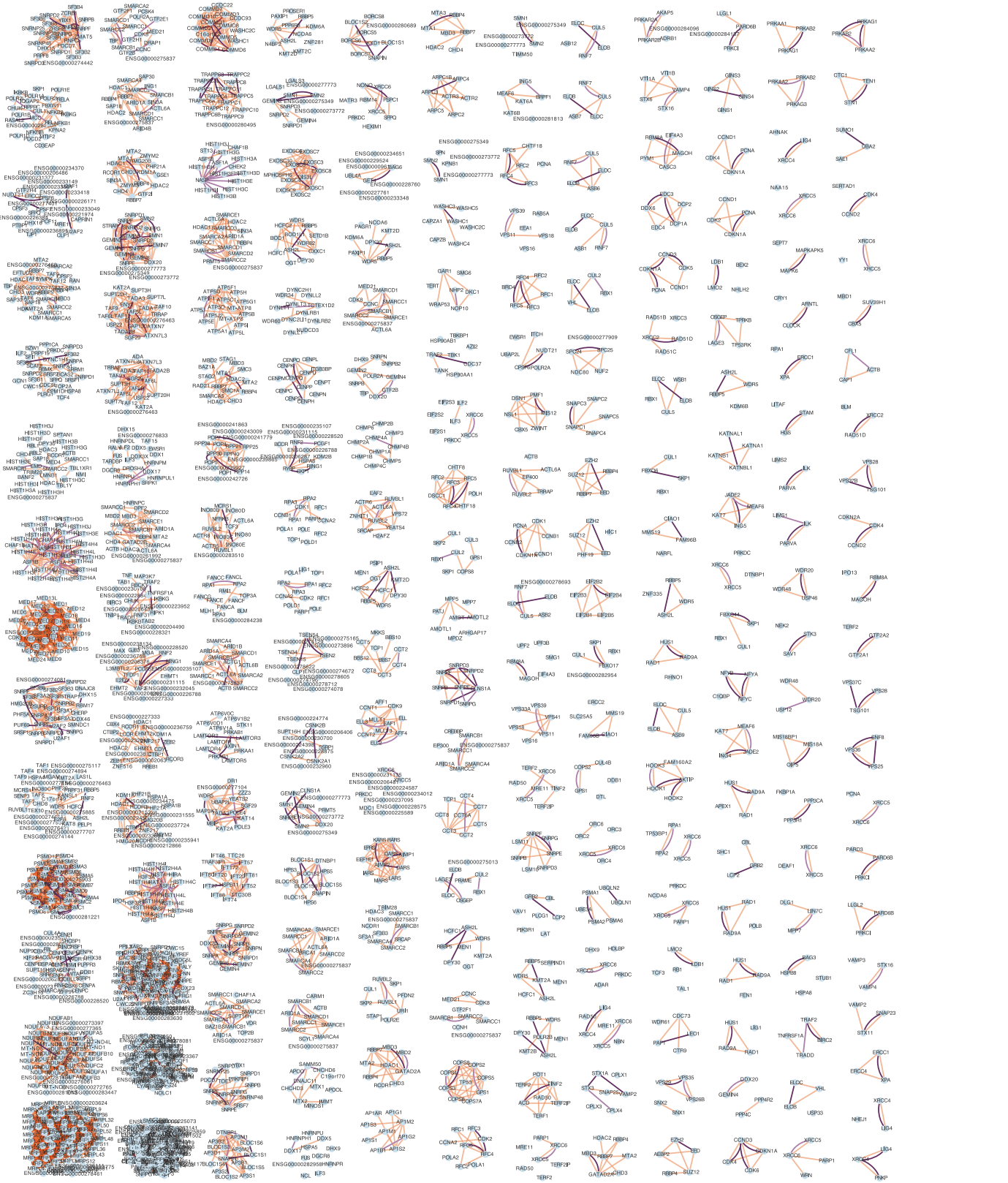
Extended Data Fig. 1 | Y2H assay development and validation of HuRI. **a**, Number of protein-coding genes in hORFeome v9.1 and GTEx (tissues), FANTOM (cell types) and HPA (cell lines) transcriptome projects. The number of genes in hORFeome v9.1 is on par with the number of genes expressed in three comprehensive individual transcriptome sequencing studies and includes 94% of the genes with robust evidence of expression in all three. **b**, Overlap between hORFeome v9.1 and intersection of transcriptomes in **a**. **c**, Individual and combined recovery of PRSv1 and RRSv1 pairs by Y2H assay versions ($n = 252, 270$). **d**, Coloured squares showing which protein pairs were detected in PRSv1 (left) and RRSv1 (right) by Y2H assay versions. **e**, Recovery rates of Lit-BM and PPIs from screens of a 2,000-by-2,000 gene test space per Y2H assay version in MAPPIT. **f**, Cumulative PPI count performing three screens with each Y2H assay version in the test space compared to nine screens with Y2H assay version 1. **g**, **h**, MAPPIT and GPCA recovery of Lit-BM and PPIs from screens of Space III when split by screen at an RRS rate of 1% (**g**) or across a range of thresholds (**h**). All error bars in **c**, **e** and **g**, are 68.3% Bayesian confidence interval; shaded error band in **h** is standard error of proportion and n = between 101 and 395 pairs successfully tested for each category. **i**, Number of proteins in HuRI, detected with each additional screen.



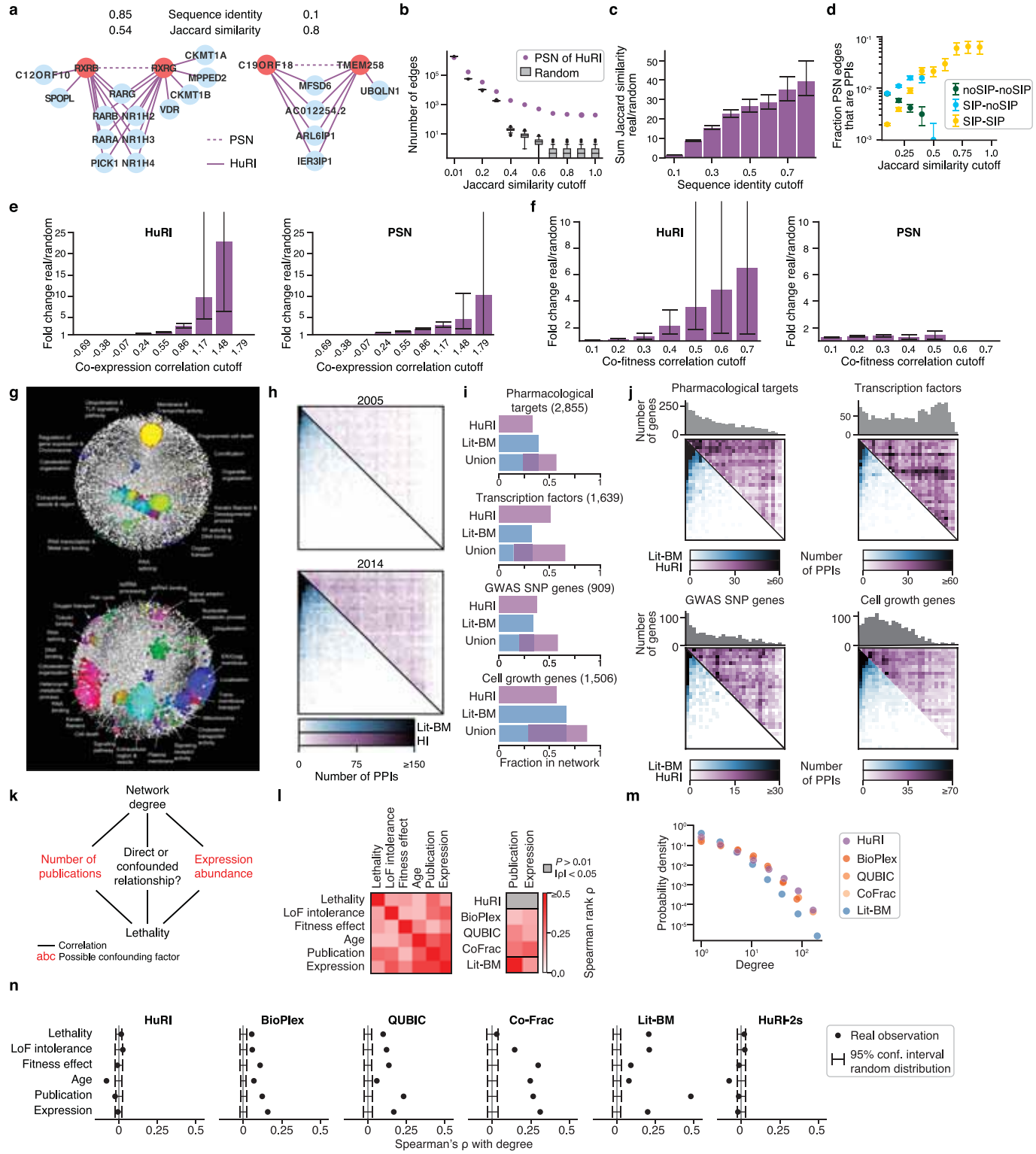
Extended Data Fig. 2 | Definition of literature-curated PPI datasets. **a**, Categorization of literature-curated PPIs into distinct subsets based on the experimental methods in which they were detected and the number of pieces of experimental evidence. **b**–**e**, Results of testing the different categories of literature-curated pairs in Y2H (**b**, **d**) and MAPPIT (**c**, **e**) in which the pairs have been further divided into high-throughput (HT) and low-throughput (LT) subsets (**b**, **c**). There were between $n = 191$ and $n = 471$ successfully tested PPIs for each category. BM, binary multiple; BS, binary singleton; NB, non-binary. Error bars are standard error of proportion.



Extended Data Fig. 3 | Stericity and interaction strength contribute to PPI detectability. a, **b**, Fraction of PPIs with N or C terminus <10 Å (**a**) or <20 Å (**b**) to PPI interface, for PPIs with known structure in and not in HuRI (n = 37–1,891 PPIs). Error bars are standard error of proportion. The structure of UBE2D3 bound to RNF115 illustrates an example of a PPI found only by Y2H assay version 3 (PDB code 5ULH). **c**, MAPPIT recovery rates of HuRI and Lit-BM PPIs that were also detected in HuRI by the number of screens each pair was detected in. Error bars are 68.3% Bayesian confidence interval (n = 22–793 PPIs successfully tested in each category). **d**, MAPPIT recovery rates of Lit-BM PPIs that were also detected in HuRI, for increasing number of pieces of experimental evidence per PPI. Error bars are 68.3% Bayesian confidence interval (n = 24–61 PPIs successfully tested in each category). **e**, **f**, Distributions of interaction interface area (**e**) or number of atomic contacts (**f**) by the number of HuRI screens in which a PPI is detected, with box plots showing median, interquartile range (IQR), and 1.5× IQR (with outliers); n = 1,004 PPIs. **g**, Left, examples of within-complex interactions detected in HuRI (purple) and BioPlex (orange). Right, fraction of HuRI PPIs between proteins of protein complexes that link proteins of the same complex, split by PPIs found in single and multiple screens (dark purple). Error bars are standard error of proportion; n = 1,042 and 775 PPIs, for single and multiple screens, respectively. **h**, Number of screens each PPI in HuRI was detected in, split by Y2H assay version. **i**, Number of Y2H assay versions each PPI in HuRI was detected in. **j**, Estimates of the size of the total binary protein interactome and the fraction covered by HuRI, right and left, respectively, as a function of the minimum number of publications per gene and the minimum number of evidence for the Lit-BM reference. Error bands are 68.3% Bayesian confidence interval; $n \geq 170$ Lit-BM PPIs.

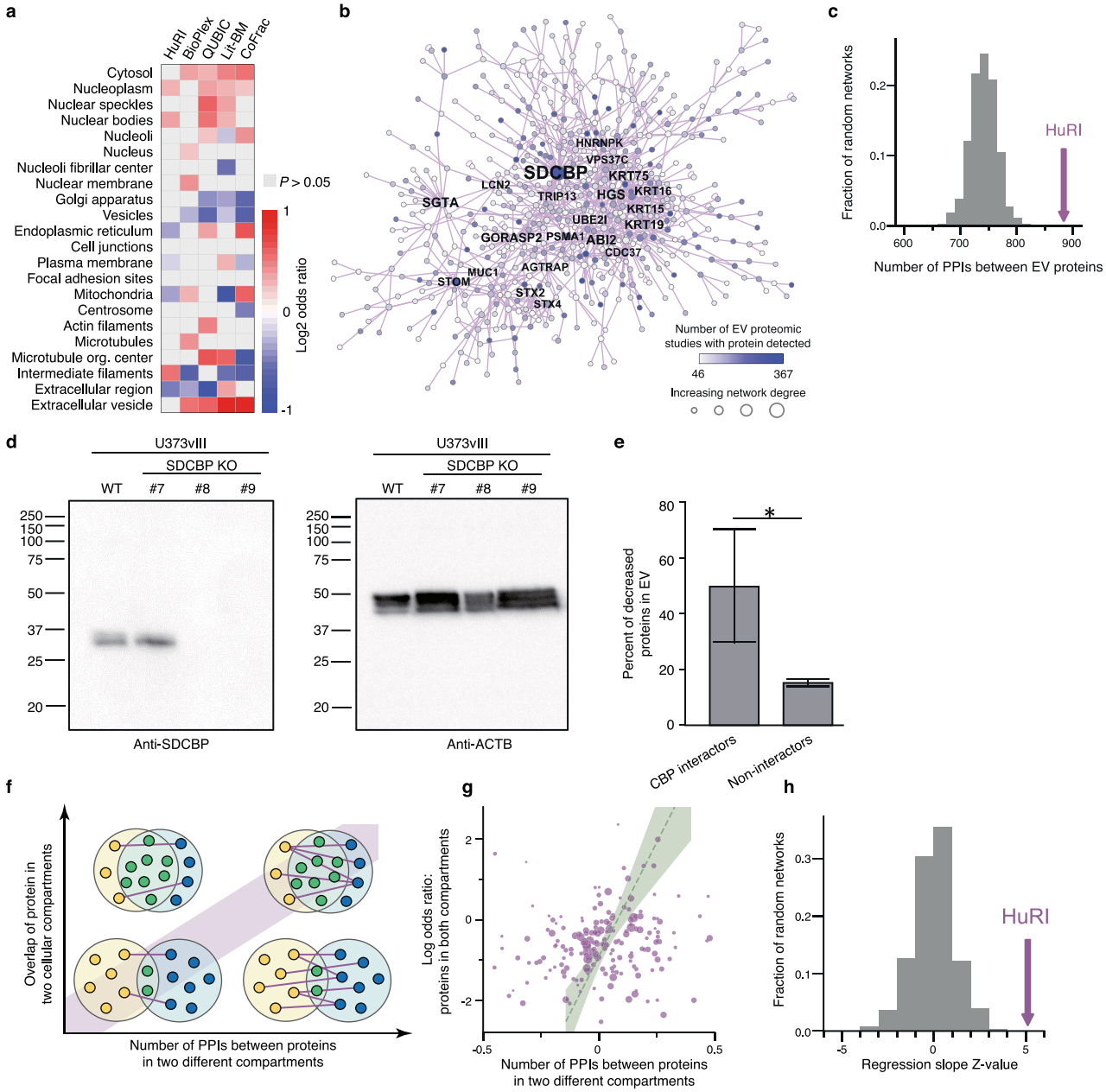


Extended Data Fig. 4 | HuRI provides direct contact information for proteins in complexes. Intra-complex PPIs are shown for protein complexes from CORUM as found in BioPlex (orange) or HuRI (purple). HuRI PPIs are further distinguished into PPIs found in single (light purple) or multiple screens (dark purple).

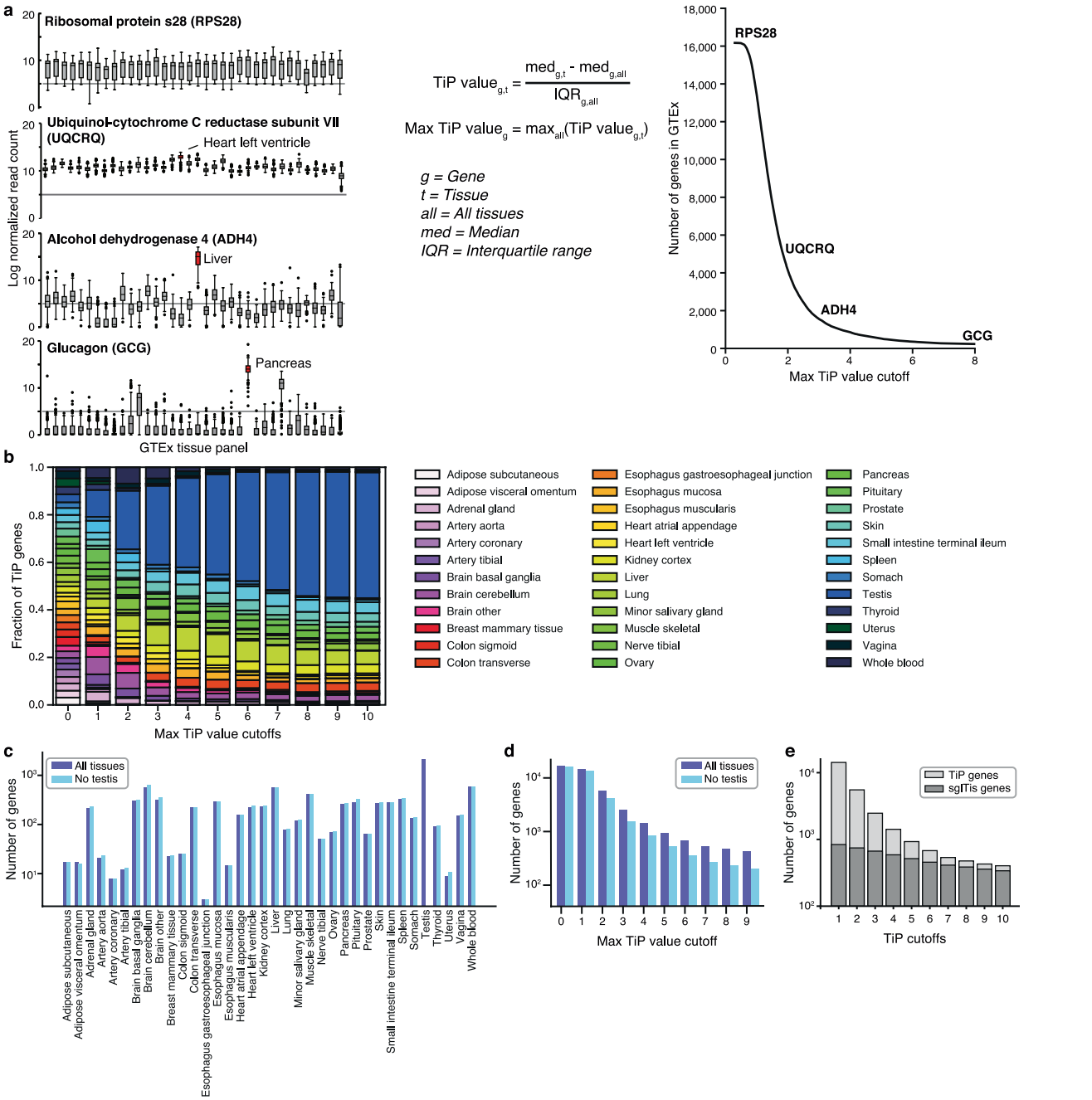


Extended Data Fig. 5 | See next page for caption.

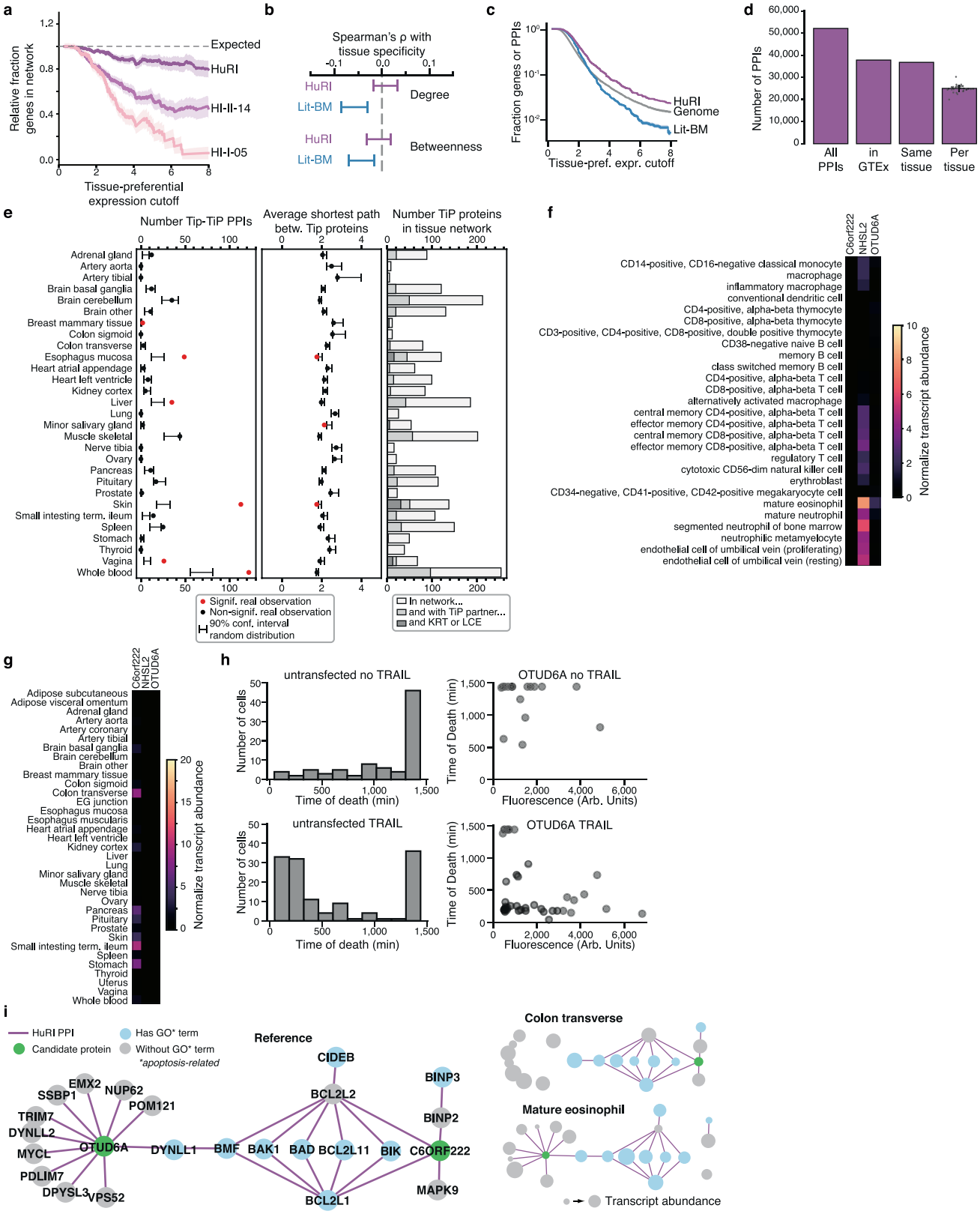
Extended Data Fig. 5 | Topological and functional significance of HuRI. **a**, Examples of protein pairs in HuRI with high interaction profile similarity and both high (left) and low (right) sequence identity. **b**, The number of pairs of proteins in HuRI and 100 random networks at increasing Jaccard similarity cutoffs. Box plots are as in Extended Data Fig. 3e. **c**, Enrichment over random networks of the sum of Jaccard similarities of pairs of proteins in HuRI at increasing thresholds of sequence identity. Error bars are 95% confidence intervals, centre is relative to mean of random networks. **d**, Fraction of PSN edges that are also PPIs in HuRI, split by the PPIs involving no, one or two self-interacting proteins (SIPs), at increasing Jaccard similarity cutoffs. Error bars are standard error of proportion. **e**, **f**, Enrichment over random networks of the PPI count (left) or sum of Jaccard similarities (right) of HuRI PPIs or PSN pairs, respectively, at increasing co-expression (**e**) and co-fitness (**f**) cutoffs. Error bars are 95% confidence interval, centre is relative to mean of random networks. **g**, Functional modules in HuRI (top) and its PSN (bottom) with functional annotations. **h**, Heat maps of PPI counts, ordered by number of publications, for our previous human interactome maps and Lit-BM **i**, Fraction of genes with at least one PPI for biomedically interesting genes. **j**, Heat maps of HuRI and Lit-BM PPI counts between proteins, ordered by number of publications, restricted to PPIs involving genes from the corresponding gene set. **k**, Schematic of relation between variables: observed PPI degree, abundance, number of publications, and lethality. **l**, Correlation matrices. PPI datasets refer to their network degree. **m**, Degree distribution of various PPI networks. **n**, Empirical determination of significance of correlation between various network degrees and gene properties. HuRI-2s, subset of HuRI found in at least two screens. $n = 13{,}441$–$53{,}704$ PPIs per network.



Extended Data Fig. 6 | Co-localization of proteins interacting in HuRI. **a**, Odds ratios of proteins in different subcellular compartments and PPI datasets. $n = 125–3,941$ proteins per compartment, two-tailed Fisher's exact test. **b**, The subnetwork of HuRI involving extracellular vesicle proteins. Names of high-degree proteins are shown. **c**, Number of PPIs in HuRI between extracellular vesicle proteins (purple arrow) compared to the distribution from randomized networks (grey). **d**, Western blots of SDCBP (left) and ACTB (loading control, right) in wild-type and three knockout (KO) cell lines (#7–#9), repeated twice in two independent laboratories. Full scanned image was displayed, obtained by ChemiDoc MP imager (Bio-Rad). Cell line #8 was used for extracellular vesicle proteomics. **e**, Fraction of proteins in which abundance in extracellular vesicles was significantly reduced in the SDCBP-knockout cell line, split by proteins interacting and not interacting with SDCBP as identified in HuRI. Error bars are standard error of proportion ($n = 6$ interactors, 638 non-interactors, $^*P = 0.042$, one-tailed empirical test). **f**, Schematic illustrating that the number of HuRI PPIs between proteins from two different compartments should correlate with the enrichment of both compartment pairs to overlap, if co-localization annotation is incomplete. **g**, Scatter plot showing, for each pair of subcellular compartments, odds ratios quantifying the enrichment for proteins located in both compartments versus the enrichment of the density of PPIs between proteins located to either compartment. Size of points is scaled by the standard error of the *x* axis variable. Regression line and 95% confidence interval are shown. **h**, The z-score of the regression slope of **g** compared to those of random networks.

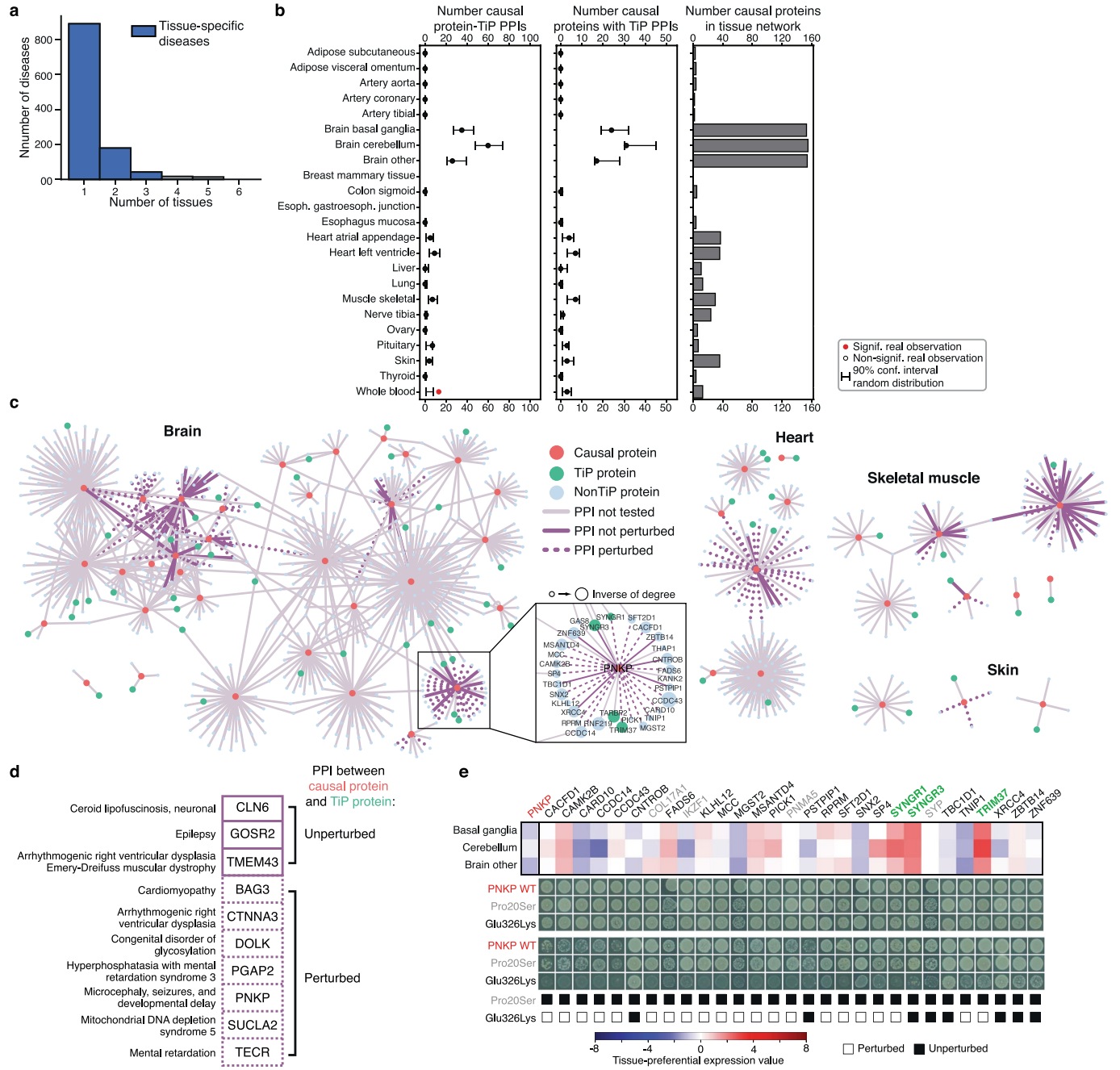


Extended Data Fig. 7 | Investigation of tissue-preferential expression data. **a**, Examples of genes displaying different levels of TiP gene expression across the GTEx tissue panel (left). Box plots are as in Extended Data Fig. 3e. $n = 90–779$ samples per tissue. Equation to calculate tissue-preferential expression for every gene–tissue pair and the maximum TiP value for every gene (middle). Number of genes showing tissue-preferential expression for increasing tissue-preferential expression cutoffs (right). **b**, Relative number of TiP genes for every tissue for increasing tissue-preferential expression cutoffs. **c**, **d**, Differences in number of TiP genes after removal of testis before TiP value calculation per tissue (TiP value cutoff = 2) (**c**) and in total for increasing tissue-preferential expression cutoffs (**d**). **e**, Number of TiP genes and number of TiP genes that are also exclusively expressed in one tissue (for increasing tissue-preferential expression cutoffs. sglTis, single tissue.

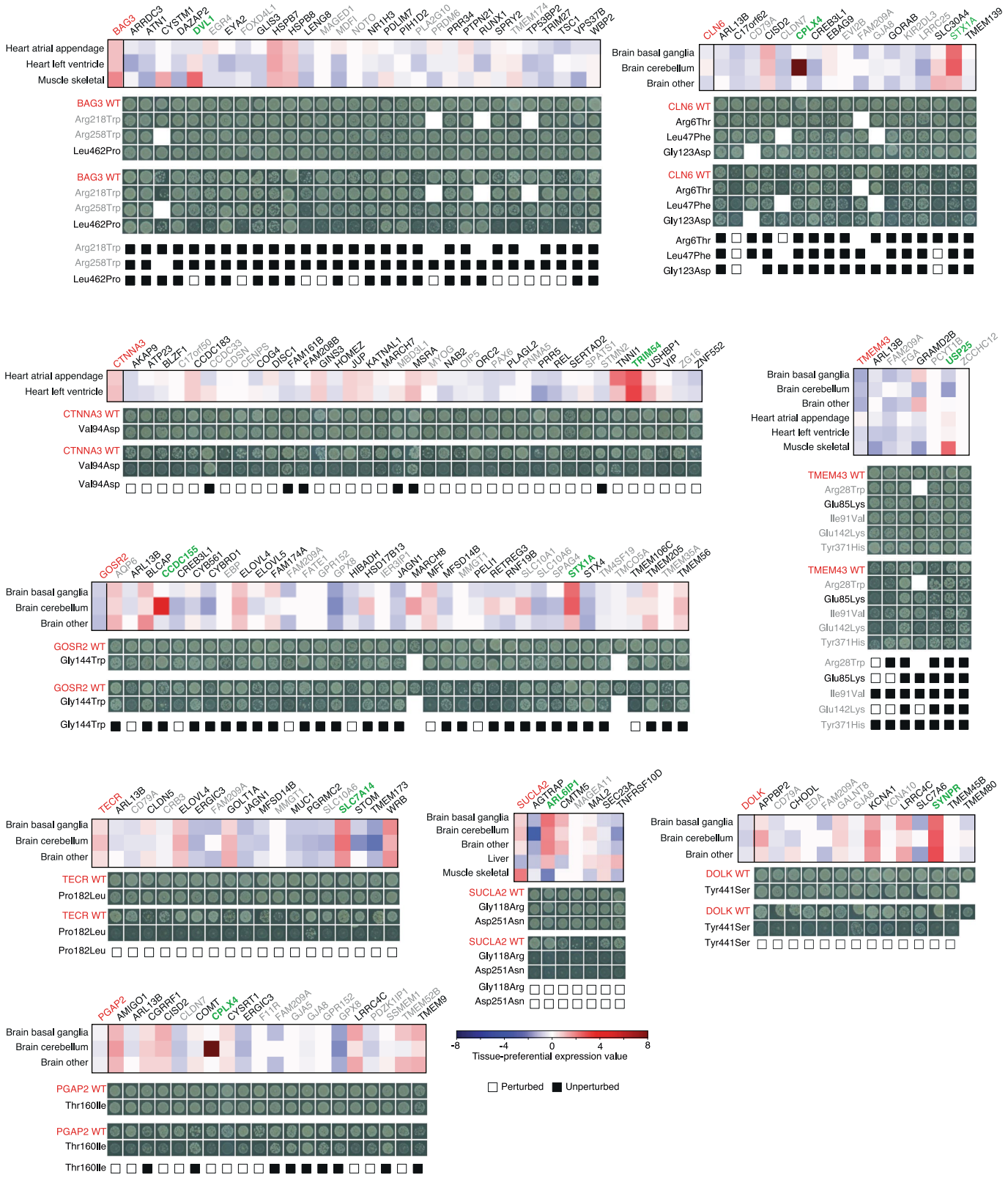


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | PPIs between TiP proteins and uniformly expressed proteins likely adapt basic cellular processes to mediate cellular context-specific functions. a, TiP protein coverage by CCSB PPI networks for increasing levels of tissue-preferential expression. Shaded error bars are proportional to standard error of proportion, $n \geq 233$ genes. **b**, Spearman correlation coefficients and 95% confidence intervals for correlations between degree or betweenness and tissue specificity for HuRI and Lit-BM ($n = 6,684$ and 4,971 proteins). **c**, Fraction of HuRI and Lit-BM that involve TiP proteins compared to fraction of genome that are TiP genes for increasing levels of tissue-preferential expression. **d**, Number of PPIs in HuRI, involving proteins in GTEx, in which both proteins are expressed in the same tissue, and the mean of the tissue-specific subnetworks. Error bar denotes s.d. **e**, Test for enrichment of TiP–TiP PPIs (left) and significance of average shortest path between TiP proteins (middle) in each tissue subnetwork, number of TiP proteins in each subnetwork, interacting with other TiP proteins, being part of keratin (KRT) or late-cornified envelope (LCE) protein family (right). **f**, **g**, Transcript expression levels across the BLUEPRINT haematopoietic cell lineage (**f**) and GTEx tissue panel (**g**) for three candidate genes predicted to function in apoptosis. EG, oesophagus gastroesophageal. **h**, Histogram of number of untransfected cells and their time of death (left) without (top) and with (bottom) addition of TRAIL. Time of death of cells expressing OTUD6A–GFP fusions versus OTUD6A expression measured as fluorescence (right) without (top) and with (bottom) addition of TRAIL. **i**, Apoptosis-related network context of OTUD6A and C6ORF222 in HuRI, unfiltered (left) and filtered using colon transverse or mature eosinophil transcript levels (right).



Extended Data Fig. 9 | Potential mechanisms of tissue-specific diseases. **a**, Histogram of the number of Mendelian diseases showing symptoms in several tissues. **b**, Test for enrichment of causal proteins associated with tissue-specific Mendelian diseases to interact with TiP proteins of affected tissues. **c**, Network neighbourhood of uniformly expressed causal proteins of tissue-specific diseases found to interact with TiP proteins in HuRI, indicating PPI perturbation by mutations. **d**, Causal genes split by mutation found to perturb PPI to TiP protein (dashed) or not (solid). **e**, Expression profile of PNKP and interactors in brain tissues and PPI perturbation pattern of disease causing (Glu326Lys) and benign (Pro20Ser) mutation. Yeast growth phenotypes on SC-Leu-Trp (top) or SC-Leu-Trp-His+3AT media (bottom) are shown; green or grey protein symbols denote preferentially or not expressed, respectively.



Extended Data Fig. 10 | Mutations in uniformly expressed causal proteins associated with tissue-specific Mendelian diseases perturb interactions to TiP proteins. Expression profile and interaction perturbation profile of nine causal proteins and their interaction partners. Top, affected tissues were selected for display. Middle, control of activation domain and Gal4 DNA-binding domain plasmid presence and cell density by spotting yeast colonies on SC-Leu-Trp media. Bottom, detection of PPIs by spotting yeast on SC-Leu-Trp-His+3AT media, in which yeast growth indicates PPIs. Red letters denote causal proteins; grey protein symbols denote interaction partners not expressed in affected tissues; black and grey alleles denote pathogenic and not pathogenic, respectively; green protein symbols denote TiP interaction partners in affected tissues.

nature research

Corresponding author(s): Michael A. Calderwood, Frederick P. Roth, Marc Vidal, David E. Hill

Last updated by author(s): Dec 6, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Bowtie2 2.2.3 samtools 1.2 Illumina bcl2fastq 2.20 BLAT 3.6.x1 Python 2.7.10 MySQL 5.6.26 Custom code for Y2H screening pipeline Maxquant 1.5.6.5 NTA 3.0 Colonyzer 2
Data analysis	YARN package downloaded on March 15th 2016 ImageJ 1.52d BLAST 2.2.30 MUSCLE 3.8.31 InterProScan 5.16-55.0 Cytoscape 3.4.0 Microsoft Excel 2016 GraphPad Prism 7 Matlab 2016a SAFE 1.5 FuncAssociate 3 Python 3.7.2

numpy 1.15.4
scipy 1.1.0
scikit-image 0.14.2
Orange-Bioinformatics 2.6.25
pandas 0.24.1
matplotlib 3.0.2
igraph 0.7.1
matplotlib-venn 0.11.5
py2cytoscape 0.7.0
R 3.5.1
downloader 0.4
Biobase 2.40.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The PPI data from this publication has been submitted to the IMEx (http://www.imexconsortium.org) consortium through IntAct and assigned the identifier IM-25472. HuRI, Lit-BM, and all previously published human interactome maps from CCSB are available at http://interactome.dfci.harvard.edu/huri/. All HuRI-related networks generated and analyzed in this study are available at NDExbio.org (https://tinyurl.com/networks-HuRI-paper). The raw and analyzed proteomic data were deposited in the PRIDE repository with the accession number PXD012321.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	1. A sample size of ~250 pairs for categories in PPI validation assay experiments was chosen. This is based on a 95% CI value of 5% inputting an expected recovery rate of 20%, from previous similar experiments, using the standard error of proportion. 2. The test space for the pilot screens was screening a random 10% of the ORFeome against a different 10% amounting to 1% of the total space. Based on previous experience with interaction density of screening, this was estimated to result in roughly 100 pairs per screen, enough to observe differences between assay versions and the saturation effects of a few percent after several screens. 3. For comparative proteomics of extracellular vesicle (EV; in Fig. 4), we produced three biological replicates (at EV purification step). 4. For cell death assay (in Extended data fig. 8h), we preformed two replicates and obtained data for 40 cells in total which are displayed together as single data points on the scatter plot. Based on our previous experience with live-cell imaging of cell death dynamics, we anticipated that imaging n = 20 to 50 cells would allow us to effectively characterize the dynamics in each sample.
Data exclusions	No data were excluded.
Replication	1. For comparative proteomics, this gives us enough reproducibility, showing that we have significantly higher correlation between replicates than that between samples (before & after SDCBP knock-out). In addition, knock-out of SDCBP was confirmed by this proteomics in all three replicates and further confirmed by Wester blot. 2. For the cell death assay (Extended Figure 8h), the data presented collates data from two independent replicates of the experiments that showed the same trends. Both experiments were carried out using the same plating and transfection conditions, using cells that were passaged fewer than 10 times after receipt from ATCC.
Randomization	Screen pairs and controls for Y2H pairwise test, MAPPIT, and GPCA were randomly arranged on well plates.
Blinding	Manual scoring of Y2H pairwise tests was blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
[ ]	[x] Antibodies
[ ]	[x] Eukaryotic cell lines
[x]	[ ] Palaeontology
[x]	[ ] Animals and other organisms
[x]	[ ] Human research participants
[x]	[ ] Clinical data

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[x]	[ ] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used

Two antibodies were used for the SDCBP KO validation

1. Rabbit anti-SDCBP antibody (ab133267 with clone number ERP8102 and lot number GR282684-7) was purchased from Abcam (Cambridge) and used with dilution of 1:2000.

2. Mouse anti-ACTB antibody (A5441 with clone name AD-15) was purchased from Sigma and used with dilution of 1:10000. Lot number is not avaliable for this antibody.

3. Goat anti-rabbit IgG (HRP-linked) antibody (7074S with lot number 28) was purchased from Cell Signaling Technology and used with dilution of 1:5000. Clone name was not provided by Cell Signaling Technology.

4. Goat anti-mouse IgG (HRP-linked) antibody (1706516 with lot number 64055911) was purchased from Bio-Rad and used with dilution of 1:5000. Clone name was not provided by Bio-Rad.

Validation

No antidbodies were validated.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

1. U373vIII: the same cell line used in the many previous studies (Al-Nedawi, et al. Nat Cell Biol. 2008 May;10(5):619-24., et al.)

2. HeLA: it was bought from ATCC (https://www.atcc.org/products/all/CCL-2.aspx)

3. HEK293T for GPCA: the same cell line used in the many previous studies (Choi, et al. Nat Commun. 2019 Aug 29;10 (1):3907.)

4. HEK293T for MAPPIT: it was obtained by our department in 1996 from Dr. Mark Hall at the Biochemistry department of the University of Birmingham.

5. Yeast strains - Y8800 and Y8930: the same strain used in the many previous studies (Rolland, et al. Cell. 2014 Nov 20;159 (5):1212-1226.)

Authentication

All the cell lines except HEK293T for MAPPIT were not re-authenticated.
HEK293T for MAPPIT was authenticated by analyzing single nucleotide polymorphism (Lin, et al. Nat Commun. 2014 Sep 3;5:4767.)

Mycoplasma contamination

All the cell lines were tested with DNA stain for mycoplasma contamination and free of contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used.