



HAL
open science

A quantitative mapping approach to identify direct interactions within complexomes

Philipp Trepte, Christopher Secker, Soon Gang Choi, Julien Olivet, Eduardo Silva Ramos, Patricia Cassonnet, Sabrina Golusik, Martina Zenkner, Stephanie Beetz, Marcel Sperling, et al.

► **To cite this version:**

Philipp Trepte, Christopher Secker, Soon Gang Choi, Julien Olivet, Eduardo Silva Ramos, et al.. A quantitative mapping approach to identify direct interactions within complexomes. 2021. pasteur-03355221

HAL Id: pasteur-03355221

<https://pasteur.hal.science/pasteur-03355221>

Preprint submitted on 27 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

A quantitative mapping approach to identify direct interactions within complexomes

Philipp Trepte^{1,2,*,#}, Christopher Secker^{1,#}, Soon Gang Choi^{3,4,5,#}, Julien Olivet^{3,4,5,6,#},
Eduardo Silva Ramos¹, Patricia Cassonnet⁷, Sabrina Golusik¹, Martina Zenkner¹,
Stephanie Beetz¹, Marcel Sperling⁸, Yang Wang^{3,4,5}, Tong Hao^{3,4,5}, Kerstin Spirohn^{3,4,5},
Jean-Claude Twizere⁶, Michael A. Calderwood^{3,4,5}, David E. Hill^{3,4,5}, Yves Jacob^{3,7}, Marc
Vidal^{3,4,*}, Erich E. Wanker^{1,*}

¹Neuroproteomics, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125, Berlin, Germany.

²Brain Development and Disease, Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Dr. Bohr-Gasse 3, 1030, Vienna, Austria.

³Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute (DFCI), 450 Brookline Avenue, Boston, MA, 02215, USA.

⁴Department of Genetics, Blavatnik Institute, Harvard Medical School (HMS), 77 Avenue Louis Pasteur, Boston, MA, 02115, USA.

⁵Department of Cancer Biology, Dana-Farber Cancer Institute (DFCI), 450 Brookline Avenue, Boston, MA, 02215, USA.

⁶Laboratory of Viral Interactomes, Unit of Molecular Biology of Diseases, Interdisciplinary Cluster for Applied Genoproteomics (GIGA Institute), University of Liège, 1 Avenue de l'hôpital, 4000, Liège, Belgium.

⁷Département de Virologie, Unité de Génétique Moléculaire des Virus à ARN (GMVR), Institut Pasteur, UMR3569, Centre National de la Recherche Scientifique (CNRS), Université de Paris, 28 rue du Docteur Roux, 75015, Paris, France.

⁸Multifunctional Colloids and Coating, Fraunhofer Institute for Applied Polymer Research (IAP), Geiselbergstr. 69, 14476 Potsdam-Golm, Germany.

#These authors contributed equally: Philipp Trepte, Christopher Secker, Soon Gang Choi, Julien Olivet.

*Correspondence should be addressed to: philipp.trepte@imba.oeaw.ac.at; marc_vidal@dfci.harvard.edu; erich.w@mdc-berlin.de

32 **ABSTRACT**

33 Complementary methods are required to fully characterize all protein complexes, or the
34 complexome, of a cell. Affinity purification coupled to mass-spectrometry (AP-MS) can identify
35 the composition of complexes at proteome-scale. However, information on direct contacts
36 between subunits is often lacking. In contrast, solving the 3D structure of protein complexes
37 can provide this information, but structural biology techniques are not yet scalable for
38 systematic, proteome-wide efforts. Here, we optimally combine two orthogonal high-
39 throughput binary interaction assays, LuTHy and N2H, and demonstrate that their quantitative
40 readouts can be used to differentiate direct interactions from indirect associations within
41 multiprotein complexes. We also show that LuTHy allows accurate distance measurements
42 between proteins in live cells and apply these findings to study the impact of the polyglutamine
43 expansion mutation on the structurally unresolved N-terminal domain of Huntingtin. Thus, we
44 present a new framework based on quantitative interaction assays to complement structural
45 biology and AP-MS techniques, which should help to provide first-approximation contact maps
46 of multiprotein complexes at proteome-scale.

47

48

49 INTRODUCTION

50 Characterizing the molecular architecture of all protein complexes, or the complexome,
51 of a cell is essential to fully decipher its biological functions^{1,2}. Protein complexes vary in sizes
52 and compositions as they often consist of different types of macromolecules including DNA,
53 RNA and lipids. For example, in human cells, ~80 proteins closely assemble with different
54 ribosomal RNAs to form a functional ribosome, which is a compact spherical particle with a
55 diameter of 250 to 300 Å^{3,4}. In comparison, the human proteasome is composed of eight
56 protein subunits that co-assemble into a stable cylindrical complex⁵. This exemplifies how
57 structurally diverse proteins can dynamically assemble in cells to perform various biological
58 functions.

59 Affinity purification coupled to mass spectrometry (AP-MS) techniques are highly
60 efficient in identifying the composition of protein complexes at proteome-scale⁶⁻⁹, but their
61 ability to elucidate direct contacts between potentially interacting subunits is limited^{10,11}.
62 However, gaining information on direct interactions, i.e. on proteins sharing a physical
63 interface, and indirect associations, i.e. proteins that do not directly interact, within complexes
64 is important to understand their biological roles. For example, clearly defining interacting
65 domains between proteins will help in the design of hypothesis-driven functional studies.
66 Additionally, defining the relationship between complex members is crucial to characterize the
67 effect of disease-associated missense mutations on protein assemblies, or for the
68 identification and validation of drug targets.

69 Structural biology technologies such as X-ray diffraction (XRD), nuclear magnetic
70 resonance (NMR), or cryo-electron microscopy (cryo-EM) provide high-resolution data on the
71 assembly of multiprotein complexes. This includes information on the folding of protein
72 domains, amino acid composition at contact sites, attachment of compounds to proteins, and
73 post-translational modifications of specific amino acids. However, in comparison to high-
74 throughput AP-MS-based methods, generation of protein interaction data using structural
75 biology techniques can be labor intensive and time consuming. In addition, such techniques

76 often use truncated instead of full-length proteins, and comprehensive, high-resolution
77 structural information is still missing for most protein complexes. Indeed, based on a recent
78 study by Drew et al, ~7,000 human protein complexes can be found in the human proteome
79 from the integration of over 15,000 published mass-spectrometry experiments^{12,13}. However,
80 only ~4% of those (i.e. 309 protein complexes) currently have a resolved structure in the
81 literature. Thus, complementary techniques are required to precisely characterize protein
82 complexes currently identified by AP-MS-based methods^{7,11,14}.

83 Previous proteome-scale investigations of binary protein-protein interactions (PPIs)
84 revealed that the yeast two-hybrid (Y2H) system preferentially detects directly interacting
85 proteins^{10,15}, and can thus provide a first-approximation view of the overall organization of
86 complexes. While relatively versatile and amenable to high-throughput strategies, binary
87 interaction assays are somewhat limited by their sensitivity^{16,17}. We have recently
88 demonstrated that this limitation can be overcome by optimally combining complementary
89 assays and/or versions thereof to increase sensitivity without affecting specificity^{18,19}.

90 Here we investigate a new combination of two binary interaction assays, the
91 bioluminescence-based two-hybrid (LuTHy) method¹⁹, and the mammalian cell expression
92 version of the NanoLuc two-hybrid (mN2H) assay¹⁸ to systematically map pairwise interactions
93 between subunits of three multiprotein complexes, LAMTOR, BRISC and MIS12, for which
94 both AP-MS^{7,11,14} and high-resolution structural data are available²⁰⁻²². We find that direct
95 interactions confer significantly higher scores than indirect associations, which allows their
96 differentiation. Finally, we provide experimental evidence that the bioluminescence resonance
97 energy transfer (BRET) readout of LuTHy can be used as a molecular ruler in live cells to
98 estimate distances between protein domains for which no high-resolution structure is currently
99 available. Our combined quantitative PPI mapping approach should thus be applicable for the
100 in-depth characterization of subunit interactions at complexome-scale.

101 RESULTS

102 Benchmarking LuTHy under maximized specificity

103 We previously demonstrated that combining multiple complementary interaction
104 assays and/or versions thereof significantly increases PPI recovery¹⁷⁻¹⁹. Here, we first
105 benchmarked the PPI detection performance of LuTHy, a bioluminescence-based double-
106 readout technology¹⁹, against an established positive reference set (PRS), hsPRS-v2, which
107 contains 60 well-characterized human PPIs^{17,18}. To control for specificity, a random reference
108 set (RRS), hsRRS-v2, made of 78 pairs of human proteins not known to interact, was also
109 tested^{17,18}. With LuTHy, the interaction between two proteins of interest, X and Y, is first
110 measured in living cells by quantification of the BRET signal (LuTHy-BRET; **Supplementary**
111 **Figure 1A**). Then, following cell lysis, the same interaction is assessed *in vitro* by a
112 quantitative luminescence-based co-precipitation readout (LuTHy-LuC; **Supplementary**
113 **Figure 1A**). Since LuTHy plasmids allow expression of each protein as N- or C-terminal
114 fusions, and as donor (NL; NanoLuc tag) or acceptor (mCit; mCitrine tag) proteins, eight
115 tagging configurations can be assessed for every protein pair of interest (**Supplementary**
116 **Figure 1B**). Thus, when all eight configurations are tested, a typical LuTHy experiment with
117 both BRET and LuC measurements generates a total of 16 data points (**Figure 1A,B; Source**
118 **Data Figures 1-2**).

119 PPI detection in hsPRS-v2 by LuTHy was measured under conditions of maximal
120 specificity, i.e. under conditions where none of the random protein pairs from hsRRS-v2 are
121 scored positive (**Supplementary Figure 1C**). In total, using all eight tagging configurations
122 with both readouts (i.e. BRET and LuC), LuTHy identified 43 of the 60 PPIs tested in hsPRS-
123 v2, i.e. 71.7% (**Figure 1C-E**), confirming our previous observation that binary PPI mapping
124 can be maximized when multiple assay versions are combined^{18,19}.

125 As reported in Choi et al¹⁸, under identical conditions of maximal specificity, the union of
126 16 versions of five distinct PPI assays, GPCA, KISS, MAPPIT, Y2H and NanoBiT, identified
127 65.0% of the hsPRS-v2 PPIs, while the union of 12 N2H assay versions reached a detection

128 of 60.0% (**Figure 1E; Supplementary Figure 2A,B**). Thus, LuTHy, with its double readout
129 system, is a highly sensitive and specific PPI detection method that recovers most of the well-
130 established hsPRS-v2 interactions.

131 **Detecting direct interactions supported by high-resolution structural data**

132 To further evaluate the extent to which different binary PPI assays can detect directly
133 interacting proteins, we compared the recovery of the 31 hsPRS-v2 PPIs for which high-
134 resolution structures are available in the RCSB protein data bank (PDB)²³ (hsPRS-PDB), to
135 that of the 29 remaining interactions (hsPRS-non-PDB) (**Figure 2A**). Our analysis revealed
136 that structurally supported PPIs are detected with higher success rates in comparison to
137 interactions for which structural information is not currently available (**Figure 2B**). This was
138 the case for: 1) our previous benchmarking of five binary assays combined (Five assays union,
139 80.6% hsPRS-PDB vs. 48.3% hsPRS-non-PDB); 2) the union of 12 N2H assay versions (N2H
140 union, 71.0% vs. 48.3%); and 3) the newly benchmarked LuTHy method (LuTHy union, 80.6%
141 vs. 62.1%) (**Figure 2B**). Again, we confirmed our previous finding that the success of
142 interaction detection significantly increases when multiple complementary PPI detection
143 assays, or assay versions, are used. Indeed, we found that combining all seven PPI assays
144 (GPCA, KISS, MAPPIT, Y2H, NanoBiT, N2H and LuTHy) allowed recovering 30 of the 31
145 structurally supported PPIs (i.e. 96.8%) under conditions of maximal specificity (Union all
146 assays, **Figure 2B,C**). This demonstrates that structurally supported, direct PPIs can be
147 recovered with nearly complete success rate when combining currently available binary PPI
148 assays.

149 **Systematically mapping interactions within distinct multiprotein complexes**

150 To generalize these findings, we decided to map PPIs between subunits of well-
151 characterized human multiprotein complexes. We selected three complexes based on the
152 following criteria: 1) identified by AP-MS-based methods^{7,11,14}; 2) contains at least four
153 subunits; 3) at least one 3D structure available in PDB²³; and 4) at least 80% of cloned open

154 reading frames (ORFs) encoding the reported subunits present in the human ORFeome 8.1
155 collection²⁴. This resulted in a list of 24 distinct protein complexes (**Supplementary Table 1**),
156 among which three structurally diverse candidates with well-characterized biological functions
157 were prioritized: 1) the LAMTOR complex, also termed “Ragulator” complex²⁰, which regulates
158 MAP kinases and mTOR activities and consists of seven subunits (LAMTOR1, LAMTOR2,
159 LAMTOR3, LAMTOR4, LAMTOR5, RRAGA and RRAGC); 2) the MIS12 complex that
160 connects the kinetochore to microtubules²¹, and is made of five subunits (CENPC1, DSN1,
161 MIS12, NSL1 and PMF1); and 3) the BRISC complex, a large deubiquitinating machinery²²
162 consisting of five proteins (ABRAXAS2, BABAM1, BABAM2, BRCC3 and SHMT2)
163 (**Figure 3A, Supplementary Table 3**).

164 To map interactions between subunits of the LAMTOR, MIS12 and BRISC complexes,
165 we systematically tested all possible pairwise combinations with the LuTHy and mN2H assays
166 (**Source Data Figures 3-5**). Except for LAMTOR5, which was not available in the human
167 ORFeome 8.1 collection, all 16 ORFs encoding the selected target proteins were sequence
168 verified and cloned into both LuTHy and N2H expression plasmids. A resulting search space
169 of 256 pairwise combinations, corresponding to a total of 16 subunits for the three complexes
170 (16x16 matrix; **Figure 3B**), was thus systematically explored with the LuTHy and mN2H
171 assays. To score a tested interaction as positive, we rationalized that true binary PPIs should
172 only be found between the respective subunits of a given complex (i.e. intra-complex pairs),
173 but not between subunits belonging to different complexes (i.e. inter-complex pairs).
174 Therefore, we treated all inter-complex pairs as negative controls, similar to protein pairs from
175 a random reference set. We observed that LuTHy and mN2H fusion constructs showed a
176 broad distribution of interaction scores among those inter-complex pairs (**Supplementary**
177 **Figure 3A-F**). Therefore, we defined construct-specific cutoffs under conditions of maximal
178 specificity, where none of the inter-complex protein pairs are scored positive with any of the
179 assay versions. Using this strategy, we found that LuTHy and mN2H recovered 35 (68.6%)
180 and 23 (45.1%) of the 51 intra-complex interactions, respectively, while 42 of those protein
181 pairs (82.4%) were detected when combining both assays (**Figure 3C; Supplementary**

182 **Figure 4**). We then compared those results to the ones from the BioPlex dataset^{7,11,14}.
183 According to BioPlex, 66.7% of all intra-complex interactions but none of the inter-complex
184 protein pairs for the LAMTOR, BRISC and MIS12 complexes were detected by AP-MS
185 (**Figure 3C**). This demonstrates that, similar to AP-MS-based techniques, binary PPI mapping
186 with LuTHy and mN2H can recover a large fraction of interactions between subunits of distinct
187 multiprotein complexes with high sensitivity and specificity.

188 **Differentiating direct interactions from indirect associations**

189 Differentiating direct interactions from indirect associations within each multiprotein
190 complex identified by AP-MS-based methods is an important, unresolved challenge. In fact, it
191 was estimated that only ~32% of the detected protein associations in the BioPlex dataset are
192 direct interactions¹¹. Therefore, we evaluated if the binary PPI detection assays, LuTHy and
193 mN2H, could identify directly interacting subunits within multiprotein complexes. To define
194 direct interactions and indirect associations between the subunits of each complex, we used
195 3D structural information available in three public databases: PDB²³, Interactome3D²⁵ and
196 PDBePISA²⁶. While directly interacting proteins share a physical interface, we defined indirect
197 associations as pairs of subunits that do not share an interaction interface (**Figure 4A**). In
198 total, we classified 31 direct interactions and 7 indirect associations between the subunits of
199 the LAMTOR, BRISC and MIS12 complexes. Noticeably, we excluded 13 separate
200 homodimer interactions from all further analyses since they were not reported in the available
201 multiprotein complex structures (**Supplementary Table 2**). When applying the same cutoffs
202 used to differentiate intra- from inter-complex protein pairs, 70.9%, 45.2%, and 48.4% of the
203 direct interactions could be detected with LuTHy-BRET, LuTHy-LuC, and mN2H, respectively
204 (**Supplementary Figure 5A**). While mN2H only detected direct interactions, LuTHy-BRET
205 and LuTHy-LuC also recovered 71.4% and 28.6% of the indirect associations, respectively
206 (**Supplementary Figure 5B**). The union of LuTHy and mN2H detected a total of 80.6% of the
207 structure-based direct interactions, and 85.7% of the indirect associations (**Supplementary**
208 **Figure 5A,B**). In comparison, according to the BioPlex dataset, 87.1% of the direct

209 interactions along with 100% of the indirect associations were recovered by AP-MS
210 (**Supplementary Figure 5A,B**).

211 To further evaluate if we could confidently differentiate between direct interactions and
212 indirect associations, we analyzed the quantitative scores obtained with LuTHy and mN2H.
213 We observed that direct interactions generate significantly higher scores compared to indirect
214 associations in the LuTHy-BRET (**Figure 4B**), LuTHy-LuC (**Figure 4C**) and mN2H (**Figure**
215 **4D**) assays. Since LuTHy recovered both direct interactions and indirect associations, we
216 applied a machine learning-based clustering algorithm on the LuTHy-BRET and LuTHy-LuC
217 scores to classify LuTHy-positive protein pairs as direct or not-direct interactions. We used the
218 detected hsPRS-PDB PPIs as a positive training set for direct interactions, and the hsRRS-v2
219 pairs as a negative training set for not-direct interactions (**Supplementary Table 4**). Next, we
220 applied the trained cluster algorithm to the 23 direct interactions and six indirect associations
221 detected by LuTHy (**Supplementary Figure 5A,B**). Using this unbiased approach, we were
222 able to classify 14 of the LuTHy-positive direct interactions (45.2%) as true direct PPIs, without
223 wrongly classifying any of the indirect associations (**Figure 4E,F, Supplementary Figure 5C**).
224 After combining these results with the ones from the mN2H assay, we were able to confidently
225 recover and classify 67.7% of the direct interactions among the LAMTOR, BRISC and MIS12
226 complexes. In comparison, protein pair associations reported in the BioPlex database are
227 currently neither classified as direct PPIs or indirect associations (**Figure 4F, Supplementary**
228 **Figure 5C**)^{7,11,14}.

229 To better understand why a broad range of quantitative scores is obtained when the
230 same pair of directly interacting proteins is tested in eight different tagging configurations
231 (**Supplementary Figure 6A-I**), we explored the structural features of all direct interactions
232 among the three complexes. Using PDBePISA²⁶ we obtained the sizes of interaction
233 interfaces as well as the total surface areas for all directly interacting subunits within the
234 LAMTOR, BRISC and MIS12 complexes (**Supplementary Table 5**). We observed a stronger
235 correlation between interaction scores and the sizes of the interaction interfaces, rather than
236 to the total surface areas of the protein pairs (**Figure 4G; Supplementary Figure 7A-F**),

237 indicating that the larger the interaction interface, i.e. the stronger the binding²⁷, the higher the
238 interaction score. Noticeably, LuTHy-BRET showed relatively high corrected BRET (cBRET)
239 ratios for small interaction interfaces (<500 Å², **Supplementary Figure 7A**), which supports
240 our previous findings that low affinity interactions can be detected with this assay¹⁹.
241 Surprisingly, a significant negative correlation between cBRET ratios and the total surface
242 areas of protein pairs was observed (**Figure 4G; Supplementary Figure 7D-F**), suggesting
243 that the larger the co-complex, the lower the resonance energy transfer between the N- or C-
244 terminally fused NanoLuc luciferase and mCitrine tags. This indicates that both affinity and
245 distance between the tagged proteins studied with binary PPI assays can be determinants for
246 the intensity of the quantitative scores, and thus for their detection.

247 To test whether the LuTHy-BRET readouts could be used to measure distances
248 between protein domains, we determined the molecular distances between all protein termini
249 among the three complexes using available structural information (**Source Data Figures 3-**
250 **5**). We then compared those reference values to the measured cBRET ratios and found a
251 significant negative correlation between them and the molecular distances of the protein
252 termini (**Supplementary Figure 7G**). Interestingly, no such relationship could be observed for
253 the non-energy transfer-based corrected LuC (cLuC) or mN2H ratios (**Supplementary**
254 **Figure 7H,I**). In addition, the distances between protein termini of indirect associations
255 identified with LuTHy-BRET were found to be within a similar range of distances than that of
256 direct interactions (**Supplementary Figure 7J**). This suggests that LuTHy-BRET can detect
257 indirect associations between proteins even in the absence of a common interaction interface
258 as long as the tagged proteins are within a distance where an energy transfer between the
259 donor and the acceptor can occur.

260 Altogether, these results show that the quantitative scores obtained with LuTHy and
261 mN2H can be used to systematically identify direct interactions within multiprotein complexes
262 and differentiate them from indirect associations. They also highlight that the measured
263 cBRET ratios are correlated to the molecular distances between the protein termini of directly

264 interacting subunits, which implies that LuTHy-BRET measurements could also be used to
265 measure distances directly in live cells.

266 **Measuring molecular distances between subunits of protein complexes in live cells**

267 Förster resonance energy transfer-based techniques have been previously used as
268 spectroscopic rulers^{28,29} to monitor conformational changes in proteins, or to estimate the
269 proximity relationships of macromolecules in complexes³⁰⁻³². Since high-resolution structural
270 data are currently unavailable or incomplete for many protein complexes¹³, we investigated
271 whether the LuTHy-BRET readout could be used to measure distances between full-length
272 subunits of protein complexes in live cells. To establish a cBRET-distance standard curve, we
273 used the structure-based molecular distances, the cBRET ratios of directly interacting subunits
274 in the LAMTOR, BRISC and MIS12 complexes, and the structurally supported hsPRS-v2 PPIs
275 (i.e. hsPRS-PDB) as references (**Supplementary Figure 8A, Source Data Figures 1-2 and**
276 **3-5**). However, many of these protein pairs are, on average, structurally covered by less than
277 75% of their respective full-length sequences (**Supplementary Figures 8B and 9;**
278 **Supplementary Tables 5 and 6, Source Data Figures 3-5**). Therefore, for the calibration,
279 we selected a subset of 44 protein pairs (i.e. eight different interactions in multiple tagging-
280 configurations) where the tagged protein termini are fully resolved in the associated structures.
281 For this subset of structurally resolved protein pairs, we generated a standard curve by plotting
282 the cBRET ratios against the measured distances. We observed a strong negative correlation
283 between the cBRET ratios and the reported distances ($R = -0.58$), which is best described by
284 a sigmoidal fit (**Supplementary Figure 8C**). This is in good agreement with previous reports
285 describing that the energy transfer efficiency is inversely proportional to the sixth power of the
286 distance between the donor and the acceptor²⁹. This result strongly suggests that cBRET
287 ratios can indeed be used to measure distances between protein domains in live cells.

288 We next compared cBRET ratios measured between full-length proteins *in cellulo*, to
289 the reported distances in the corresponding high-resolution structures where truncated
290 proteins were used. For example, in the structure of the DR1-DRAP1 PPI, all 176 amino acids

291 of DR1 are resolved, while only the first 77 of the 205 amino acids of DRAP1 are structurally
292 defined³³. Thus, ~62% of the DRAP1 protein, including its entire C-terminal domain, does not
293 appear in the co-complex structure. The molecular distances reported in the structure suggest
294 that the resolved termini of DRAP1 (N-terminus and glutamic acid 77) are both ~20 Å apart
295 from the N-terminus of DR1 (**Supplementary Figure 8D**). When comparing these reported
296 distances to the LuTHy-BRET results, we observed higher cBRET ratios when DRAP1 and
297 DR1 were both N-terminally tagged, compared to C-terminally fused DRAP1 and N-terminally
298 fused DR1 (**Supplementary Figure 8E**). This suggests that the physical distance between
299 the C-terminus of full-length DRAP1 and the N-terminus of DR1 is indeed larger in live cells
300 than the ~20 Å reported in the structure where a truncated DRAP1 protein was used. To
301 estimate the distance between the structurally unresolved termini in the full-length co-complex,
302 we inferred the distance from the LuTHy-BRET data by using the calibrated cBRET-distance
303 standard curve (**Supplementary Figure 8C**). The measured cBRET ratios indicated that the
304 DRAP1 C-terminus is ~61 Å apart from the N-terminus, and ~74 Å apart from the C-terminus
305 of DR1 (**Figure 4H,I**). To generalize these findings, we extended the analysis to 30 additional
306 PPIs that scored positive in LuTHy-BRET, and for which structural information on the tagged
307 protein termini is currently missing (**Figure 4J**). As expected, we found that distances between
308 full-length proteins in live cells inferred from LuTHy-BRET ratios do not generally match with
309 the structurally reported distances (**Supplementary Figure 8F, Supplementary Figure 10**),
310 confirming the results obtained for the DR1-DRAP1 interaction.

311 Together, these findings confirm that LuTHy-BRET measurements can provide
312 valuable, complementary information about molecular distances between the subunits of
313 protein complexes in live cells.

314 **Assessing disease mutation-induced dynamics in structurally unresolved protein** 315 **domains of the Huntingtin-HAP40 interaction**

316 To evaluate if the LuTHy-BRET readout could be used to analyze the effect of a disease
317 mutation on the distance between protein domains, we studied the Huntington's disease

318 related Huntingtin-HAP40 complex. Huntington's disease is caused by a CAG-repeat
319 expansion located in exon-1 of the Huntingtin gene (*HTT*) that is translated into an elongated
320 polyglutamine (polyQ) stretch within the first 83 amino acids of the >3000 amino acids HTT
321 protein^{34,35}. Recently, the structure of full-length HTT, in complex with its binding partner
322 HAP40, was solved at high-resolution by cryo-EM (**Figure 5A**)³⁶. However, the polyQ stretch
323 located at the N-terminus of HTT was not structurally solved, probably due to its high
324 flexibility³⁶. Thus, the structure and relative localization of the N-terminal, polyQ-containing
325 domain remains elusive.

326 To investigate if the expansion mutation could influence the relative arrangement and
327 molecular distance between the N-terminal, polyQ-containing domain of HTT and HAP40, we
328 assessed this interaction by LuTHy-BRET. We observed a significant reduction in cBRET
329 ratios when comparing the HTT-HAP40 interaction where HTT contains either a pathogenic
330 (Q145), or a physiological (Q23) polyQ tract (**Figure 5B**). To determine the distance between
331 the two proteins, we used the cBRET-distance standard curve (**Supplementary Figure 8C**)
332 and found that the N-terminus of HAP40 is ~53 Å apart from the N-terminus of HTT with the
333 non-pathogenic Q23 tract (**Figure 5A; Supplementary Table 7**). However, the distance
334 increases to ~58 Å when HTT contains the elongated, pathogenic Q145 stretch (**Figure 5A;**
335 **Supplementary Table 7**). This suggests that the polyQ expansion mutation alters the domain
336 arrangements within the HTT-HAP40 complex, and that LuTHy-BRET is a powerful tool to
337 measure such subtle changes in live cells.

338 To assess if the effect of the elongated polyQ stretch on the relative arrangement of the
339 N-terminal HTT domain is also detectable in the absence of HAP40, we analyzed the HTT
340 protein in its apo form. The structure of the apo-HTT protein was recently solved and it was
341 proposed that its C-terminal HEAT (Huntingtin/Elongation factor 3/protein phosphatase
342 2A/TOR) domain alters its position relative to the N-terminal HEAT domain upon polyQ
343 expansion³⁷. To monitor such a dynamic structural change within the HTT protein in live cells,
344 we generated an intramolecular LuTHy-BRET sensor, where the NanoLuc (NL) luciferase is
345 fused to the N-terminus, and the fluorescent acceptor, mCitrine (mCit), is incorporated into an

346 unstructured loop at leucine 2680 (L2680, **Figure 5C-E**). We then quantified LuTHy-BRET
347 signals for the resulting NL-HTTQ23-mCit(L2680) and NL-HTTQ145-mCit(L2680) sensor
348 proteins (**Figure 5F**). Using the cBRET-distance standard curve, we interpolated that the
349 HTTQ23 N-terminus is ~ 83 Å apart from the loop at L2680, while this distance increases to
350 ~ 93 Å in the HTTQ145 protein (**Figure 5D; Supplementary Table 7**).

351 Interestingly, it was also shown that the apo-HTT conformation significantly differs from
352 the HTT-HAP40 complex³⁷. Upon HAP40 binding, the C-terminal HEAT domain undergoes a
353 rearrangement, swinging into closer proximity to the N-terminus of HTT (**Figure 5E**). To
354 monitor this dynamic rearrangement in live cells, we used the NL-HTTQ23-mCit(L2680) and
355 NL-HTTQ145-mCit(L2680) LuTHy-BRET sensors and quantified the BRET ratios in presence
356 or absence of exogenous HAP40 protein. When co-expressing HAP40 and the HTT sensors,
357 the BRET ratio significantly increased, indicating that HTT also undergoes this structural
358 rearrangement upon HAP40 binding in live cells (**Figure 5F**). In addition, in the presence of
359 HAP40, the NL-HTTQ23-mCit(L2680) sensor showed a significantly higher BRET signal
360 compared to the NL-HTTQ145-mCit(L2680) sensor, confirming that the polyQ expansion
361 alters the domain arrangements of the HTT-HAP40 complex (**Figure 5E**). Using these data,
362 we interpolated that in the presence of HAP40, the HTTQ23 N-terminus and the structural
363 loop at L2680 are ~ 61 Å apart, whereas they are ~ 71 Å apart with the expanded polyQ tract
364 (**Figure 5D; Supplementary Table 7**). Taken together, these results show that both in the
365 presence or absence of HAP40, the pathogenic polyQ stretch-containing N-terminal domain
366 of HTT is extending ~ 10 Å further away from the core protein, a phenomenon that could not
367 be resolved in the cryo-EM structures. Notably, this extension can make the N-terminal region
368 of the HTT mutant more accessible to engage in additional interactions, which is in good
369 agreement with a previous report³⁸.

370 Overall, these findings demonstrate that the LuTHy-BRET assay can be used to monitor
371 dynamic changes and rearrangements (e.g. induced by mutations) within protein complexes
372 in living cells.

373

374 **DISCUSSION**

375 To globally understand cellular processes, it is essential to characterize complexomes
376 to an extent where all interacting protein subunits of a complex and their relationships to one
377 another are clearly defined. Currently, this knowledge is usually obtained when the 3D
378 structure of a protein complex is resolved. In parallel, the Y2H system has been widely used
379 to generate large PPI networks based on relatively simple yes/no growth readouts. However,
380 many binary PPI assays can generate highly quantitative scores. For example, such assays
381 have been previously used to determine relative binding affinities or to assess the effects of
382 mutations and drugs on specific PPIs^{19,39,40}. Here, we demonstrate that the quantitative
383 readouts of two versatile PPI assays, LuTHy and mN2H, can be used to systematically detect
384 protein interactions within distinct multiprotein complexes, and confidently differentiate direct
385 PPIs from indirect associations. The presented quantitative mapping approach can thus
386 complement high-throughput AP-MS-based and low-throughput structural biology techniques
387 to perform systematic, in-depth characterizations of diverse protein complexes.

388 We provide different standards to control for the quality of such quantitative mapping
389 efforts: 1) the published hsPRS-v2 and hsRRS-v2, which contain direct interactions and
390 random protein pairs, respectively¹⁸, and 2) the direct interactions and indirect associations
391 from three multiprotein complexes, i.e. LAMTOR, BRISC, and MIS12. These pairs of full-
392 length proteins can systematically be introduced in future screening pipelines to calibrate PPI
393 assays and experimentally determine scoring thresholds that will maximize detection of direct
394 interactions, while minimizing recovery of indirect associations, inter-complex or random
395 protein pairs. Interestingly, while LuTHy also identified indirect associations at a threshold
396 where none of the inter-complex pairs were scored positive, the mN2H assay did not recover
397 any of those pairs. This could potentially be explained by the fact that LuTHy assays are
398 performed under conditions of very low donor protein expression¹⁹. Therefore, it is likely that
399 endogenously expressed members of the complex might bridge the indirectly associating
400 proteins detected by LuTHy. However, direct interactions could confidently be distinguished

401 from indirect associations using a machine learning-based clustering algorithm that should be
402 applicable to analyze the quantitative readouts of other binary PPI assays.

403 Our quantitative mapping approach should also help prioritizing high-confidence, direct
404 interactions for further hypothesis-driven experiments such as those where PPIs are used as
405 drug targets in chemical screens, or as candidates in mutagenesis studies. For example,
406 identifying directly interacting subunits within disease-relevant protein complexes can pave
407 the way for the development of small-molecule inhibitors or stabilizers of the studied complex.
408 In addition, when a direct interaction is identified, libraries of protein fragments⁴¹ could easily
409 be tested in order to identify minimal interacting domains between the two partners. This could,
410 for example, guide structural biology efforts to solve specific 3D co-complex structures.

411 Since many macromolecular assemblies still remain unresolvable by traditional
412 structural biology techniques, it is necessary to integrate complementary experimental data⁴²
413 and theoretical approaches⁴³⁻⁴⁸. In addition, 3D structures often use truncated protein
414 constructs and provide snapshots of the studied complexes outside their cellular
415 environment^{28,43}. Indeed, all direct interactions occurring in live cells might not always be
416 captured in the 3D model(s) currently available, and they are thus limited in fully representing
417 the *in vivo* dynamics of complexes with full-length proteins. However, data obtained by binary
418 interaction assays can provide important, complementary information on protein complex
419 architectures in a more physiological environment. For example, single-molecule Förster
420 resonance energy transfer (smFRET) has provided valuable results as an integrative
421 structural biology technique by delivering information on distances between protein subunits
422 within complexes in live cells^{28,31}. We are now adding LuTHy-BRET to the list of available
423 integrative structural biology tools since we show that this readout can be used to measure
424 distances between protein subunits and domains as well as to monitor dynamic
425 conformational changes in live cells. Since LuTHy-BRET measurements are obtained from a
426 population of molecules and not from a single molecule, it detects averaged effects within the
427 studied cell population. Furthermore, BRET-based approaches offer three advantages over
428 techniques like smFRET: 1) they can be applied *in vivo*⁴⁹; 2) they have higher signal-to-noise

429 ratios due to the absence of autofluorescence and photodestruction⁵⁰; and 3) they do not
430 require specialized equipment for in-cell measurements and can thus be easily implemented
431 in any lab equipped with microtiter plate readers.

432 Finally, we have shown that the polyQ expansion mutation in the Huntington's disease
433 related HTT protein gives rise to an abnormal domain rearrangement that can be quantified in
434 live cells with LuTHy-BRET. This suggests that this assay can potentially be used in future
435 experiments to identify chemical modulators that reverse the effect of this and other mutations.

436 **AUTHOR CONTRIBUTIONS**

437 P.T., S.G.C., J.O., M.V. and E.E.W. conceived the study. P.T., C.S., S.G.C., J.O., P.C., Y.W.,
438 E.S.R., S.G., M.Z., S.B., K.S. and Y.J. designed and performed the experiments and collected
439 the data. P.T. analyzed the majority of the results, with significant contributions from C.S.,
440 S.G.C., J.O., M.S., T.H. while E.S.R., P.C., Y.W., J.C.T., M.A.C., D.E.H. and Y.J. provided
441 critical insights. All authors discussed the results and P.T., C.S., S.G.C., J.O., M.V. and E.E.W.
442 wrote the manuscript.

443 **ACKNOWLEDGEMENTS**

444 The authors would like to thank all members of the Wanker, Vidal, Jacob, and Twizere
445 laboratories for helpful discussions throughout this project. This work was supported by the
446 Helmholtz Association, iMed and Helmholtz-Israel Initiative on Personalized Medicine
447 (Germany); the Federal Ministry of Education and Research and e:med Systems Medicine –
448 IntegraMent 01GS0844 (Germany) all to E.E.W.; as well as by the CHDI Foundation (USA);
449 the German Cancer Consortium DKTK (Germany) and the Deutsche Krebshilfe, ENABLE
450 (Germany) to E.E.W. and P.T. This work was also supported by a Claudia Adams Barr Award
451 to S.G.C., a Fonds de la Recherche Scientifique (F.R.S.-FNRS)-Télévie Grant (FC27371,
452 Credit no 7454518F) and a Wallonia-Brussels International (WBI)-World Excellence
453 Fellowship to J.O. Additional supports were provided by NIH grants P50HG004233 and

454 R01GM130885 awarded to M.V. and U41HG001715 awarded to M.V., D.E.H. and M.A.C. This
455 work was also supported by the LabEx IBEID (grant 10-LABX-0062). M.V. is a Chercheur
456 Qualifié Honoraire, and J.C.T. a Maître de Recherche of the Fonds de la Recherche
457 Scientifique (F.R.S.-FNRS, Wallonia-Brussels Federation, Belgium).

458 **CONFLICT OF INTEREST**

459 The authors declare that they have no conflict of interest.

460 **DATA AND CODE AVAILABILITY**

461 The protein interactions from this publication have been submitted to the IMEx
462 (<http://www.imexconsortium.org>) consortium through IntAct⁵¹ and assigned the identifier
463 IM-29174. All UniProt and RCSB-PDB accession codes are provided in the Source Data.
464 Python and R codes used for data analyses are available upon request.

465 **METHODS**

466 **ORF sequencing and plasmid generation**

467 For hsPRS-v2 and hsRRS-v2 proteins, the corresponding sequence-verified entry vectors
468 published in Choi et al¹⁸ were Gateway cloned into the different LuTHy destination plasmids.
469 ORFs for subunits of the LAMTOR, MIS12 and BRISC complexes were taken from the CCSB
470 human ORFeome 8.1, which is a sequence confirmed clonal collection of human ORFs in a
471 Gateway entry vector system²⁴. In total, 16 entry plasmids were picked from the collection,
472 single clones were isolated, and ORFs were PCR-amplified and confirmed by bi-directional
473 Sanger DNA sequencing. Entry clones were shuttled into LuTHy and N2H destination vectors
474 using the Gateway Cloning Technology. All resulting vectors were analyzed by PCR-
475 amplification of cloned ORFs and DNA gel electrophoresis (N2H plasmids), or restriction
476 digestion and sequence validation (LuTHy plasmids). For the LuTHy assay, additional control

477 plasmids (PA-NL, Addgene #113445; PA-mCit-NL, Addgene #113444; PA-mCit, Addgene
478 #113443; NL, Addgene #113442) were used, as previously described¹⁶. The pcDNA3.1
479 plasmids encoding HTTQ23 and HTTQ145 (glutamines encoded by CAG/CAA triplets) were
480 gifts from the CHDI foundation that were subcloned into pDONR221 entry vectors
481 (ThermoFisher, #12536017). The pDONR221-F8A1 entry plasmid encoding HAP40 was
482 obtained from Source BioSciences (OCAAo5051D1091D). *HTT* and *HAP40* entry clones were
483 shuttled into LuTHy destination plasmids, and *HAP40* was also cloned into a pDEST26-cmyc
484 destination plasmid, which was a kind gift from Matthias Selbach. All amino acid positions in
485 the HTT protein refer to a sequence containing 17 glutamines with a total length of 3138 amino
486 acids. To generate the intramolecular LuTHy-BRET HTT sensors, the LuTHy plasmids
487 pcDNA3.1 NL-HTTQ23 and pcDNA3.1 NL-HTTQ145 were digested with SfiI and SfoI that cut
488 within the HTT sequence and the Kan/neoR resistance sequence of the plasmid backbone.
489 The resulting larger fragment (12161 bp for Q23, and 12563 bp for Q145) were agarose gel
490 purified using the Invisorb[®] Fragment Cleanup kit from Invitex Molecular. Next, the HTT
491 sequence from Q2500 to L2680 was PCR amplified using the primers 353-FWD and 341-
492 REV. The 353-FWD primer contained a 30 bp overhang into the *HTT* sequence, and the 341-
493 REV primer an 18 bp GSGSGS-linker sequence as well as a 7 bp overhang into the 5'-
494 mCitrine sequence. The coding sequence for mCitrine was amplified without the start and stop
495 codons, using the 342-FWD and the 343-REV primers. The 342-FWD and 343-REV primers
496 both contained an 18 bp GSGSGS-linker sequence and a 7 bp overhang into the respective
497 5' or 3' *HTT* sequences. The *HTT* sequence from amino acid P2681 to the Kan/neoR cassette
498 of the plasmid backbone was amplified using the following primers: 340-FWD and 354-REV.
499 The 340-FWD primer contained an 18 bp GSGSGS-linker sequence and a 7 bp overhang into
500 the 3'-mCitrine sequence, and the 354-REV primer a 30 bp overhang into the Kan/neoR
501 cassette of the plasmid backbone. All PCR products were column purified using the MSB[®]
502 Spin PCRapace kit from Invitex Molecular. Finally, the digested and gel purified plasmid
503 backbones were assembled, together with the three purified PCR products, by Gibson
504 assembly according to the manufacturer's protocol (New England Biolabs, E2611). The

505 resulting colonies carrying the final plasmids were analyzed by restriction digestion and
506 Sanger DNA sequencing. All primer sequences can be found in **Supplementary Table 8**.

507 **LuTHy assay procedure**

508 The LuTHy assay was performed as previously described¹⁹. In brief, HEK293 cells were
509 reversely transfected in white 96-well microtiter plates (Greiner, #655983) at a density of 4.0-
510 4.5×10^4 cells per well with plasmids encoding donor and acceptor proteins. After incubation
511 for 48 h, mCitrine fluorescence was measured in intact cells (Ex/Em: 500 nm/530 nm). For
512 LuTHy-BRET assays, coelenterazine-h (pjk, #102182) was added to a final concentration of
513 5 μ M (5 mM stock dissolved in methanol). Next, cells were incubated for an additional 15 min
514 and total luminescence as well as luminescences at short (370-480 nm) and long (520-570
515 nm) wavelengths were measured using the Infinite[®] microplate readers M200, M1000, or
516 M1000 PRO (Tecan). After luminescence measurements, the luminescence-based co-
517 precipitation (LuC) assay was performed. Cells were lysed in 50-100 μ L HEPES-phospho-
518 lysis buffer (50 mM HEPES, 150 mM NaCl, 10% glycerol, 1% NP-40, 0.5% deoxycholate, 20
519 mM NaF, 1.5 mM MgCl₂, 1 mM EDTA, 1 mM DTT, 1 U Benzonase, protease inhibitor cocktail
520 (Roche, EDTA-free), 1 mM PMSF, 25 mM glycerol-2-phosphate, 1 mM sodium orthovanadate,
521 2 mM sodium pyrophosphate) for 30 min at 4°C. Lysates (7.5 μ L) were transferred into small
522 volume 384-well microtiter plates (Greiner, #784074) and fluorescence (mCit_{IN}) was measured
523 as previously described¹⁹. To measure the total luminescence (NL_{IN}), 7.5 μ L of 20 μ M
524 coelenterazine-h in PBS was added to each well and the plates incubated for 15 more minutes.
525 For LuC, small volume 384-well microtiter plates (Greiner, #784074) were coated with sheep
526 gamma globulin (Jackson ImmunoResearch, #013-000-002) in carbonate buffer (70 mM
527 NaHCO₃, 30 mM Na₂CO₃, pH 9.6) for 3 h at room temperature, and blocked with 1% BSA in
528 carbonate buffer before being incubated overnight at 4°C with rabbit anti-sheep IgGs in
529 carbonate buffer (Jackson ImmunoResearch, #313-005-003). 15 μ L of cell lysate was
530 incubated for 3 h at 4°C in the IgG-coated 384-well plates. Then, all wells were washed three
531 times with lysis buffer and mCitrine fluorescence (mCit_{OUT}) was measured as described¹⁹.

532 Finally, 15 μ L of PBS buffer containing 10 μ M coelenterazine-h was added to each well and
533 luminescence (NL_{OUT}) was measured after a 15 min incubation period.

534 **LuTHy data analysis**

535 Data analysis was performed as previously described¹⁹. In brief, the LuTHy-BRET and LuTHy-
536 LuC ratios from BRET and co-precipitation measurements are calculated as follows:

$$537 \quad (1) \text{ BRET ratio} = \frac{LWL}{SWL} - Cf \text{ with } Cf = \frac{LWL_{PA-NL}}{SWL_{PA-NL}}$$

538 with LWL and SWL being the detected luminescences at long (520–570 nm) and short (370–
539 480 nm) wavelengths, respectively. The correction factor (Cf) represents the donor bleed-
540 through value from the PA-NL only construct. The corrected BRET (cBRET) ratio is calculated
541 by subtracting the maximum BRET ratios of control 1 (NL/PA-mCit-Y), or of control 2 (NL-
542 X/PA-mCit) from the BRET ratio of the studied interaction (NL-X/PA-mCit-Y).

543 For the LuC readout, the obtained luminescence precipitation ratio (PIR) of the control protein
544 PA-NL (PIR_{PA-NL}) is used for data normalization, and is calculated as follows:

$$545 \quad (2) \text{ PIR}_{PA-NL} = \frac{NL_{OUT}}{2 * NL_{IN}}$$

546 With NL_{OUT} being the total luminescence measured after co-IP and NL_{IN} the luminescence
547 measured in the cell extracts, directly after lysis. Subsequently, LuC ratios are calculated for
548 all interactions of interest, and normalized to the PIR_{PA-NL} ratio:

$$549 \quad (3) \text{ LuC ratio} = \frac{NL_{OUT}/2 * NL_{IN}}{PIR_{PA-NL}}$$

550 Finally, a corrected LuC (cLuC) ratio is calculated by subtracting either the LuC ratio of
551 control 1 (NL/PA-mCit-Y), or of control 2 (NL-X/PA-mCit) from the LuC ratio of the studied
552 interaction (NL-X/PA-mCit-Y). The calculated LuC ratios obtained for controls 1 and 2 are then
553 compared to each other, and the highest value is used to correct the LuC ratio of the respective
554 interaction.

555 **Mammalian cell-based version of the N2H assay (mN2H)**

556 HEK293T cells were seeded at 6×10^4 cells per well in 96-well, flat-bottom, cell culture
557 microplates (Greiner Bio-One, #655083), and cultured in Dulbecco's modified Eagle's medium
558 (DMEM) supplemented with 10% fetal calf serum at 37 °C and 5% CO₂. 24 h later, cells were
559 transfected with 100 ng of each N2H plasmid (pDEST-N2H-N1, -N2, -C1 or -C2) using linear
560 polyethylenimine (PEI) to co-express proteins fused with complementary NanoLuc fragments,
561 F1 and F2. The stock solution of PEI HCl (PEI MAX 40000; Polysciences Inc; Cat# 24765)
562 was prepared according to the manufacturer's instructions. Briefly, 200 mg of PEI HCl powder
563 were added to 170 mL of water, stirred until complete dissolution, and pH was adjusted to 7
564 with 1 M NaOH. Water was added to obtain a final concentration of 1 mg/mL, and the stock
565 solution was filtered through a 0.22 µm membrane. The DNA/PEI ratio used for transfection
566 was 1:3 (mass:mass). 24 h after transfection, the culture medium was removed and 50 µL of
567 100x diluted NanoLuc substrate (Furimazine, Promega Nano-Glo, N1120) was added to each
568 well of a 96-well microplate containing the transfected cells. Plates were incubated for 3 min
569 at room temperature. Luciferase enzymatic activity was measured using a TriStar or CentroXS
570 luminometer (Berthold; 2 s integration time).

571 **Processing publicly available interaction data**

572 Publicly available binary protein interaction datasets used in this study came from the original
573 Choi et al publication¹⁸. The hsPRS-v2 PPI recovery rates were extracted as published, in
574 conditions where none of the hsRRS-v2 pairs are scored positive. To generate hsPRS-PDB,
575 structurally supported interactions from hsPRS-v2 were identified using interactome insider
576 (<http://interactomeinsider.yulab.org>)⁵², and using both, co-crystal structures and homology
577 modeling (**Source Data Figures 1-2**). The BioPlex dataset^{7,11,14} was downloaded from
578 <https://bioplex.hms.harvard.edu/interactions.php> on May 10, 2021.

579 **Scoring hsPRS-v2 PPIs under conditions where no hsRRS-v2 pair is scored positive**
580 **for LuTHy**

581 Source data for each assay version performed in this study are presented in the **Source Data**
582 **Figures 1-2**. Each binary PPI experiment with LuTHy-BRET and LuTHy-LuC was performed
583 six times, with two biological and three technical replicates. For each protein pair X-Y, we
584 calculated the corrected BRET and LuC ratios as described above. To identify protein pairs
585 that scored positive at a threshold of no hsRRS-v2 detection for a given tagging configuration,
586 only cBRET or cLuC ratios higher than the highest hsRRS-v2 score for that tagging
587 configuration were considered. All hsPRS-v2 pairs that did not meet this criterion were not
588 scored positive and therefore defined as not detected in the corresponding assay version.

589 **Complex selection and definition of direct interactions and indirect associations**

590 Human protein complexes used in this study were selected based on the following criteria.
591 First, human protein complexes should have at least one experimentally determined structure
592 in PDB²³ (<https://www.rcsb.org>). Second, the complex should have at least four subunits.
593 Third, at least 80% of entry clones for individual subunits of a complex should be present in
594 the human ORFeome 8.1 collection²⁴. A total of 24 distinct complexes (**Supplementary**
595 **Table 1**) with different PDB structures met those criteria, and three protein complexes with
596 well-documented biological functions were selected from this list: LAMTOR, BRISC and
597 MIS12. These complexes were also picked as they appear in AP-MS interactome maps,
598 according to BioPlex^{7,11,14}. Finally, direct interactions and indirect associations within protein
599 complexes were determined by interactome 3D²⁵ and PDBePISA²⁶.

600 **Scoring interactions within multiprotein complexes**

601 Data for LuTHy and mN2H mapping of multiprotein complexes can be found in **Source Data**
602 **Figures 3-5**. Recovery of interactions by LuTHy-BRET, LuTHy-LuC and mN2H was
603 calculated based on construct-specific cutoffs. Therefore, the quantitative scores were

604 determined as described in the LuTHy and mN2H methods sections, respectively. For each
605 construct, X or Y, involved in the X-Y interaction, a cutoff was defined as the highest inter-
606 complex pair score for each of the two respective constructs (X and Y). The final cutoff for a
607 tested protein pair between constructs X and Y was then determined as the maximum of the
608 two cutoff values.

609 **Cluster analysis of LuTHy data**

610 For supervised classification of direct and not-direct interaction clusters, a Gaussian finite
611 mixture model using the R package mclust⁵³ was used. First, the LuTHy-BRET and LuTHy-
612 LuC quantitative scores (i.e. cBRET and cLuC ratios) from 63 LuTHy-positive interactions
613 among hsPRS-PDB and from 585 protein pairs in hsRRS-v2 (**Supplementary Table 3**) were
614 used as positive and negative training sets, respectively. The MclustDA discriminant analysis
615 function was applied to those training sets using the eigenvalue decomposition discriminant
616 analysis (EDDA) method. The training set performed at a 2.36% classification error rate. Next,
617 the trained algorithm was used to cluster 117 intra-complex protein pairs (**Supplementary**
618 **Table 3**) that scored above the construct-specific cutoffs in LuTHy assays. Using the function
619 predict.MclustDA, 47 direct protein pairs (i.e. multiple configurations corresponding to 14 out
620 of the 23 LuTHy-positive direct interactions) were correctly classified as direct interactions and
621 all 12 indirect protein pairs (i.e. multiple configuration corresponding to all 6 LuTHy-positive
622 indirect associations) were accurately classified as not-direct. In addition, 58 direct protein
623 pairs (i.e. multiple configuration corresponding to 9 out of the 23 LuTHy-positive direct
624 interactions) could not be classified as direct interactions using this supervised classification
625 analysis.

626 **Determining interaction interface areas, complex surface areas and molecular** 627 **distances between protein termini**

628 Interaction interface and total complex surface areas were determined according to
629 PDBePISA²⁶ (https://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) on April 16, 2021. Protean

630 3DTM (DNASTAR[®]) was used to measure molecular distances between protein termini of
631 interactions in the hsPRS-PDB as well as between direct interactions and indirect associations
632 within the multiprotein complexes: LAMTOR (RCSB PDB: 6EHR), MIS12 (RCSB PDB: 5LSJ)
633 and BRISC (RCSB PDB: 6H3C). The determined molecular distances can be found in **Source**
634 **Data Figures 3-5**. Nonlinear and linear regressions between cBRET ratios and molecular
635 distances between protein termini were calculated using GraphPad Prism 9 and the fits were
636 compared using the extra-sum-of-squares F test with a bottom constraint to 0. GraphPad
637 Prism 9 was also used to interpolate molecular distances and their 90% confidence intervals
638 using the measured cBRET ratios.

639

640

641 **REFERENCES**

642

- 643 1. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell*
644 **144**, 986–998 (2011).
- 645 2. Gorka, M. *et al.* Protein Complex Identification and quantitative complexome by CN-
646 PAGE. *Sci Rep-uk* **9**, 11523 (2019).
- 647 3. Anger, A. M. *et al.* Structures of the human and Drosophila 80S ribosome. *Nature* **497**,
648 80–85 (2013).
- 649 4. Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I. & Klaholz, B. P. Visualization
650 of chemical modifications in the human 80S ribosome structure. *Nature* **551**, 472–477
651 (2017).
- 652 5. Lu, Y. *et al.* Conformational Landscape of the p28-Bound Human Proteasome Regulatory
653 Particle. *Mol Cell* **67**, 322–333.e6 (2017).
- 654 6. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by
655 stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
- 656 7. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human
657 Interactome. *Cell* **162**, 425–440 (2015).
- 658 8. Luck, K., Sheynkman, G. M., Zhang, I. & Vidal, M. Proteome-Scale Human Interactomics.
659 *Trends Biochem Sci* **42**, 342–354 (2017).
- 660 9. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein
661 identifies proximal and interacting proteins in mammalian cells. *J Cell Biology* **196**, 801–
662 810 (2012).
- 663 10. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**,
664 402–408 (2020).
- 665 11. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the
666 human interactome. *Cell* **184**, 3022–3040.e28 (2021).
- 667 12. Drew, K. *et al.* Integration of over 9,000 mass spectrometry experiments builds a global
668 map of human protein complexes. *Mol Syst Biol* **13**, 932 (2017).
- 669 13. Drew, K., Wallingford, J. B. & Marcotte, E. M. hu.MAP 2.0: integration of over 15,000
670 proteomic experiments builds a global compendium of human multiprotein assemblies.
671 *Mol Syst Biol* **17**, (2021).
- 672 14. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities
673 and disease networks. *Nature* **545**, 505 (2017).
- 674 15. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**,
675 1212–1226 (2014).

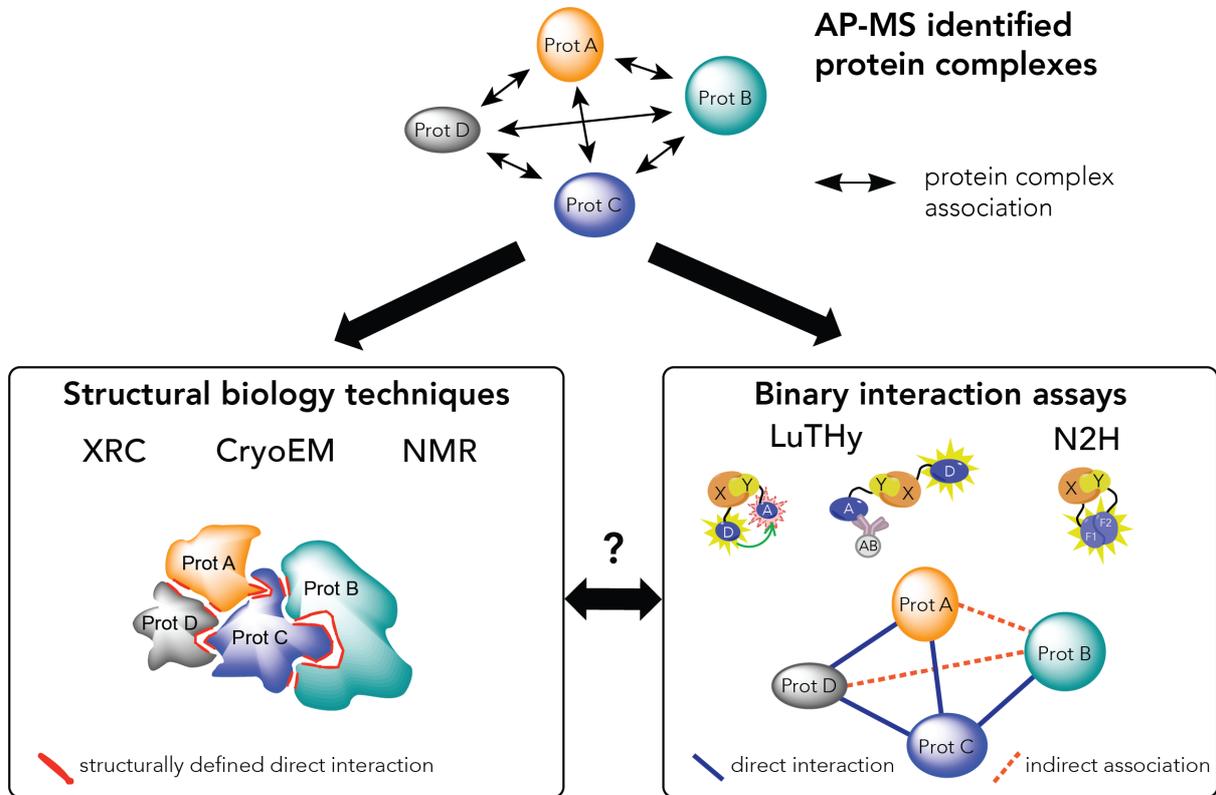
- 676 16. Yu, H. *et al.* High-Quality Binary Protein Interaction Map of the Yeast Interactome
677 Network. *Science* **322**, 104–110 (2008).
- 678 17. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat*
679 *Methods* **6**, 83–90 (2009).
- 680 18. Choi, S. G. *et al.* Maximizing binary interactome mapping with a minimal number of
681 assays. *Nat Commun* **10**, 3907 (2019).
- 682 19. Trepte, P. *et al.* LuTHy: a double-readout bioluminescence-based two-hybrid technology
683 for quantitative mapping of protein–protein interactions in mammalian cells. *Mol Syst Biol*
684 **14**, e8071 (2018).
- 685 20. Araujo, M. E. G. de *et al.* Crystal structure of the human lysosomal mTORC1 scaffold
686 complex and its impact on signaling. *Science* **358**, 377–381 (2017).
- 687 21. Petrovic, A. *et al.* Structure of the MIS12 Complex and Molecular Basis of Its Interaction
688 with CENP-C at Human Kinetochores. *Cell* **167**, 1028–1040.e15 (2016).
- 689 22. Rabl, J. *et al.* Structural Basis of BRCC36 Function in DNA Repair and Immune
690 Regulation. *Mol Cell* **75**, 483–497.e9 (2019).
- 691 23. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
- 692 24. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat*
693 *Methods* **8**, 659–661 (2011).
- 694 25. Mosca, R., Céol, A. & Aloy, P. Interactome3D: adding structural details to protein
695 networks. *Nat Methods* **10**, 47–53 (2012).
- 696 26. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline
697 State. *J Mol Biol* **372**, 774–797 (2007).
- 698 27. Chen, J., Sawyer, N. & Regan, L. Protein–protein interactions: General trends in the
699 relationship between binding affinity and interfacial buried surface area. *Protein Sci* **22**,
700 510–515 (2013).
- 701 28. Lerner, E. *et al.* Toward dynamic structural biology: Two decades of single-molecule
702 Förster resonance energy transfer. *Science* **359**, eaan1133 (2018).
- 703 29. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc National Acad*
704 *Sci* **58**, 719–726 (1967).
- 705 30. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem* **47**,
706 819–846 (1978).
- 707 31. Asher, W. B. *et al.* Single-molecule FRET imaging of GPCR dimers in living cells. *Nat*
708 *Methods* 1–9 (2021) doi:10.1038/s41592-021-01081-y.
- 709 32. Quast, R. B. & Margeat, E. Single-molecule FRET on its way to structural biology in live
710 cells. *Nat Methods* 1–2 (2021) doi:10.1038/s41592-021-01084-9.

- 711 33. Kamada, K. *et al.* Crystal Structure of Negative Cofactor 2 Recognizing the TBP-DNA
712 Transcription Complex. *Cell* **106**, 71–81 (2001).
- 713 34. MacDonald, M. E. *et al.* A novel gene containing a trinucleotide repeat that is expanded
714 and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
- 715 35. Scherzinger, E. *et al.* Huntingtin-Encoded Polyglutamine Expansions Form Amyloid-like
716 Protein Aggregates In Vitro and In Vivo. *Cell* **90**, 549–558 (1997).
- 717 36. Guo, Q. *et al.* The cryo-electron microscopy structure of huntingtin. *Nature* **555**, 117–120
718 (2018).
- 719 37. Jung, T. *et al.* The Polyglutamine Expansion at the N-Terminal of Huntingtin Protein
720 Modulates the Dynamic Configuration and Phosphorylation of the C-Terminal HEAT
721 Domain. *Structure* **28**, 1035–1050.e8 (2020).
- 722 38. Harding, R. J. *et al.* Design and characterization of mutant and wild-type huntingtin
723 proteins produced from a toolkit of scalable eukaryotic expression systems. *J Biol Chem*
724 jbc.RA118.007204 (2019) doi:10.1074/jbc.ra118.007204.
- 725 39. Trepte, P. *et al.* DULIP: A Dual Luminescence-Based Co-Immunoprecipitation Assay for
726 Interactome Mapping in Mammalian Cells. *J Mol Biol* **427**, 3375–3388 (2015).
- 727 40. Saraon, P. *et al.* A drug discovery platform to identify compounds that inhibit EGFR triple
728 mutants. *Nat Chem Biol* 1–10 (2020) doi:10.1038/s41589-020-0484-2.
- 729 41. Boxem, M. *et al.* A protein domain-based interactome network for *C. elegans* early
730 embryogenesis. *Cell* **134**, 534–545 (2008).
- 731 42. Braberg, H. *et al.* Genetic interaction mapping informs integrative structure determination
732 of protein complexes. *Science* **370**, eaaz4910 (2020).
- 733 43. Ward, A. B., Sali, A. & Wilson, I. A. Integrative Structural Biology. *Science* **339**, 913–915
734 (2013).
- 735 44. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**,
736 1384–1403 (2019).
- 737 45. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep
738 learning. *Nature* **577**, 706–710 (2020).
- 739 46. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-
740 track neural network. *Science* eabj8754 (2021) doi:10.1126/science.abj8754.
- 741 47. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human
742 proteome. *Nature* 1–9 (2021) doi:10.1038/s41586-021-03828-1.
- 743 48. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 1–11
744 (2021) doi:10.1038/s41586-021-03819-2.
- 745 49. Mezzanotte, L., Root, M. van 't, Karatas, H., Goun, E. A. & Löwik, C. W. G. M. In Vivo
746 Molecular Bioluminescence Imaging: New Tools and Applications. *Trends Biotechnol* **35**,
747 640–652 (2017).

- 748 50. Boute, N., Jockers, R. & Issad, T. The use of resonance energy transfer in high-
749 throughput screening: BRET versus FRET. *Trends Pharmacol Sci* **23**, 351–354 (2002).
- 750 51. Orchard, S. *et al.* The MIntAct project—IntAct as a common curation platform for 11
751 molecular interaction databases. *Nucleic Acids Res* **42**, D358–D363 (2014).
- 752 52. Meyer, M. J. *et al.* Interactome INSIDER: a structural interactome browser for genomic
753 studies. *Nature Methods* **15**, 107–114 (2018).
- 754 53. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification
755 and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289–317 (2016).
- 756
- 757

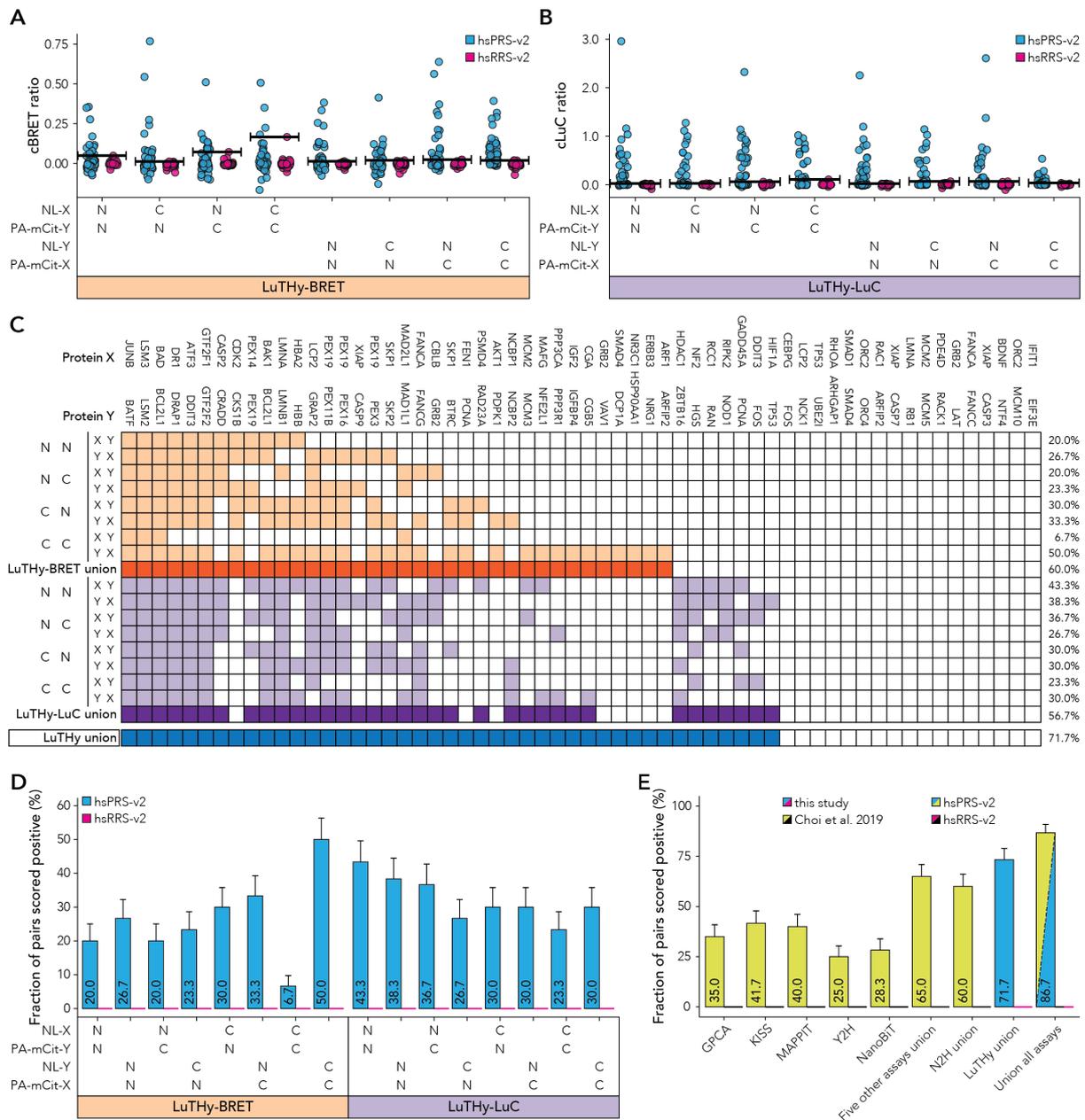
758 **FIGURES**

759 **Graphical Abstract**



760

761

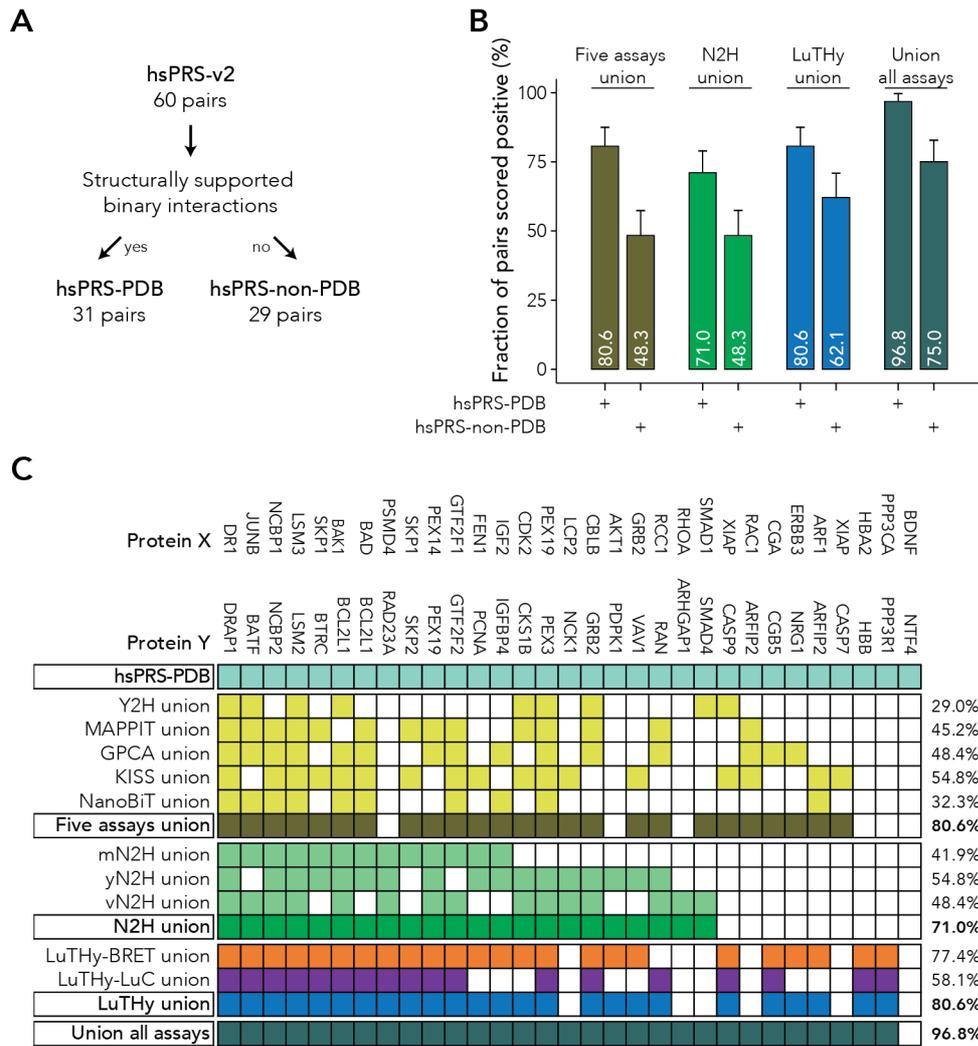


762

763 **Figure 1. Benchmarking the LuThy assay against hsPRS-v2 under conditions where none of the hsRRS-**
 764 **v2 pairs are scored positive.**

765 Quantitative scores for hsPRS-v2 and hsRRS-v2 pairs differentiated by tagging configurations for **(A)** LuThy-BRET
 766 and **(B)** LuThy-LuC. Horizontal lines in **(A)** and **(B)** indicate the scoring cutoffs for each configuration above which
 767 no hsRRS-v2 pair is scored positive. **(C)** Overview of the positively scored interactions from hsPRS-v2. Positive
 768 LuThy-BRET interactions in different orientations (light orange) and their union (dark orange) are shown, as well
 769 as positive LuThy-LuC interactions in the eight different configurations (light purple) and their union (dark purple).
 770 LuThy union (blue) summarizes results for all 16 versions. The percentage (%) at the end of each row represents
 771 the fraction of hsPRS-v2 PPIs scored positive. **(D)** Recovery rates of hsPRS-v2 interactions at no hsRRS-v2

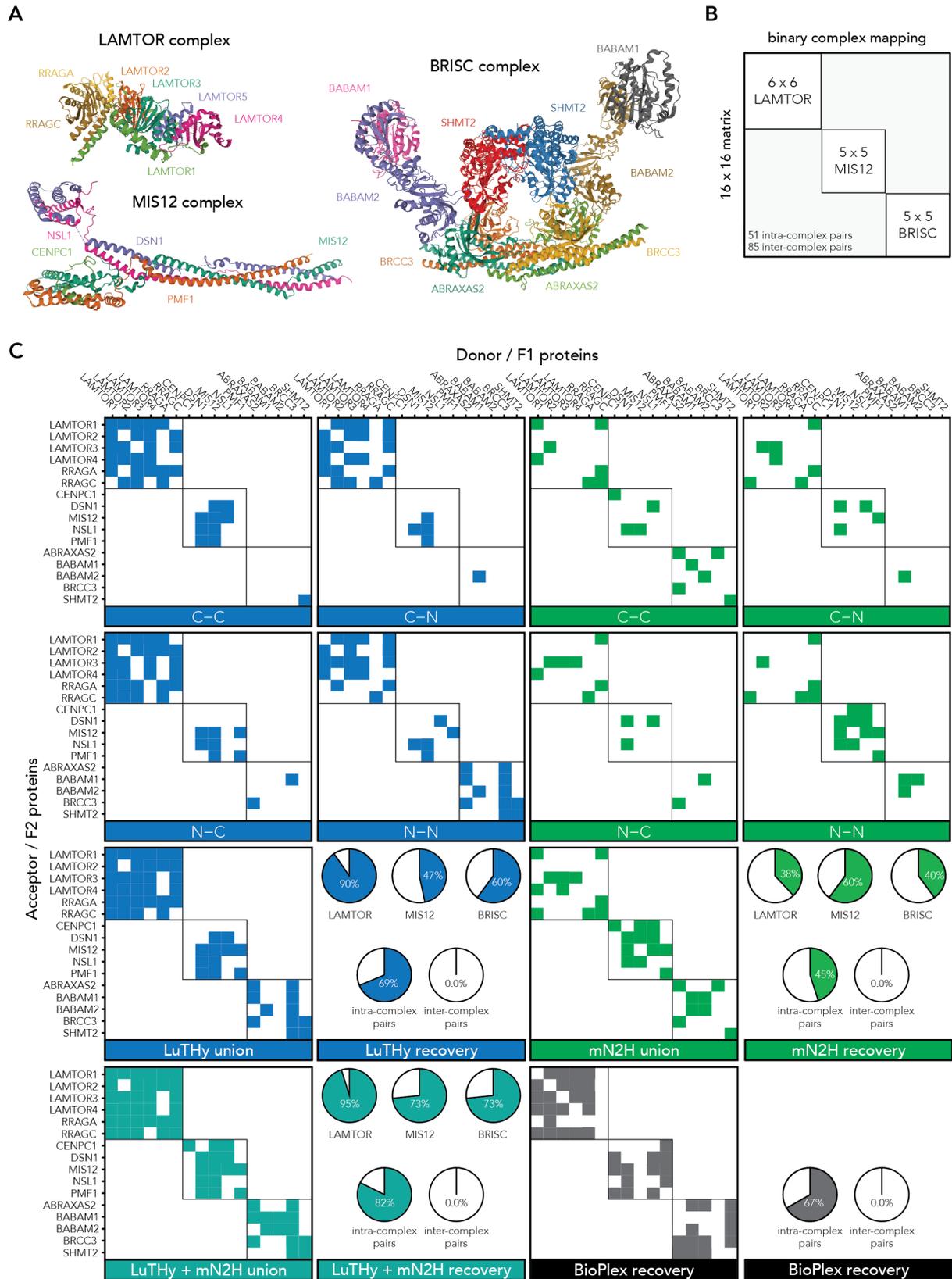
772 detection for the eight different tagging configurations in LuTHy-BRET and LuTHy-LuC. **(E)** Comparison of hsPRS-
773 v2 recovery rates for LuTHy (LuTHy union) to other binary interaction methods benchmarked in Choi et al¹⁸. Error
774 bars indicate standard errors of the proportion in **(D)** and **(E)**.



775

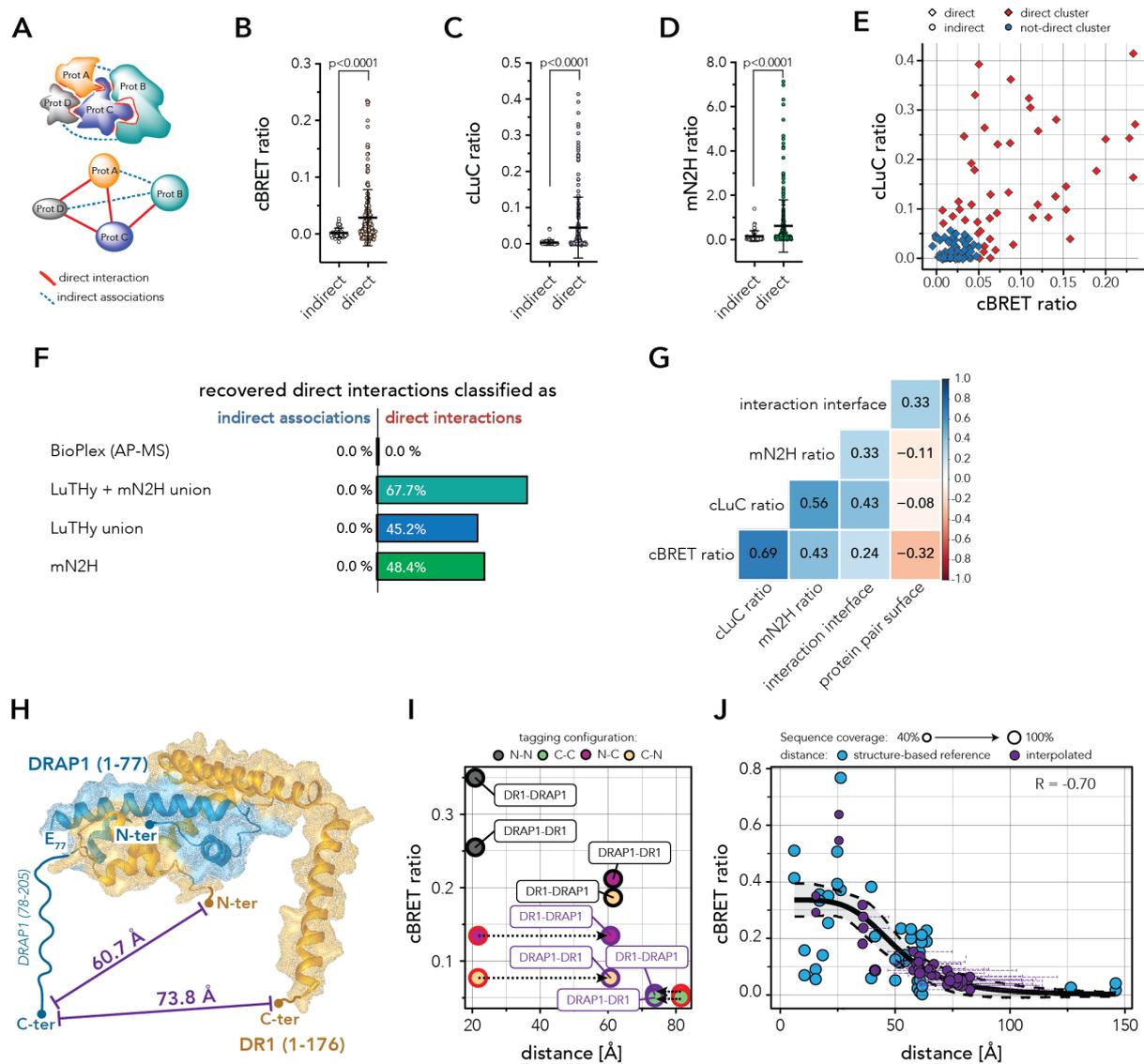
776 **Figure 2. Benchmarking binary PPI assays against structurally supported interactions in hsPRS-v2.**

777 (A) hsPRS-v2 contains 60 binary PPIs among which 31 pairs are supported by structural data in PDB (hsPRS-
 778 PDB). The 29 remaining pairs are not currently supported by 3D structures (hsPRS-non-PDB). (B,C) Comparison
 779 of hsPRS-PDB and hsPRS-non-PDB recovery rates from different binary interaction methods when none of the
 780 hsPRS-v2 pairs are scored positive. Binary PPI assays combined ("Union all assays") recover all but one of the
 781 hsPRS-PDB interactions. The percentage (%) at the end of each row in (C) indicates the fraction of hsPRS-v2 PPIs
 782 scored positive at no hsRRS-v2 detection. Error bars in (B) indicate standard errors of the proportion.



784 **Figure 3. Mapping interactions within multiprotein complexes using the LuTHy and mN2H assays.**

785 (A) Structures of the protein complexes analyzed in this study: LAMTOR (PDB: 6EHR), MIS12 (PDB: 5LSK), and
786 BRISC (PDB: 6H3C). (B) Binary interaction approach to systematically map PPIs within distinct complexes. Every
787 protein subunit from each complex was screened against every other one (all-by-all, 16x16 matrix). (C) Results of
788 the all-by-all interaction screen for the selected multiprotein complexes. Each protein pair was systematically tested
789 in every possible configuration (i.e. C-C, C-N, N-C, N-N), and as donor and acceptor in LuTHy, or as F1 and F2
790 NanoLuc fusions in mN2H. LuTHy union corresponds to the combination of both LuTHy-BRET and LuTHy-LuC
791 results. Protein pairs tested in the LuTHy assays were scored positive if at least one of the two readouts (i.e.
792 LuTHy-BRET and/or LuTHy-LuC) was positive. LuTHy + mN2H union corresponds to the combination of results
793 from LuTHy union and mN2H union. In the pie charts, the top panels show recovery rates of intra-complex pairs
794 within the LAMTOR, MIS12 and BRISC complexes, while the bottom panels show recoveries of all intra-complex
795 and inter-complex pairs. Published BioPlex results, corresponding to AP-MS data, were used for the interaction
796 matrix and pie charts. Proteins used as baits in BioPlex are indicated on the x-axis, whereas identified interaction
797 partners are indicated on the y-axis.

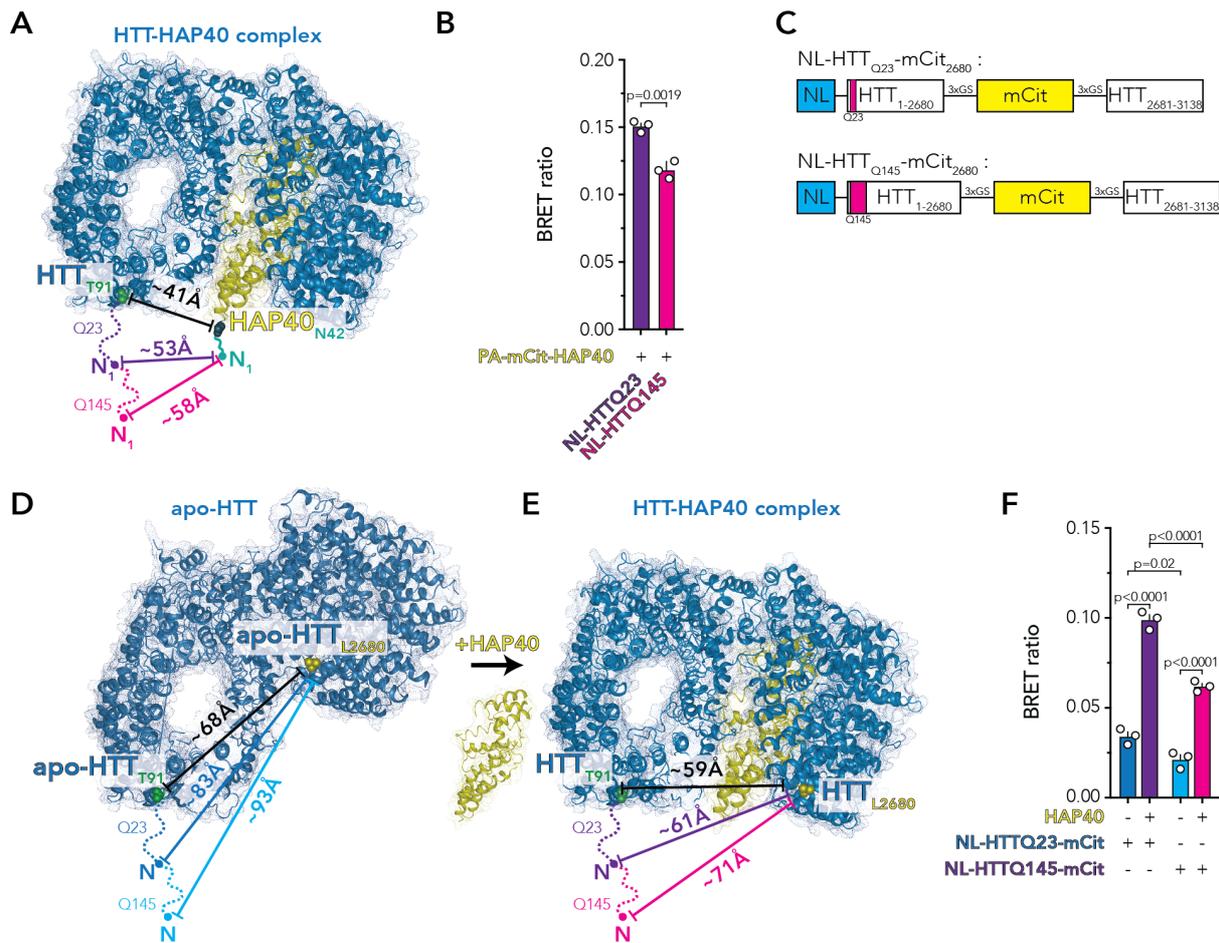


798

799 **Figure 4. Differentiating direct interactions from indirect associations with LuThy and mN2H assays.**

800 (A) Schematic for the definition of direct interactions and indirect associations within complexes using structural
 801 biology data. (B-D) Distribution of the quantitative scores and comparison between direct interactions and indirect
 802 associations. Direct interactions produce significantly higher LuThy-BRET (B), LuThy-LuC (C), and mN2H (D)
 803 scores compared to indirect associations (Welch's two-tailed t-test). (E) Scatter plot of cBRET and cLuC ratios
 804 showing direct interactions (diamonds) and indirect associations (circles) detected by LuThy-BRET or LuThy-LuC.
 805 Protein pairs were clustered by supervised expectation-maximization clustering as direct (red) or not-direct (blue)
 806 interactions using hsPRS-PDB and hsRRS-v2 as training sets. (F) Recovery and classification of structurally
 807 defined direct interactions as true direct interactions or as indirect associations by BioPlex (AP-MS), LuThy and
 808 mN2H. (G) Pearson correlation matrix for direct interactions comparing cBRET, cLuC and mN2H ratios to the
 809 interaction interface areas (\AA^2), or to the total complex surface areas (\AA^2). (H) 3D structure of the DR1-DRAP1

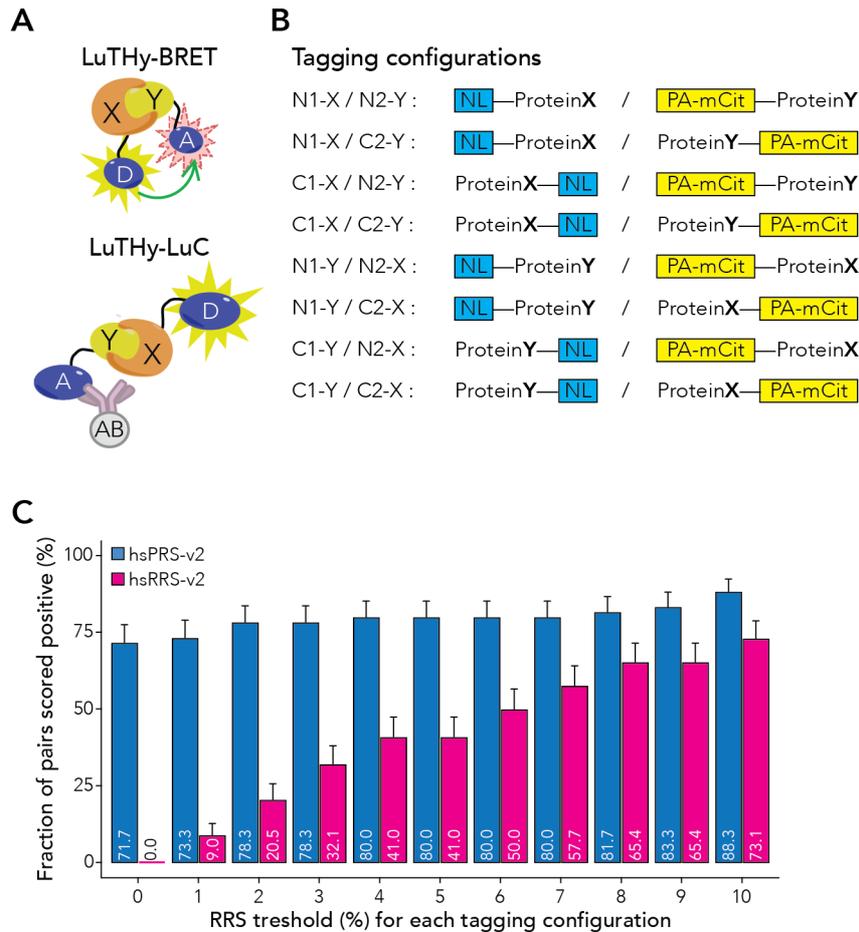
810 interaction (PDB: 1JFI). Full-length DR1 but only the N-terminal region of DRAP1 (1-77, amino acids 78-205
811 lacking) are reported in the 3D structure. The cBRET ratios were used to interpolate distances between the C-
812 terminus of DRAP1 and the N-terminus (60.7 Å), or C-terminus (73.8 Å) of DR1 based on the cBRET-distance
813 standard curve. (I) cBRET ratios for the DR1-DRAP1 PPI are plotted against the molecular distances obtained
814 from the 3D structure. Tagging configurations are colored by N-N (grey), C-C (green), N-C (purple) or C-N (yellow).
815 Each data point on the graph is labeled (framed text) according to the tested tagging configuration: the protein
816 indicated first is tagged with NanoLuc (NL) luciferase, while the second protein is tagged with PA-mCitrine (PA-
817 mCit) (e.g. DR1+DRAP1 (N-N, grey) corresponds to NL-DR1/PA-mCit-DRAP1). Tagging configurations where tags
818 are fused to the structurally unresolved termini in the current 3D structure for one of the two proteins are outlined
819 in red (e.g. DR1+DRAP1 (N-C, purple) corresponds to NL-DR1/DRAP1-mCit-PA). Interaction values outlined in
820 purple represent cBRET ratios against interpolated molecular distances. Differences between structurally
821 determined and interpolated distances are indicated by dotted arrows. (J) cBRET ratios plotted against the
822 structurally determined molecular distances for the 44 protein pairs used as references (blue). The 30 PPIs with
823 tags fused to protein termini not currently resolved in the structures are plotted against interpolated distances
824 between full-length proteins (purple). For each tested pair, the purple horizontal error bar corresponds to the 95%
825 confidence interval of the interpolated molecular distances. Sigmoidal fits were performed on the 44 tested pairs
826 outlined in black, together with the 30 pairs outlined in purple ($R = -0.70$).



827

828 **Figure 5. Interpolating molecular distances within the HTT-HAP40 complex with LuThy-BRET.**

829 (A) Cryo-EM structure of the HTT-HAP40 complex (PDB: 6EZ8). The distances between HTT(T91) and
 830 HAP40(N42) inferred from the structure are indicated (black), along with the distances between the N-terminal
 831 domains of HAP40 and HTTQ23 (purple) or HTTQ145 (magenta) interpolated from the measured BRET ratios. (B)
 832 The BRET ratio between HAP40 and HTTQ23 is significantly higher than the one between HAP40 and HTTQ145
 833 (the bar plots show means \pm standard deviations (sd), statistical significance was calculated by a two-tailed t-test,
 834 $n=3$). (C) Schematic representations of the HTTQ23 and HTTQ145 intramolecular LuThy-BRET sensors. (D,E)
 835 Cryo-EM structures of the apo-HTT protein (D, PDB: 6RMH) and the HTT-HAP40 complex (E, PDB: 6EZ8) where
 836 the distances between residues T91 and L2680 are indicated (black) either in the absence or presence of HAP40.
 837 (F) BRET ratios measured for the intramolecular LuThy-BRET sensors in the absence or presence of HAP40, for
 838 HTTQ23 and HTTQ145 (the bar plots show means \pm sd, statistical significance was calculated by two-way ANOVA
 839 with Tukey's multiple comparisons test, $n=3$). The BRET ratios were used to interpolate distances between the N-
 840 terminus of HTT and residue L2680 for HTTQ23 and HTTQ145, which are indicated in D and E.



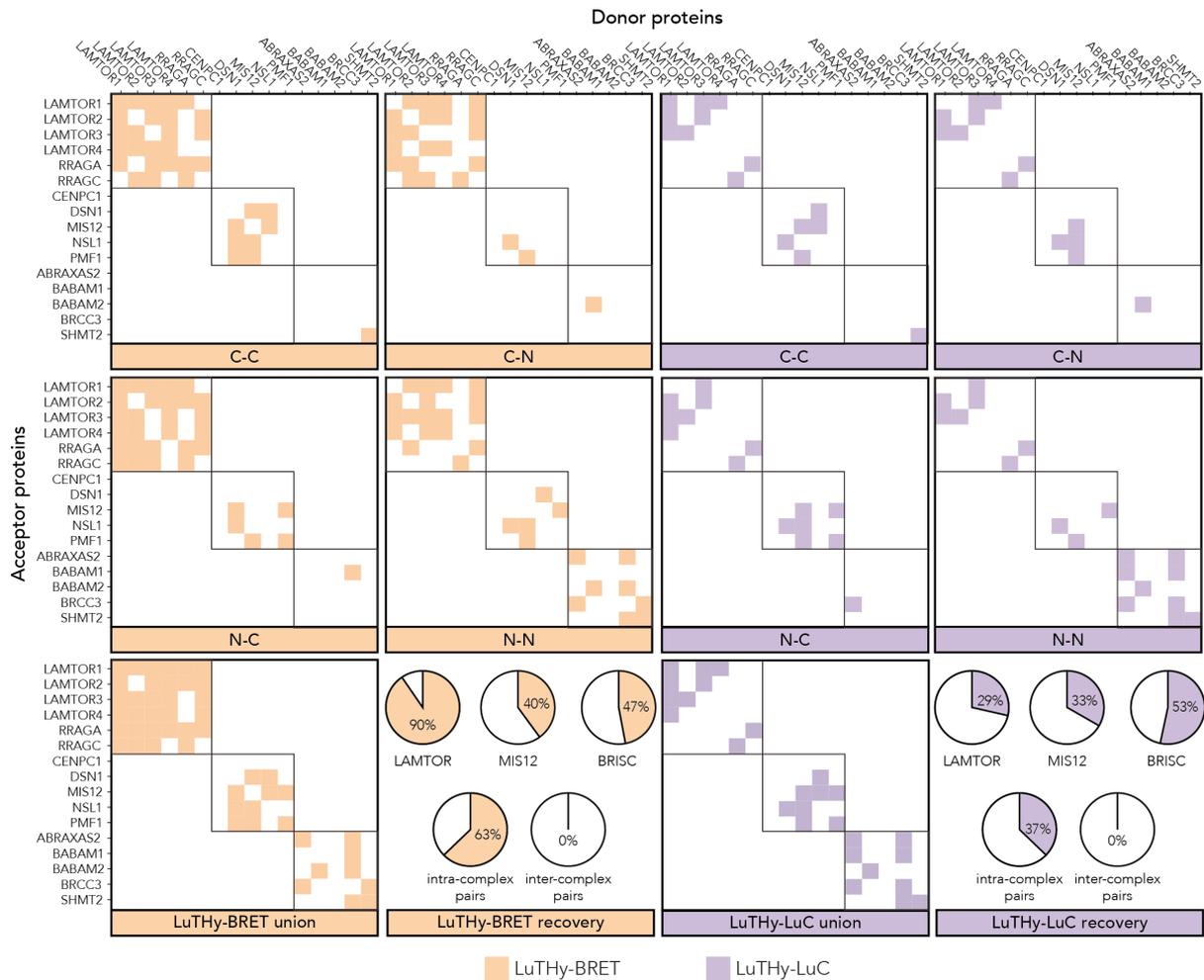
841

842 **Supplementary Figure 1 (related to Figure 1). Tagging configurations and relationship between the hsRRS-**
 843 **v2 threshold and detection of hsPRS-v2 and hsRRS-v2 pairs for LuTHy.**

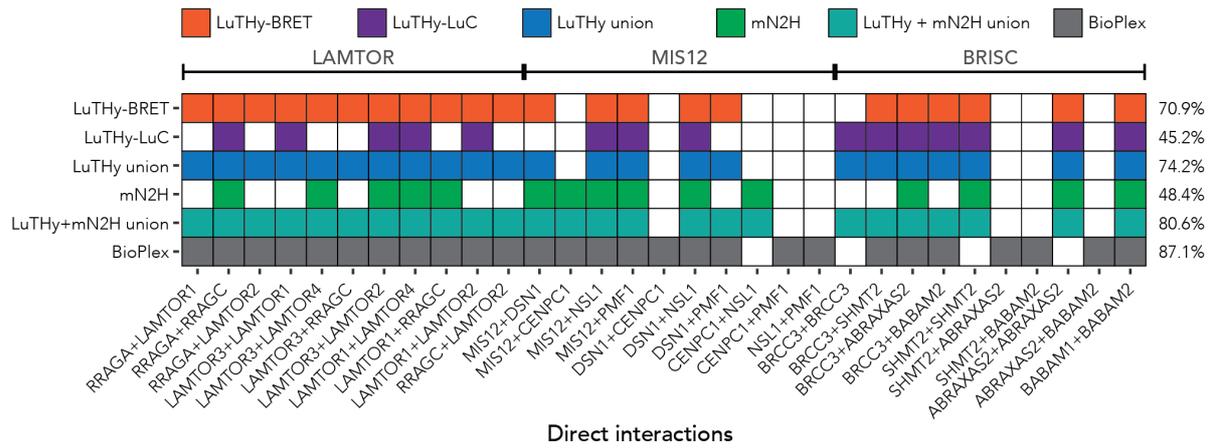
844 (A) Schematic overview of the LuTHy-BRET and LuTHy-LuC assays. X: Protein X, Y: Protein Y, D: NanoLuc donor,
 845 A: mCitrine acceptor, AB: antibody. (B) With the LuTHy assay, each protein pair X-Y can be tested in four possible
 846 configurations (N- vs. C-terminal fusion for each protein), and proteins can be swapped from one tag to the other
 847 for a total of 8 possible LuTHy versions resulting in 16 quantitative scores for each protein pair (i.e. eight for LuTHy-
 848 BRET and eight for LuTHy-LuC). (C) Impact of scoring random protein pairs on PPI recovery for all LuTHy versions:
 849 cumulative detection rates of hsPRS-v2 and hsRRS-v2 pairs when increasing, identical hsRRS-v2 thresholds are
 850 applied for each individual assay version. Error bars in (C) indicate standard errors of the proportion.

852 **Supplementary Figure 2 (related to Figure 1). Performance and complementarity of different binary PPI**
853 **assay versions under conditions where none of the hsRRS-v2 pairs are scored positive.**

854 **(A,B)** Performance of individual binary PPI assay versions benchmarked in Choi et al¹⁸ (Y2H, MAPPIT, GPCA,
855 KISS, NanoBiT, N2H), and in this study (LuTHy). In total, 42 assay versions were tested, reaching ~87% detection
856 of hsPRS-v2 PPIs. The 31 hsPRS-PDB pairs correspond to hsPRS-v2 PPIs currently supported by at least one 3D
857 structure in PDB. The percentage (%) at the end of each row represents the fraction of hsPRS-v2 PPIs scored
858 positive when none of the hsRRS-v2 pairs are recovered. Error bars in **(A)** indicate standard errors of the
859 proportion.

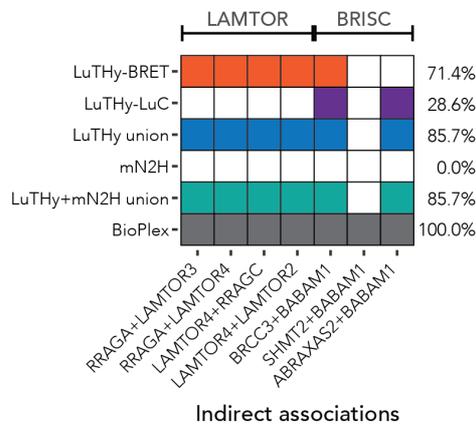


A



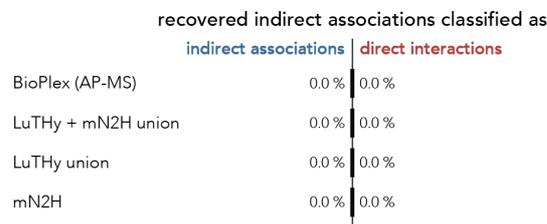
Direct interactions

B



Indirect associations

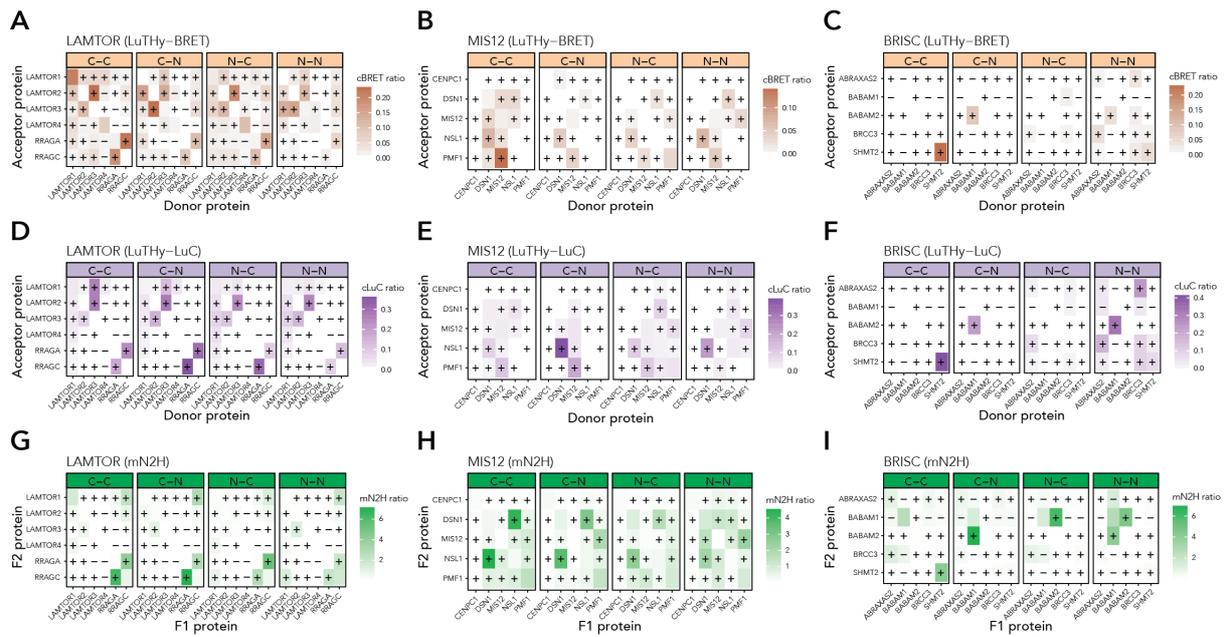
C



879

880 **Supplementary Figure 5 (related to Figure 3 and 4). Summary of the direct interactions and indirect**
 881 **associations detected by different approaches.**

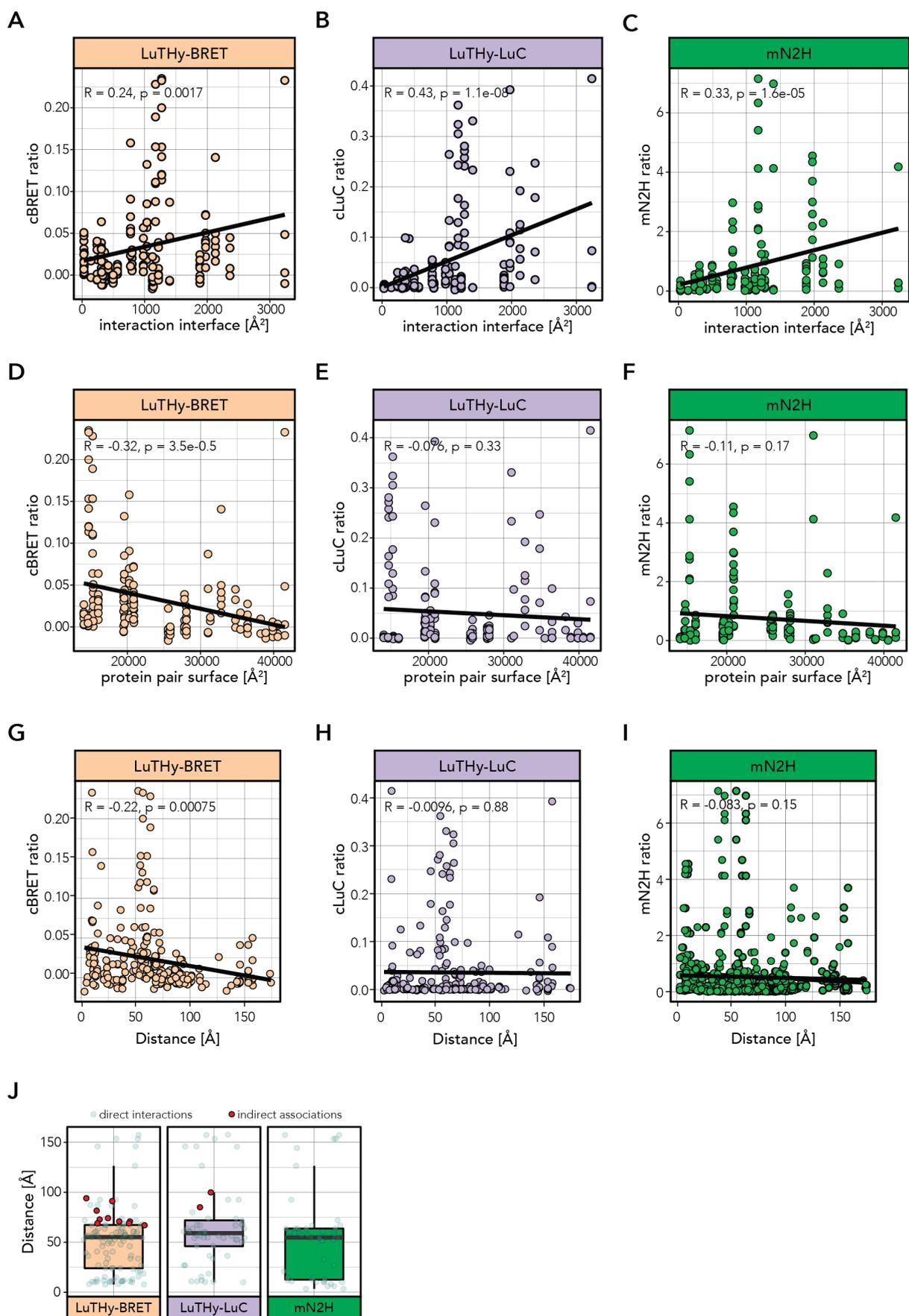
882 Detected direct interactions (**A**) and indirect associations (**B**) by the LuTHy and mN2H binary PPI assays, and by
 883 AP-MS-based (BioPlex) techniques. The percentage (%) at the end of each row represents the fraction of protein
 884 pairs recovered (for LuTHy and mN2H: at a threshold where no inter-complex pair is scored positive). LuTHy +
 885 mN2H union summarizes the LuTHy and mN2H results. Published reference data from the AP-MS BioPlex dataset
 886 are shown. Structural biology data were used to define direct interactions and indirect associations. (**C**) Recovery
 887 and classification of structurally defined indirect associations as true indirect associations or as direct interactions
 888 for BioPlex, LuTHy and mN2H.



889

890 **Supplementary Figure 6 (related to Figure 4). Quantitative scores for direct interactions and indirect**
 891 **associations according to the tested tagging configurations.**

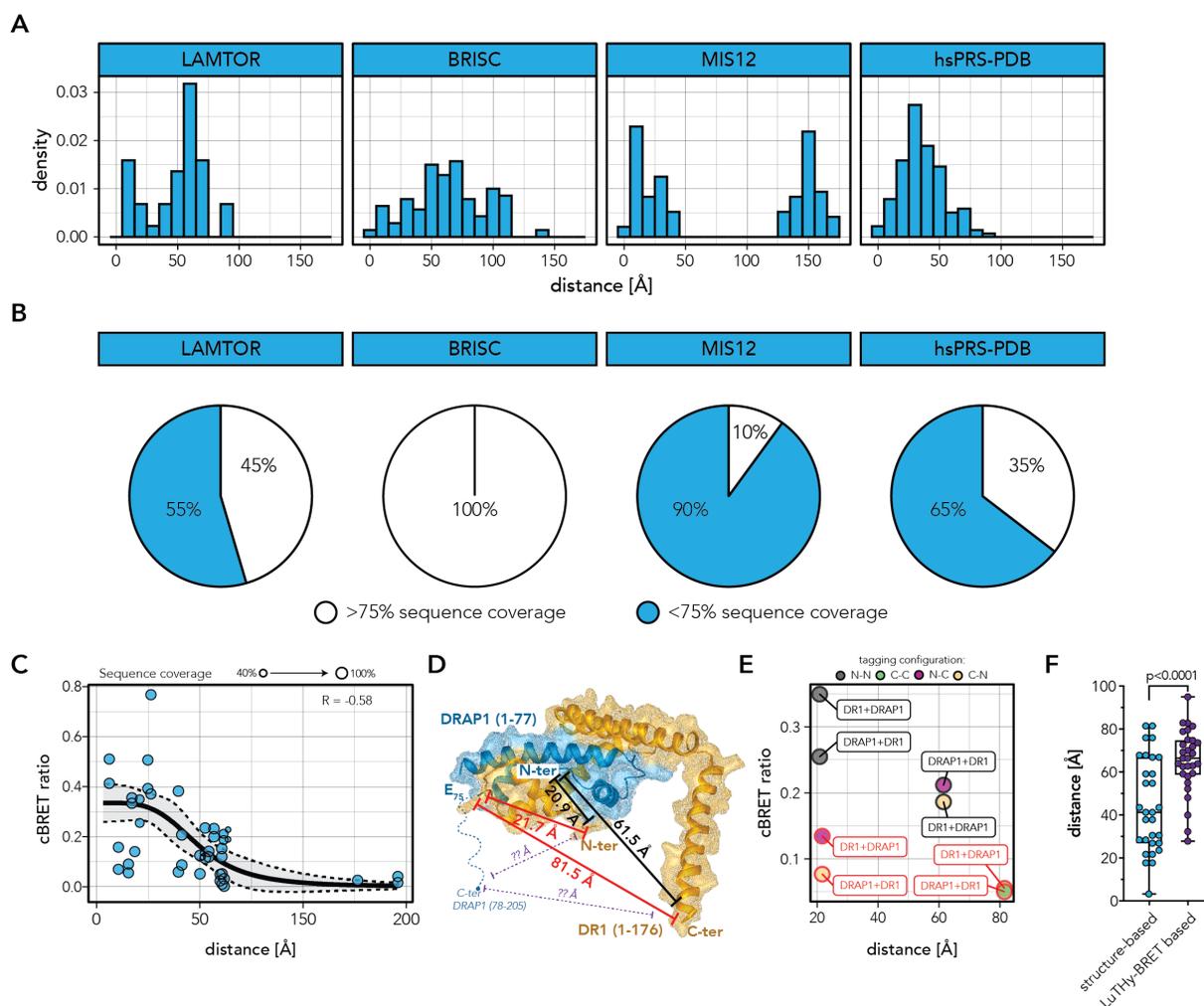
892 Heat maps of all intra-complex protein pairs in the LAMTOR, MIS12 and BRISC complexes, differentiated by
 893 tagging configurations for LuThy-BRET (A-C), LuThy-LuC (D-F) and mN2H (G-I). The range of quantitative scores
 894 is depicted from smaller, or equal to zero (white) to maximal (orange, purple or green for LuThy-BRET, LuThy-
 895 LuC or mN2H, respectively) values. A “+” symbol indicates a direct interaction, a “-” symbol an indirect association,
 896 and an empty cell a separate homodimer interaction.



897

898 **Supplementary Figure 7 (related to Figure 4). Correlation between the LuTHy and mN2H quantitative**
899 **scores and the structural properties of the analyzed complexes.**

900 Correlation between the interaction interfaces between proteins and the cBRET (**A**), cLuC (**B**), and mN2H (**C**) ratios
901 for the corresponding PPIs. Correlation between the total surface areas of interacting proteins and the cBRET (**D**),
902 cLuC (**E**), and mN2H (**F**) ratios. Correlation between the molecular distances (Å) of the tagged protein termini and
903 the cBRET (**G**), cLuC (**H**), and mN2H (**I**) ratios of the corresponding PPIs. Pearson correlation coefficients (R) with
904 p-values are displayed. (**J**) Box plots of the molecular distances (Å) between the tagged protein termini for the
905 detected direct interactions (transparent green) and indirect associations (red) in the LuTHy-BRET, LuTHy-LuC
906 and mN2H assays (displayed are the median, lower and upper hinges showing the 25th and 75th percentiles, lower
907 and upper whiskers extending from the hinges with 1.5x the inter-quartile range).



908

909 **Supplementary Figure 8 (related to Figure 4). Protein sequence coverage and molecular distances between**
 910 **subunits of the studied complexes used to generate the cBRET-distance standard curve.**

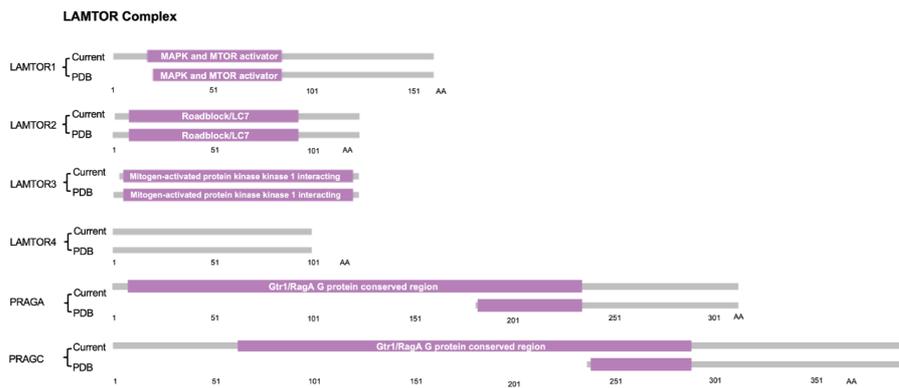
911 **(A)** Relative distribution of structure-based distances for direct interactions within the LAMTOR, BRISC and MIS12
 912 complexes, as well as in the hsPRS-PDB PPIs. **(B)** Proportion of directly interacting proteins reported in the 3D
 913 structures that contain, on average, more (white) or less (green) than 75% of the full-length protein sequences. **(C)**
 914 Correlation between the cBRET ratios and the molecular distances for the 44 protein pairs where at least six
 915 tagging configurations out of the eight tested are scored positive in LuTHy-BRET, and for which the respective
 916 tagged protein termini are structurally resolved. Goodness of the sigmoidal regression fit is indicated on the graph
 917 ($R = -0.58$). **(D)** 3D structure of the DR1-DRAP1 interaction where molecular distances between the last amino
 918 acids of the structurally solved proteins are indicated in red. The dotted blue lines indicate the protein fractions that
 919 are not structurally resolved (DRAP1 C-terminus). Unknown molecular distances between the missing termini in
 920 the 3D structure are indicated by dotted purple lines. **(E)** cBRET ratios for the DR1-DRAP1 interaction are plotted
 921 against the structure-based molecular distances. Tagging configurations are colored by N-N (grey), C-C (green),
 922 N-C (purple) or C-N (yellow). Each data point on the graph is labeled (framed text) according to the tested tagging

923 configuration: the protein indicated first is tagged with NanoLuc (NL) luciferase, while the second protein is tagged
924 with PA-mCitrine (PA-mCit) (e.g. DR1+DRAP1 (N-N, grey) corresponds to NL-DR1/PA-mCit-DRAP1). Tagging
925 configurations where tags are fused to the structurally unresolved termini in the current 3D structure for one of the
926 two proteins are outlined in red (e.g. DR1+DRAP1 (N-C, purple) corresponds to NL-DR1/DRAP1-mCit-PA). (F)
927 Structure-based and predicted distances for the 30 PPIs with tags fused to protein termini not currently resolved in
928 the structures (box and whisker visualizing the median, lower and upper hinges showing the 25th and 75th
929 percentiles, lower and upper whiskers extending from min to max; all data points are shown). Statistical significance
930 was calculated by a two-tailed, paired t-test, n=30.

A



B



C

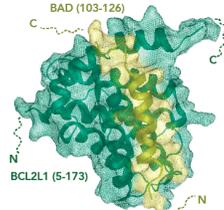


931

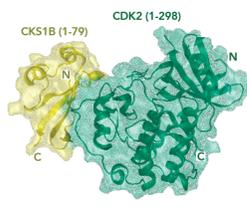
932 **Supplementary Figure 9 (related to Figure 4). Protein sequence coverage of the constructs used in the**
 933 **present binary interaction study (“Current”) compared to those found in the published structures (“PDB”).**

934 Schematic representations of the protein constructs used in the BRISC (**A**), LAMTOR (**B**), and MIS12 (**C**)
 935 complexes. The numbers indicate the lengths of the protein constructs in amino acids (AA). (**A**) The structural study
 936 characterizing the BRISC complex (PDB: 6H3C) used full-length proteins whereas four of the constructs
 937 (ABRAXAS2, BABAM1, BABAM2, BRCC3) used in the current study are C-terminally truncated. The current binary
 938 interaction study uses proteins containing >90% of the BABAM1, BABAM2 and BRCC3, and ~80% of the
 939 ABRAXAS2 full-length reference sequences. (**B**) The structural study characterizing the LAMTOR complex (PDB:
 940 6EHR) uses smaller fragments for the RRAGA and RRAGC subunits. Except for a single AA missing on the N-

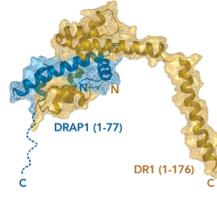
941 terminus of LAMTOR3, the current study uses the full-length sequences for subunits of the LAMTOR complex. (C)
942 Overall, the structural study of the MIS12 complex (PDB: 5LSJ) lacks >50% of the AA sequences for the different
943 subunits.



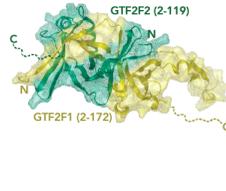
Based on PDB structure 1G5J
BAD (23/168 amino acids covered = 14%)
BCL2L1 (168/233 amino acids covered = 72%)



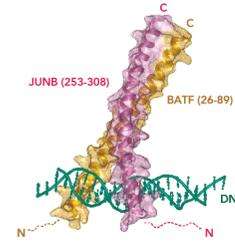
Based on PDB structure 1RUH
CKS1B (79/79 amino acids covered = 100%)
CDK2 (298/298 amino acids covered = 100%)



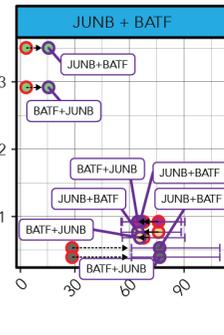
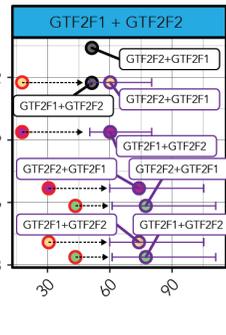
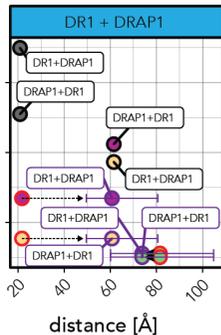
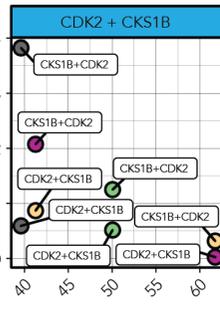
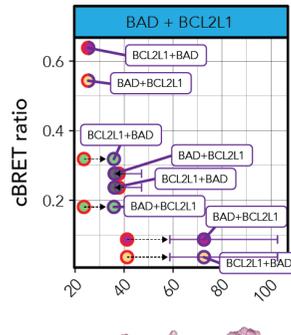
Based on PDB structure 1JFI
DR1 (176/176 amino acids covered = 100%)
DRAP1 (77/205 amino acids covered = 38%)



Based on PDB structure 1E3U
GTF2F1 (171/517 amino acids covered = 33%)
GTF2F2 (118/249 amino acids covered = 47%)



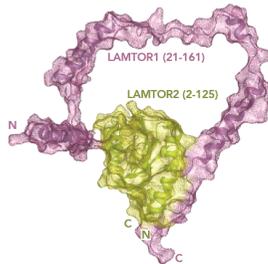
Based on homology PDB structure 1A02
BATF (63/125 amino acids covered = 50%)
bZIP domain from FOS
JUNB(55/331 amino acids covered = 17%)



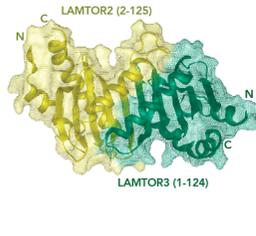
Sequence coverage
○ → ●
0% → 100%

tagging configuration
● N-N
● C-C
● N-C
● C-N

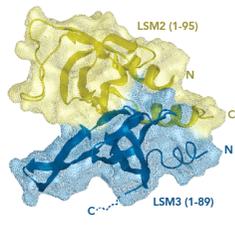
distance
● structure-based reference
● incomplete structural information
● interpolated



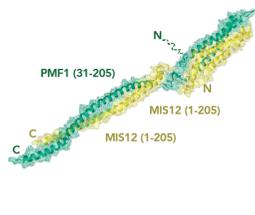
Based on PDB structure 6EHR
LAMTOR1 (140/161 amino acids covered = 87%)
LAMTOR2 (124/125 amino acids covered = 99%)



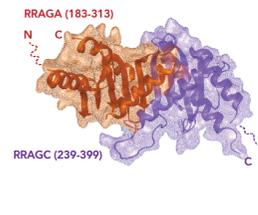
Based on PDB structure 6EHR
LAMTOR2 (124/125 amino acids covered = 99%)
LAMTOR3 (124/124 amino acids covered = 100%)



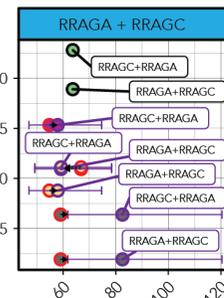
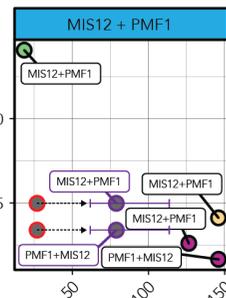
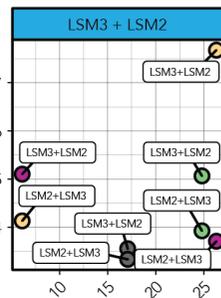
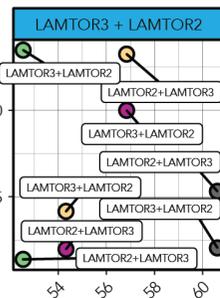
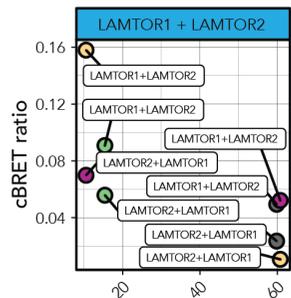
Based on homology PDB structure 4M7D
LSM2 (95/95 amino acids covered = 100%)
Lsm2 from *S. cerevisiae*
LSM3 (89/102 amino acids covered = 87%)
Lsm3 from *S. cerevisiae*



Based on PDB structure 5LSJ
MIS12 (205/205 amino acids covered = 100%)
PMF1 (174/205 amino acids covered = 85%)



Based on PDB structure 6EHR
RRAGA (130/313 amino acids covered = 42%)
RRAGC (160/399 amino acids covered = 40%)



distance [Å]

945 **Supplementary Figure 10 (related to Figure 4). Subset of interactions where at least six out of the eight**
946 **tested tagging configurations scored positive in the LuThy-BRET assay.**

947 The following 3D structures are represented: 1G5J (BAD-BCL2L1), 1BUH (CDK2-CKS1B), 1JFI (DR1-DRAP1),
948 1F3U (GTF2F1-GTF2F2), 1A02 (JUNB-BATF), 6EHR (LAMTOR1-LAMTOR2 and LAMTOR3-LAMTOR2), 4M7D
949 (LSM3-LSM2), 5LSJ (MIS12-PMF1), and 6EHR (RRAGA-RRAGC). On each structure, the protein termini are
950 indicated by N or C. Protein sequence coverages used in the structures are also indicated (%). Protein regions that
951 are missing in the structures are indicated by dotted lines. For all interactions, cBRET ratios are plotted against the
952 molecular distances measured in the corresponding 3D structures or interpolated using the cBRET standard curve.
953 The quantitative scores outlined in black correspond to tagging configurations where the tagged protein termini are
954 resolved in the structures. These scores were used as reference data points for distance interpolations. Interaction
955 scores and structure-based molecular distances for tagging configurations where the tagged protein termini are
956 absent from the 3D structure are outlined in red. Interaction scores outlined in purple correspond to cBRET ratios
957 against interpolated molecular distances. Changes in structurally measured distances versus the interpolated
958 distances are indicated by dotted arrows. The purple horizontal error bars correspond to the 95% confidence
959 intervals of the interpolated molecular distances. Tagging configurations are color-coded: N-N (grey), C-C (green),
960 N-C (purple), and C-N (yellow). Each data point on the graph is labeled (blank frame) according to the associated
961 tagging configuration tested: the protein indicated first is tagged with NanoLuc (NL) luciferase, while the second
962 protein is tagged with PA-mCitrine (PA-mCit) (e.g. DR1-DRAP1 (N-N, grey) corresponds to NL-DR1/PA-mCit-
963 DRAP1). Tagging configurations where tags are fused to the missing termini in the current 3D structure for one of
964 the two proteins are outlined in red (e.g. DR1-DRAP1 (N-C, purple) corresponds to NL-DR1/DRAP1-mCit-PA). The
965 size of the dots indicates the average percentage of protein sequence coverage in the corresponding 3D structure
966 of the PPI.

967 **APPENDIX**

968 **Supplementary Table 1:** List of the multiprotein complexes meeting the prioritization criteria in this study.

969 **Supplementary Table 2:** Number of inter-complex pairs, direct interactions and indirect associations for the three
970 selected protein complexes: LAMTOR, BRISC and MIS12 as defined by their 3D structures. The numbers of
971 separate homodimer interactions not reported in the structures are also indicated.

972 **Supplementary Table 3:**

973 **Sheet 1 - training sets:** Positive and negative training sets for LuTHy cluster analysis.

974 **Sheet 2 - clustering results:** Supervised classification of direct and not-direct interaction clusters.

975 **Supplementary Table 4:** Surface and interface areas (\AA^2) between directly interacting proteins in the LAMTOR,
976 BRISC and MIS12 complexes.

977 **Supplementary Table 5:** Names of protein complexes, protein subunits, and PDB IDs used in this study. The
978 amino acid sequences of the proteins used for binary PPI mapping or in the PDB structures are indicated.

979 **Supplementary Table 6:** Characterization of protein family (pfam) domains.

980 **Supplementary Table 7:** Structure-based and BRET-interpolated distances for the HTT-HAP40 interaction, and
981 for the intramolecular HTT sensors for HTTQ23 and HTTQ145.

982 **Supplementary Table 8:** Primer sequences used in this study.

983 **Supplementary Table 9:** Statistical reports for relevant figures.

984

985 **Source Data Figures 1-2:** Raw LuTHy data for interactions in hsPRS-v2.

986 **Source Data Figures 3-5:**

987 **Sheet 1 - Multi-protein complexes:** Raw LuTHy and mN2H mapping data for the selected multiprotein
988 complexes. Molecular distances measured between structurally supported interactions between
989 protein pairs within the LAMTOR, BRISC and MIS12 complexes.

990 **Sheet 2 - hsPRS-PDB:** Molecular distances measured between structurally supported heterodimer
991 interactions of the hsPRS-v2 (hsPRS-PDB).