BFG-PCA: tools and resources that expand the potential for binary protein interaction discovery

Daniel Evans-Yamamoto[1-4], François D. Rouleau[1,2,5,6], Piyush Nanda[2,Æ], Koji Makanae[2], Yin Liu[2,‡], Philippe C. Després[1,2,5,6], Hitoshi Matsuo[2,œ], Motoaki Seki[2,§], Alexandre K. Dube[1,5-7], Diana Ascencio[1,5-7], Nozomu Yachie[2-4,‡,¶] & Christian R. Landry[1,5-7,¶]

[1] Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC, Canada

[2] Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan

[3] Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Japan

[4] Institute for Advanced Biosciences, Keio University, Fujisawa, Japan

[5] Regroupement Québécois de Recherche sur la Fonction, l'Ingénierie et les Applications des Protéines, (PROTEO), Université Laval, Québec, QC, Canada

[6] Département de biochimie, microbiologie et bio-informatique, Université Laval, Québec, QC, Canada

[7] Département de Biologie, Université Laval, Québec, QC, Canada

[Æ] Current affiliation: Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, United States of America

[‡] Current affiliation: School of Biomedical Engineering, University of British Columbia, Vancouver, Canada

[œ] Current affiliation: Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan

[§] Current affiliation: Department of Molecular Oncology, Graduate School of Medicine, Chiba University, Chiba, Japan

[¶] Correspondence should be addressed to N.Y. (nozomu.yachie@ubc.ca) and C.R.L. (christian.landry@bio.ulaval.ca)

Abstract

Barcode fusion genetics (BFG) utilizes deep sequencing to improve the throughput of protein-protein interaction (PPI) screening in pools. BFG has been implemented in Yeast two-hybrid (Y2H) screens (BFG-Y2H). While Y2H requires test protein pairs to localize in the nucleus for reporter reconstruction, Dihydrofolate Reductase Protein-Fragment Complementation Assay (DHFR-PCA) allows proteins to localize in broader subcellular contexts and proves to be largely orthogonal to Y2H. Here, we implemented BFG to DHFR-PCA (BFG-PCA). This plasmid-based system can leverage ORF collections across model organisms to perform comparative analysis, unlike the original DHFR-PCA that requires yeast genomic integration. The scalability and quality of BFG-PCA were demonstrated by screening human and yeast interactions of >11,000 protein pairs. BFG-PCA showed high-sensitivity and high-specificity for capturing known interactions for both species. BFG-Y2H and BFG-PCA capture distinct sets of PPIs, which can partially be explained based on the domain orientation of the reporter tags. BFG-PCA is a high-throughput protein interaction technology to interrogate binary PPIs that exploits clone collections from any species of interest, expanding the scope of PPI assays.

Running title

Pooled matrix PPI screening with BFG-PCA

Keywords

Barcode Fusion Genetics/DHFR-PCA/DNA barcodes/Protein-protein interaction

Introduction

In cellular systems, proteins form functional modules and/or complexes that underlie most biological processes by physically interacting with each other (Alberts 1998; Vidal, Cusick, and Barabási 2011). Discovering such interaction networks is one of the main goals of systems biology. Two major approaches to detect protein-protein interactions (PPIs) have contributed the bulk of the current data, affinity purification followed by mass spectrometry (AP/MS), and methods such as Yeast two-hybrid (Y2H) and protein complementation assay (PCA). The former approach detects biomolecular association among group of proteins from cellular fractions (Rigaut et al. 1999; Ren et al. 2003; Roux et al. 2012; Gillet et al. 2012), whereas the latter detect direct “binary” or pairwise PPIs, by tagging each interaction partner, the bait and the prey, using reporter protein fragments (Fields and Song 1989; Hu, Chinenov, and Kerppola 2002; Tarassov et al. 2008). Other approaches such as proximity-dependent biotinylation *in vivo* (Roux et al. 2012; Rhee et al. 2013; Go et al. 2021), co-elution and co-fractionation (Salas et al. 2020)(Kristensen, Gsponer, and Foster 2012) and protein-cross linking (C. Yu and Huang 2018; Bruce 2012; Sinz 2010) also contribute to the dissection of PPI networks, with varying degrees of resolution.

Binary interaction screenings are powerful approaches owing to their relatively simple implementation in terms of instrumentation. Up to now, systematic high-quality Y2H screening (Venkatesan et al. 2009) has revealed the largest binary interactome network to date, covering the entire human and yeast proteomes (Luck et al. 2020). Because of their scalability, Y2H has also been applied to a large number of model organisms, including for instance Arabidopsis and Drosophila (Arabidopsis Interactome Mapping Consortium 2011; Li et al. 2004; Rajagopala et al. 2014; H. Yu et al. 2008). Despite such efforts, we are far from a complete interactome map when considering various “proteoforms” (Smith, Kelleher, and Consortium for Top Down Proteomics 2013; Aebersold et al. 2018), disease mutations (Sahni et al. 2013) and protein polymorphisms that can have distinct biophysical interaction profiles (Corominas et al. 2014).

One of the limiting factors associated with binary detection methods is the need to perform pairwise tests between baits and preys in a comprehensive manner because these pairwise tests are dominated by negative results. However, the application of next generation sequencing (NGS) has played a key role in increasing throughput, and thus, interactome coverage (Yachie et al. 2016; Trigg et al. 2017; Schlecht et al. 2017, 2012; J.-S. Yang et al. 2018; F. Yang et al. 2018). Combined with methods that involve cell survival as detection signals, NGS facilitates the exploration of the search space of PPIs because of the enrichment

of positive PPIs. One of the first studies that have implemented a pooled screening approach using NGS, Stitch-seq, allowed the identification of interacting baits and preys in pooled assays through fusion PCR of bait and prey ORFs after selecting for interacting pairs (Rolland et al. 2014; H. Yu et al. 2011). Several other approaches exploiting this principle have been implemented. They include Barcode Fusion Genetics (BFG) (Yachie et al. 2016), PPi-seq (Schlecht et al. 2017; Liu et al. 2020), which use synthetic DNA barcodes to tag their gene of interests and CrY2H-seq (Trigg et al. 2017), rec-YnH (J.-S. Yang et al. 2018), and RLL-Y2H (F. Yang et al. 2018), which use the Open Reading Frame (ORF) sequences itself to identify the protein pairs. Using ORF sequences as identifiers offers simplicity to the design but DNA barcodes may be more reliable in terms of accuracy and performance, and may reduce sequencing costs, although they may require more investment upstream of the screens.

BFG was recently adapted to Y2H screening. In BFG experiments, bait and prey plasmids contain DNA barcodes that are fused through intra-cellular recombination in cells that survive selection on a specific media. Sequencing of the fused barcodes allows the identification of the interacting pairs in bulk (**Fig 1A**). Because the barcode fusion technology is portable to other approaches in yeast genetics, it could be used to adapt other binary mapping methods to pooled screening and thus enable a better coverage of PPI networks. Indeed, different assays have little but significant overlap of positive interactions, and it is important to assay PPIs with multiple orthogonal assays to comprehensively map interactomes (Braun et al. 2009; Choi et al. 2019). For instance, systematic benchmarking of various complementary assays in yeast and human cells has reported that each method captures only ~35% of the confident positive reference PPI set (HsPRSv1) (Braun et al. 2009). Even more revealing for this study, for binary assays, the currently reported *S. cerevisiae* PPIs by Y2H and PCA share only 525 unique interactions (Y2H: 12,995; PCA: 6,739; Union: 19,209; Jaccard Index: 0.027), despite each method having proteome wide PPI mapping efforts of similar quality (H. Yu et al. 2008; Tarassov et al. 2008; Stark et al. 2006). There are many reasons why different methods cover different parts of PPI networks. For instance, reporter proteins or protein fragments are fused at either the N or C termini and some may require the localization of proteins to specific cell compartments (Buntru et al. 2016) **(Figs 1B and 1C)**.

BFG enables pooled matrix screening that exploits various selection markers affecting growth (Yachie et al. 2016). DHFR-PCA is a binary PPI detection method based on growth via the reconstitution of a mutant DHFR in yeast cells, which provides resistance to the drug methotrexate (Tarassov et al. 2008). Contrary to Y2H, DHFR-PCA does not require the addition

of a nuclear localization for reporter activation, and in principle enables PPI detection in the protein pair's native subcellular context. Until now, efforts to map PPIs by DHFR-PCA have focused on interactions present *in vivo* by tagging DHFR fragments to genomic loci, even when barcodes are used for pooled based assays (Schlecht et al. 2017, 2012). Although this comes with many advantages, it also comes with limitations depending on the questions being addressed. For instance, protein expression levels are largely regulated by the environment, making interactions of weakly expressed/unexpressed proteins difficult to detect in some conditions. Having bait and prey proteins expressed from plasmids could help alleviate this limitation. Controlling or uniformalizing expression level may help differentiate transcriptional versus post transcriptional effects on PPIs in experiments comparing different growth conditions (Celaj et al. 2017; Schlecht et al. 2017). Another advantage of plasmid-based screening is that it allows for screening PPIs for protein variants, or among proteins from other species or between species, provided the coding sequences can be cloned and expressed in yeast. Here, we developed and made publicly available affordable resources for BFG-DHFR-PCA (henceforth BFG-PCA) and we used this resource to demonstrate the efficacy of BFG-PCA by screening 11,232 bait-prey pairs (**Fig 1D**). We show that BFG-PCA enables the detection of *in vivo* PPIs and the comparison side-by-side with BFG-Y2H demonstrates that they capture distinct sets of PPIs.

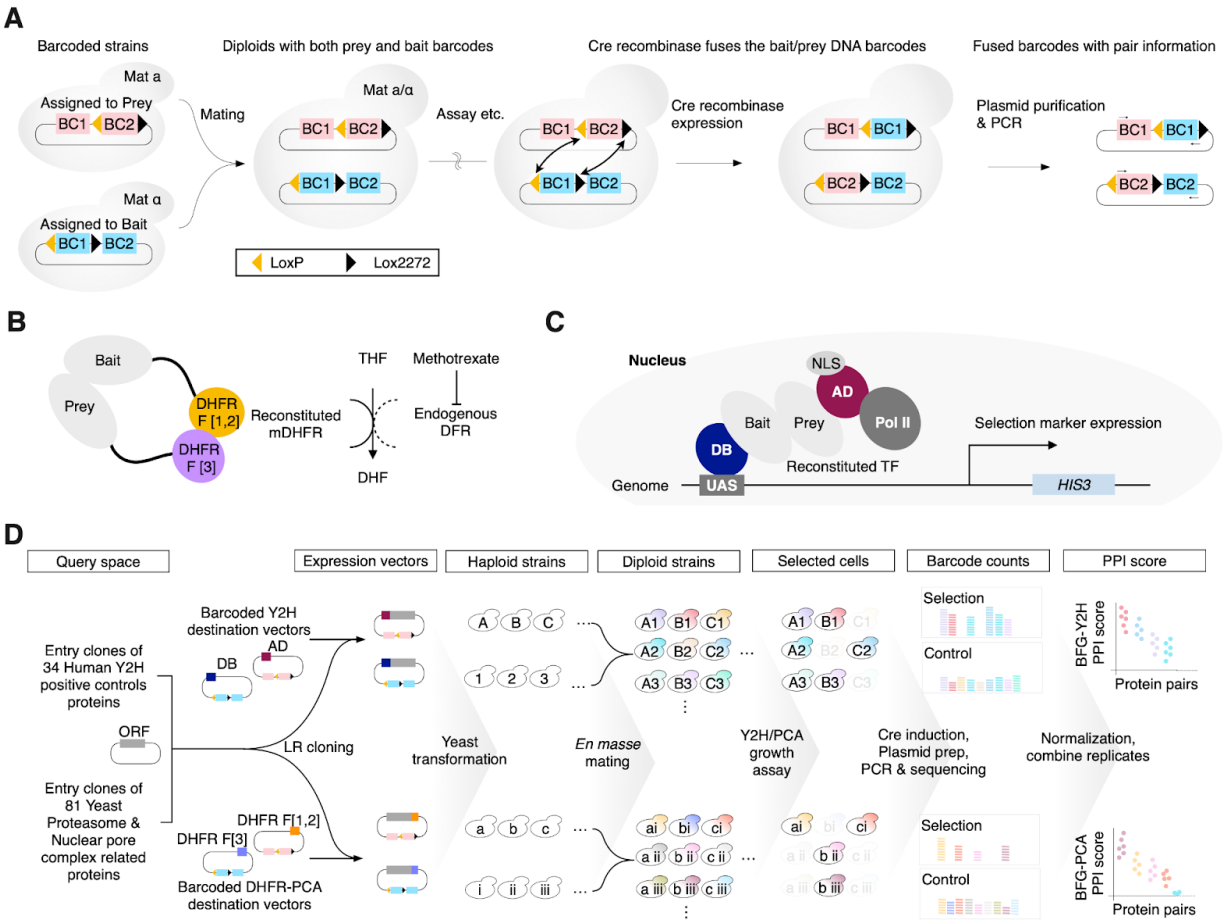

Figure 1. Overview of this study.

A. Schematic of Barcode Fusion Genetics (Yachie et al. 2016). BFG barcode cassettes are assigned to each of the bait and prey with 2 DNA barcodes (BC1 and BC2), with one of them flanked by Cre recombination site LoxP and Lox2272. The bait and prey plasmids with BFG barcode cassettes are introduced to MATα and MATa cells, respectively. Upon mating, the diploid cells have both the bait and prey plasmids and their barcode cassettes. By inducing expression of the Cre recombinase, the $BC1_{bait}$ and $BC2_{prey}$ are swapped between the plasmids, resulting in barcode fusions $BC1_{prey}$-$BC1_{bait}$ and $BC2_{bait}$-$BC2_{prey}$, having the bait-prey pair information. Each of the bait and prey barcodes have a common primer site flanking the barcode region unique to the type of barcode, enabling specific PCR amplification of the $BC1_{prey}$-$BC1_{bait}$ and $BC2_{prey}$-$BC2_{bait}$ products. By counting the number of barcode pairs by deep sequencing, one can estimate the relative abundance of diploids in the pool.

B. Illustration of the DHFR-PCA reporter. In DHFR-PCA, DHFR F[1,2] and DHFR F[3] are fused to bait and prey proteins, respectively. Upon interaction, the DHFR fragments come in proximity, reconstituting the methotrexate-resistant murine DHFR enzyme (mDHFR) while the conditionally essential endogenous DHFR is inhibited by the drug methotrexate (MTX).

C. Illustration of the Y2H reporter. In Y2H, the DNA binding domain (DB) and the activator domain (AD) of the Gal4 transcription factor (TF) are fused to bait and prey proteins,

respectively. The fused proteins are localized to the nucleus by the nuclear localization signal (NLS). The DB domain will bind to the upstream activation sequence (UAS). When the bait and prey proteins interact, the Gal4TF is reconstituted, recruiting RNA polymerase II, expressing the selection marker of choice. We used the *HIS3* marker with medium lacking histidine for Y2H selection throughout this study.

D. Overview of the BFG screening. We queried 115 proteins from Human and Yeast, and Gateway cloned them to barcoded Y2H and DHFR-PCA destination vectors with 2 barcode replicates each. We individually transformed haploid strains with the barcoded expression vectors. The haploid strains were pooled for *en masse* mating, generating all possible bait-prey pairs of diploids. After selecting diploid cells, we performed pooled selection for each method. After selection, we induced Cre expression for barcode fusion, purified the plasmids, and PCR amplified the barcodes for illumina sequencing. We counted barcodes, normalized them by the barcode counts in the control condition and background auto-activity of the strains. The replicates were combined for each protein pair, generating the final PPI score for each method to call PPIs.

Results and discussion

Adapting DHFR-PCA for plasmid-based PPI detection

Gateway cloning compatible destination vectors and yeast strains were generated for BFG-PCA and are available through Addgene. We constructed a collection of 1,483 centromeric Gateway compatible plasmids with unique barcodes (893 for DHFR F[1,2] and 590 for DHFR F[3]), enabling assays of up to 526,870 protein pairs in pools using barcode fusion and sequencing (**Table EV1**). The functionality of these plasmids for DHFR-PCA and strains was first examined by performing a growth assay on 8 protein pairs consisting of 4 known PPI and 4 pairs not reported to interact. As expected, all pairs showed similar growth on the non-selective (−methotrexate) condition and pairs with reported interactions showed growth on the selective (+methotrexate) condition (**Fig EV1A**).

To further examine the performance of plasmid based DHFR-PCA, we performed a DHFR-PCA assay on 300 protein pairs using the established protocol on solid media (Rochette et al. 2015) (**Fig EV1B and C**) with these plasmids. The selected space included DHFR-PCA expected positives and likely negatives for quality assessment (**Fig EV1B**). Likely negative pairs were selected based on the BioGRID database (version 3.4.157) (Stark et al. 2006), with criteria including 1) no reported physical or genetic interaction for the yeast proteins or their orthologs in *Schizosaccharomyces pombe* and *Homo sapiens,* 2) no shared gene ontology terms, and 3) a distance greater than 2 edges in the PPI network. The PPI score of each pair was calculated based on colony sizes estimated from plate images (**Fig EV1C**), and sorted to examine agreement with known PPIs (**Fig EV1D**). Protein pairs with reported interactions were enriched for high PPI score pairs. We evaluated this by Mathew's Correlation Coefficient (MCC), giving a value of 0.462, comparable to reported PPIs in BioGRID with either Y2H (MCC = 0.488) or PCA (MCC = 0.403). The raw PPI score and details of each protein pair are shown in Tables EV1 and EV2.

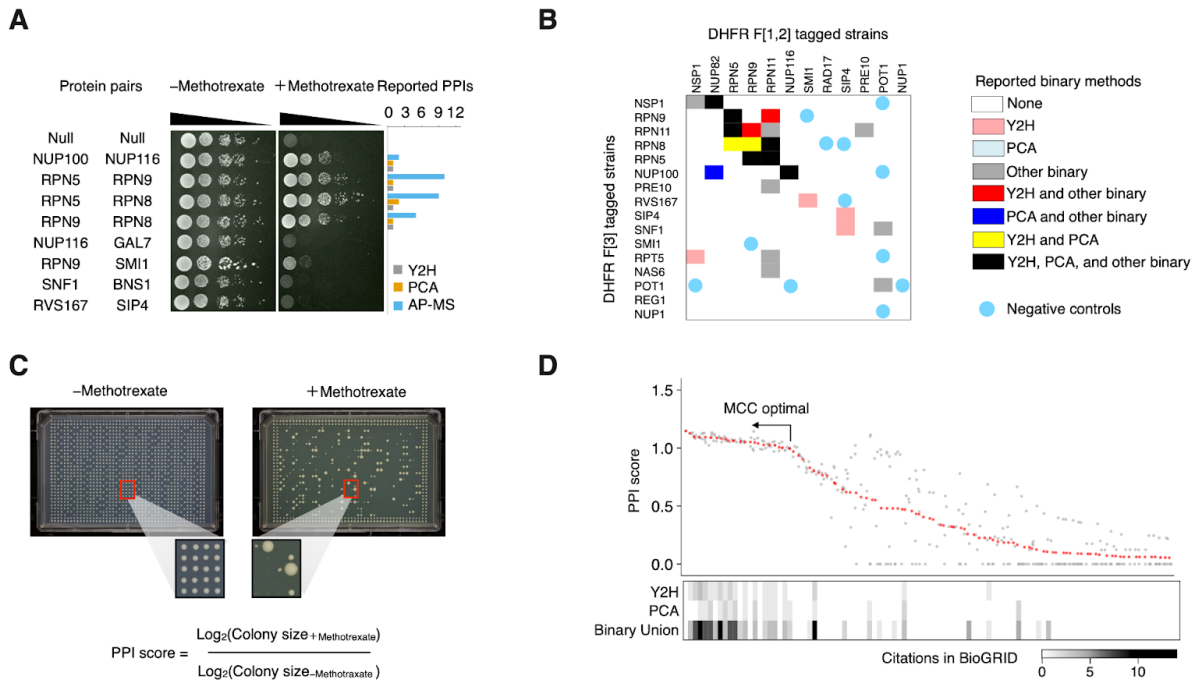

Figure EV1. Plasmid-based DHFR-PCA captures known PPIs.

A. Plasmid based DHFR-PCA spot assays were performed on 8 protein pairs. Null-Null represents empty vector (destination plasmid) control. Reported PPIs in BioGRID are shown in the barplot. While all pairs grow under no methotrexate control (−Methotrexate), only protein pairs expected to have interactions show growth with presence of methotrexate (+Methotrexate).

B. A subset of the query space to demonstrate the quality of plasmid-based DHFR-PCA, which was also tested by BFG-PCA (see following section). Previously known interactions are indicated in colors according to the method.

C. An example of a DHFR-PCA high density plate. The colony formed from replicating the same cell sample is grown on control (−Methotrexate) and selection (+Methotrexate) plates. Colony size is measured based on plate images, log-transformed and used to calculate PPI scores by fold-change between selection and control.

D. Result of an assay on 300 protein pairs ordered by PPI score rank. Only the top 100 protein pairs are shown. Grey dots represent replicates, and the red dot represents the 50th percentile threshold used to call the ranks. The heatmap shows previously reported interactions in the BioGRID database. Binary union consists of interactions reported by Y2H, PCA, Biochemical activity, Affinity Capture-Luminescence, Reconstituted Complex, Co-crystal Structure, and FRET.

BFG-PCA screening condition optimization on 192 protein pairs

We performed a proof-of-concept BFG-PCA screening on a subset of 192 protein pairs assayed by plasmid based DHFR-PCA (**Figs EV1B, 2, DatasetEV1**), with the exception of those dropped out during barcoding and cloning. Previous DHFR-PCA conditions used a methotrexate concentration of 200 µg/mL (Tarassov et al. 2008). However, Yachie et al (Yachie et al. 2016) have shown that BFG-Y2H performs better when selecting under less stringent conditions than standard Y2H (Yachie et al. 2016). Therefore, four concentrations (200, 100, 10 and 1 µg/mL) of methotrexate were tested to examine the optimal concentration for BFG-PCA (see **Fig 1D** for selection step). As expected, higher concentrations resulted in fewer and smaller colonies (**Appendix Fig S2A**). Deep sequencing confirmed that the signal-to-noise ratio (SNR) increases with increasing concentration (**Appendix Figs S2B and 2A**). We compared the standard DHFR-PCA (based on colony growth) with the BFG-PCA scores (computed based on fused barcode counts). As expected, there was no relationship between the colony-based signal and BFG-PCA signal for non-interacting pairs (low colony-based signal) but a strong one above a given threshold, which corresponds to expected positive PPIs. This led to an overall positive rank correlation for all tested BFG-PCA selection conditions (Kendall rank coefficient : 1µg/mL, 72 h= 0.141; 96 h = 0.211; 10µg/mL, 72 h= 0.228; 96 h = 0.252; 100µg/mL, 72 h = 0.287; 96 h = 0.243; 200µg/mL, 72 h= 0.309; 96 h = 0.270; $p<0.01$, **Figs 2B,G**). The increasing correlation at higher concentrations of methotrexate and longer incubation periods contributed to higher SNRs (**Fig 2F,H**).

Among the BFG-PCA conditions tested, 10 µg/mL of methotrexate and 72 hours of selection yielded the best agreement with reported binary PPIs (**Figs 2C and 2D**), with a MCC of 0.61. One exception is an interaction between Pre10 and Rpt5 within the 26S proteasome complex, which had not been reported previously by any binary PPI detection method. The two proteins are neighbouring within the complex when mapped to the crystal structure (**Fig 2E**), suggesting this is a true positive interaction that has been missed from previous experiments. These conditions therefore appear to be optimal among the ones tested, for BFG-PCA screenings.

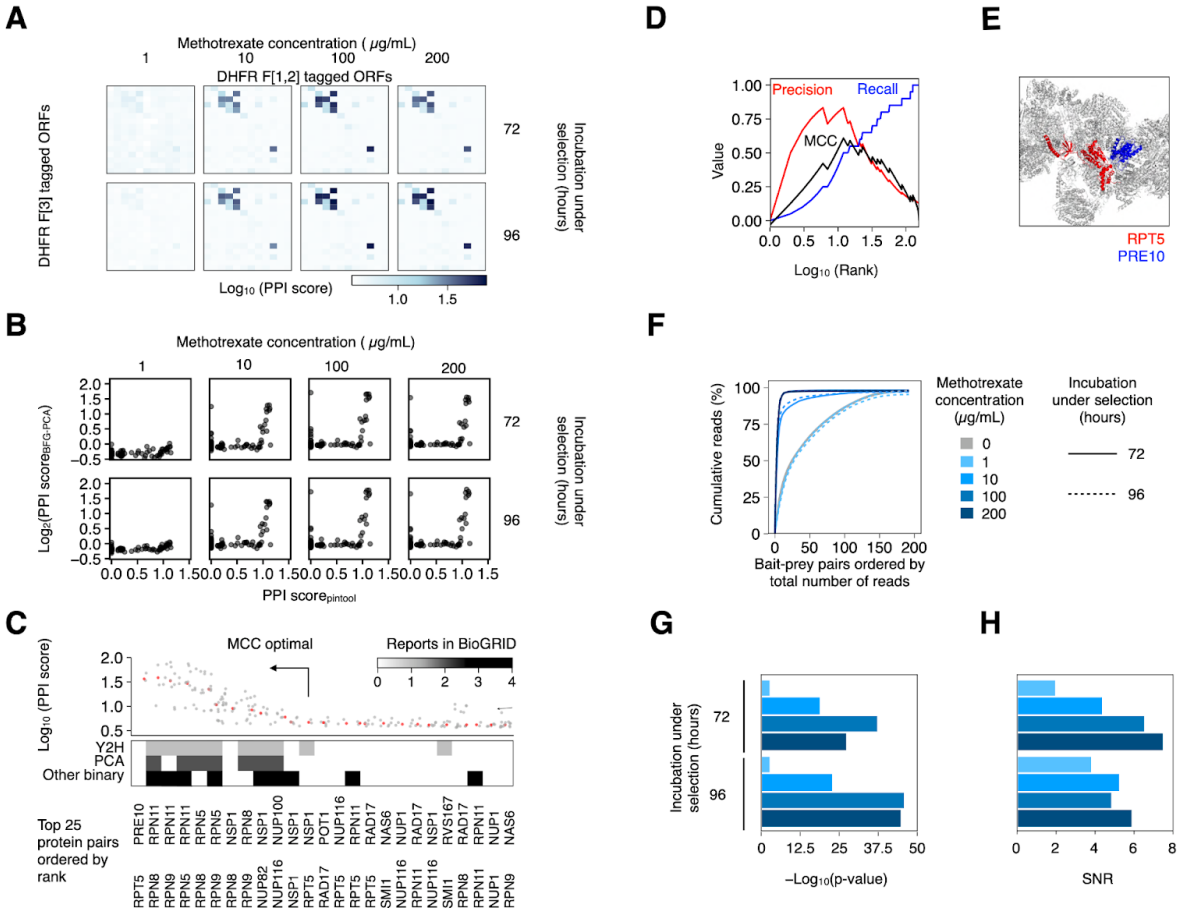

Figure 2. BFG-PCA screening quality is on par with one-by-one assay.

A. Heatmap of the BFG-PCA PPI score for each of the selection conditions. Baits and preys are ordered as in the query matrix shown in Figure EV1B.

B. PPI scores obtained from one-by-one DHFR-PCA high density plate assay and BFG-PCA.

C. PPI scores ordered by rank. Methotrexate concentration of 10 μg/mL with 72 h selection is shown. Grey dots represent replicates, and the red dots represent the 50th quantile of replicates used to call the rank. The heatmap represents reported interactions in the BioGRID database for Y2H, PCA, and other binary PPI detection methods (Biochemical activity, Affinity Capture-Luminescence, Reconstituted Complex, Co-crystal Structure, and FRET).

D. Precision/recall curve of the BFG-PCA data with methotrexate concentration of 10 μg/mL with 72 h of selection.

E. The top rank Pre10 and Rpt5, which had no previous binary interaction reported, are highlighted on the crystal structure of the yeast 26S proteasome (PDB: 6J2X).

F. Cumulative plot of raw barcode counts per protein pair under each selection condition, showing the number of protein pairs represented after sequencing.

G. & H. The Kendall rank correlation coefficient (G), and signal to noise ratio (H) for each BFG-PCA condition against one-by-one DHFR-PCA. To compute the signal to noise ratio, the PPI scores of 12 negative control pairs and the top 10 ranked scores were averaged and used as background and signal, respectively.

BFG-PCA and BFG-Y2H screening on the Proteasome and Nuclear pore complex related proteins

Since BFG has only been implemented for Y2H, and Y2H is the most frequently used method for binary PPI screening, we compared BFG-PCA and BFG-Y2H side-by-side. We examined a space consisting of 120 proteins (34 human proteins as Y2H positive controls previously used in BFG-Y2H, 16 *S. cerevisiae* proteins used for the first demonstration of BFG-PCA, and 80 *S. cerevisiae* proteins associated with the proteasome complex and/or nuclear pore complex) with 2 barcode replicates. We performed two screenings for both BFG-PCA and BFG-Y2H (**Fig 1D**), each covering 11,232 and 10,545 bait-prey pairs, respectively. The number of barcode replicates per ORF detected in each screening were mostly 2, with some having only 1 due to loss during the cloning process (**Appendix Fig S3A**). The distribution of bait/prey barcode abundance in the non-selective conditions, representing relative abundance of haploid strains before mating, followed a log-normal distribution in each screening as expected (**Appendix Fig S3B**). Similarly, the relative abundance of bait-prey barcodes, representing diploid strains, followed a log-normal distribution of barcodes in the non-selective conditions (**Appendix Fig S4**).

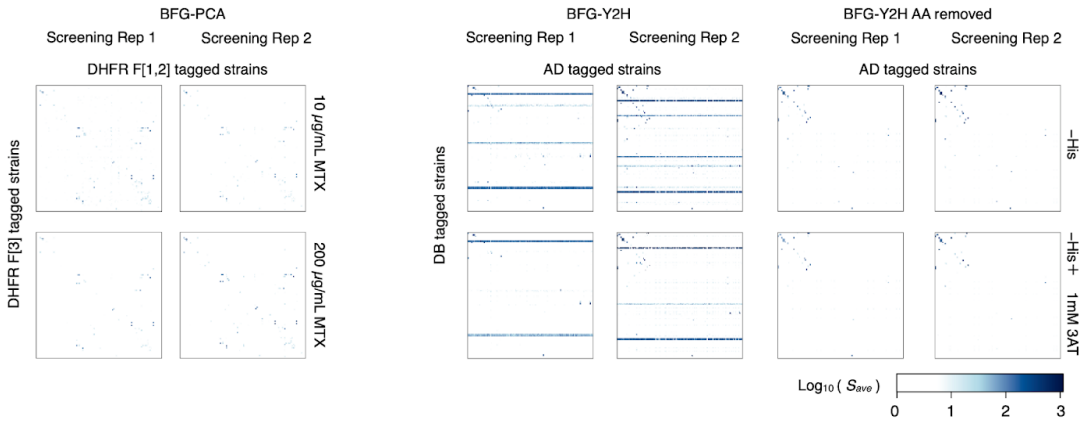

Figure EV2. Enrichment of bait-prey barcodes in selective conditions compared to their respective non-selection conditions.

Heatmap of enrichment signal (*s*) shown for each bait-prey barcode in BFG-PCA and BFG-Y2H screenings. Note that the barcode fusion replicates (BC1-BC1 and BC2-BC2) were averaged for each bait-prey barcode. AA: Auto-activator.

We computed the enrichment score '*s*' as growth enrichment of each bait-prey barcode in selective conditions compared to non-selective conditions. Under selection for both BFG-PCA and BFG-Y2H, some of the bait/prey barcodes exhibited strong background noise (**Fig EV2,**

Appendix Fig S5, and **DataEV2**). This is a commonly known phenomenon for DB strains in Y2H where some proteins directly recruit the transcription machinery without the presence of an interaction partner (Dreze et al. 2010) (**Fig 1B**). While BFG-Y2H involves a normalization for the auto-activity of problematic baits, we observed that several bait ORFs occupied >68% of all reads sequenced from selective condition libraries, which is wasted sequencing effort. Therefore, we performed an additional screening in duplicate with these strains removed for better assessment between BFG-PCA and BFG-Y2H. After removal of BFG-Y2H auto-activators, BFG-PCA and BFG-Y2H each covered 11,232 and 9,546 bait-prey pairs, respectively. When examining BFG-PCA signal data, we observed a similar but less intense auto-activity background (**Appendix Fig S5**) which we systematically normalized when computing PPI scores. This is a known phenomenon, where some proteins interact with the DHFR fragment or the linker alone, contributing to systematic background noise (Tarassov et al. 2008). For implementation of either BFG-PCA or BFG-Y2H examining new baits and preys, it may therefore be necessary to first screen for auto-activity.

The enrichment signals within replicates, both internal and screening replicates, correlated strongly in each method (**Appendix Fig S6**), demonstrating their reproducibility. For each pair, we performed normalization of the enrichment scores based on auto-activity backgrounds (**Appendix Fig S7A,** see methods) to obtain PPI scores. For each of the 2 screening replicates performed, each protein pair had multiple levels of internal replicates corresponding to tagging orientation, barcoded strains, and barcode fusions (**Appendix Fig S6A**). When combining all the screening replicates and internal replicates, the average number of replicates for each protein pair was 23.8 and 21.6 for BFG-PCA and BFG-Y2H, respectively. Both screenings had over 99% of the protein pairs with more than 8 replicates (**Appendix Fig S6**). To call positives, we examined the best scoring method by testing average, median, and various percentile thresholds amongst the normalized score of replicates and by computing the best agreement against reported binary interactions in BioGRID (**Appendix Fig S7B**).

As a result, we detected 92 (MCC = 0.315) and 35 (MCC = 0.296) PPIs for BFG-PCA and BFG-Y2H, respectively (**Fig 3, DatasetEV3**). Although the overlap between the two methods was limited, for known binary PPIs, the PPI scores are correlated (**Fig 3E**) (R = 0.12, p-value = 6.4×10^{-3}, Kendall rank correlation). Since we included both human and yeast proteins, we performed the same analysis on the Human protein and yeast protein subsets individually (**Figs**

EV3, EV4). On the Human protein subset, BFG-PCA and BFG-Y2H detected 15 (MCC = 0.462), and 34 (MCC = 0.619) PPIs, respectively (**Fig EV3**). The difference between BFG-PCA and BFG-Y2H can be explained by the fact that known Y2H positive pairs have been deliberately included in the space as positive control. In addition, no PPI data is available on human proteins screened by DHFR-PCA so we had no a priori expectation for the performance of BFG-PCA on these. On the yeast protein subset, BFG-PCA and BFG-Y2H detected 80 (MCC = 0.311) and 8 (MCC = 0.166) PPIs, respectively (**Fig EV4**). In conclusion, while BFG-Y2H had a higher overall ability for capturing the human PPIs, which was tailored as being positive controls for this method, BFG-PCA performed better when tested against a larger, non-tailored set of yeast protein pairs that are part of protein complexes, and had fewer issues of auto-activator proteins for this particular set of proteins.

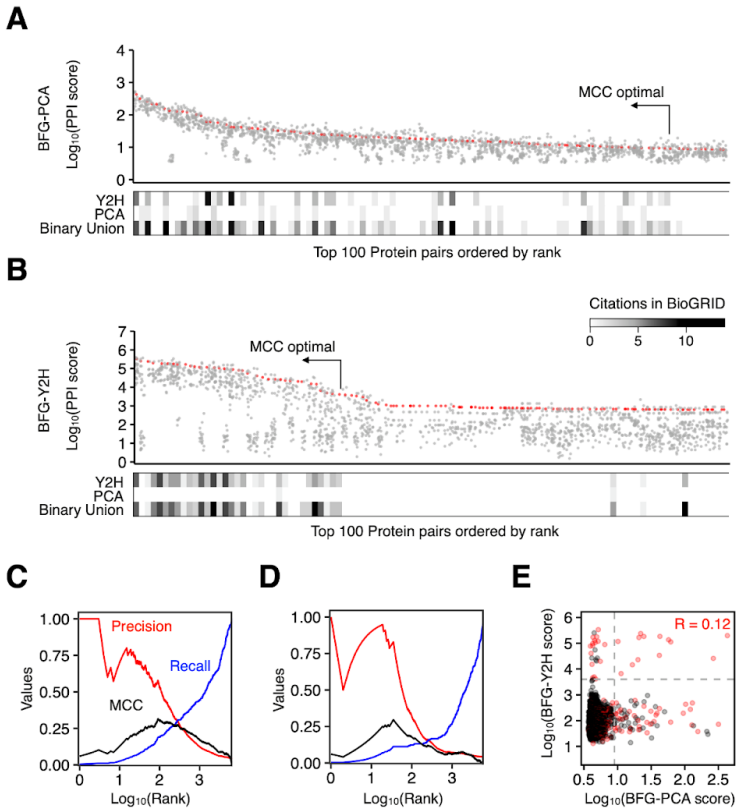

Figure 3. Quality of BFG-PCA and BFG-Y2H screenings.

A. & B. Results of assay ordered by PPI score rank for BFG-PCA (A), and BFG-Y2H (B). Grey dots represent replicates, and the red dots represent the percentile threshold used to call the ranks. The heatmap shows previously reported PPIs in the BioGRID database. Binary union consists of interactions reported by Y2H, PCA, Biochemical activity, Affinity Capture-Luminescence, Reconstituted Complex, Co-crystal Structure, and FRET.

C. & D. Precision, recall, and MCC values for BFG-PCA (C) and BFG-Y2H (D) with a given rank threshold.

E. Scatter plot between BFG-PCA and BFG-Y2H scores. Red represents PPIs in the BioGRID database reported by binary PPI detection methods. Grey dashed lines represent the threshold to call positives. R represents Kendall correlation coefficient for previously reported binary interactions.

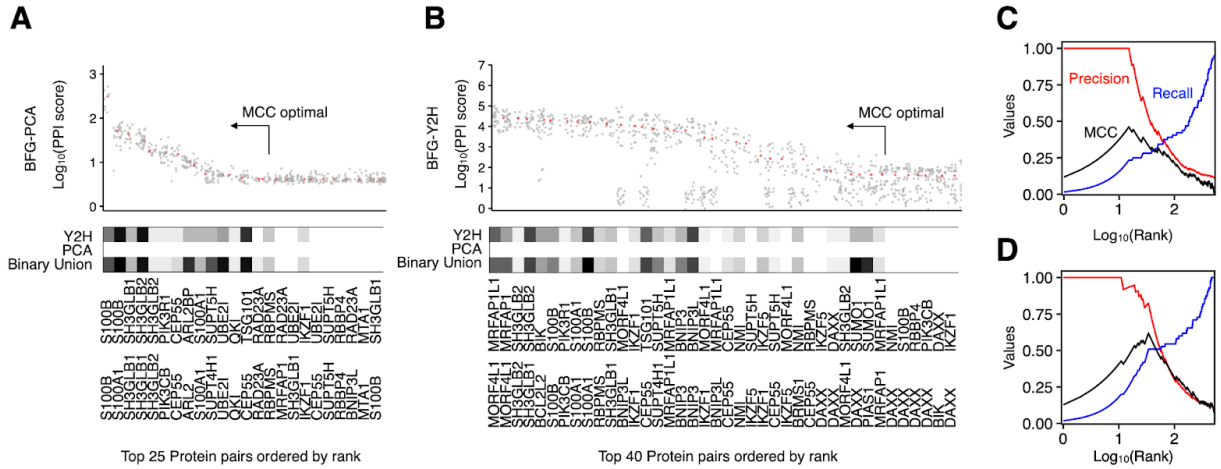

Figure EV3. Detected interactions in the Human protein subset.

A. & B. Result of assay ordered by PPI score rank on the Human proteins for BFG-PCA (A)/BFG-Y2H (B). Grey dots represent replicates, and the red dots represent replicates on the percentile threshold used to call the ranks. The heatmap shows previously reported interactions in the BioGRID database. Binary union consists of interactions reported by Y2H, PCA, Biochemical activity, Affinity Capture-Luminescence, Reconstituted Complex, Co-crystal Structure, and FRET.

C. & D. Precision, recall, and MCC curve for BFG-PCA (C) and BFG-Y2H (D) on Human protein pairs.

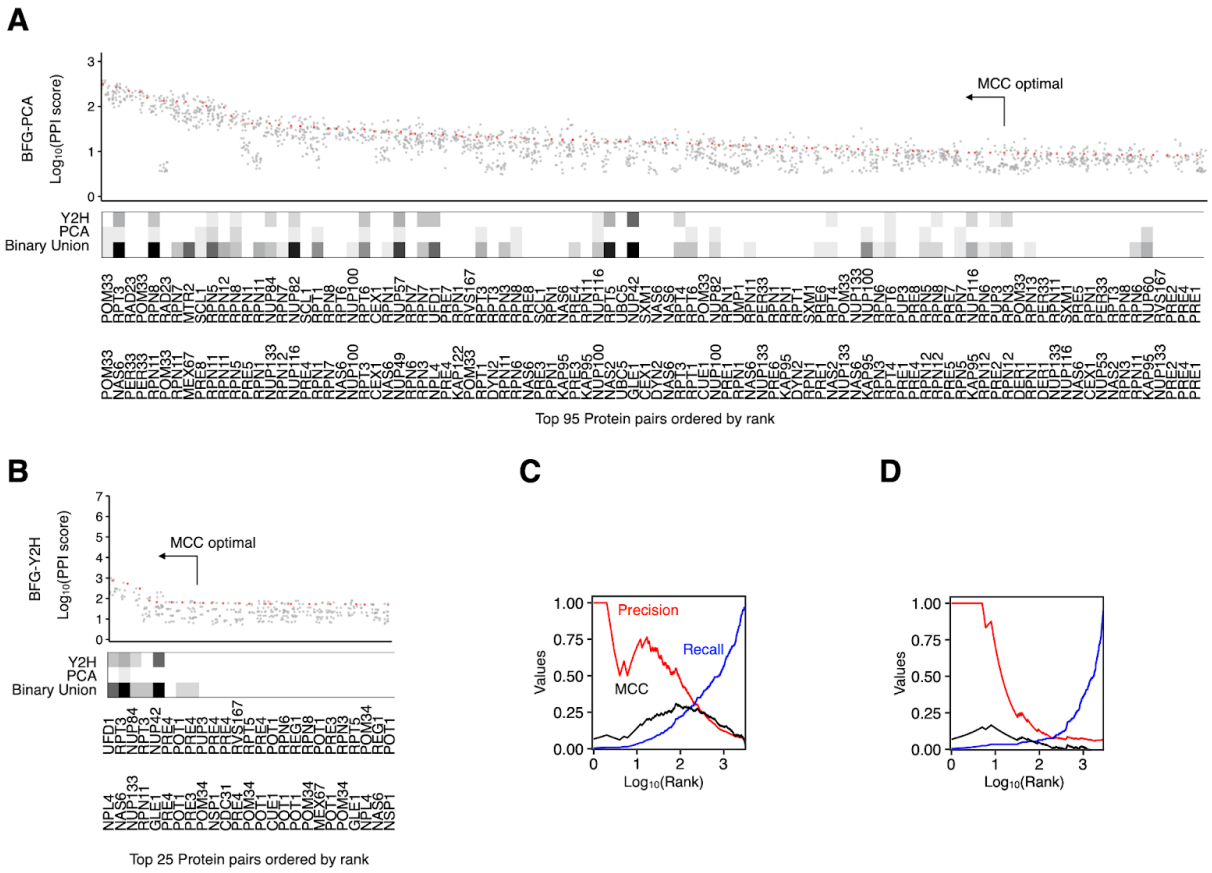

Figure EV4. Detected interactions in the Yeast protein subset.

A. & B. Result of assay ordered by PPI score rank on the Yeast proteins for BFG-PCA (A)/BFG-Y2H (B). Grey dots represent replicates, and the red dots represent replicates on the percentile threshold used to call the ranks. The heatmap shows previously reported interactions in the BioGRID database. Binary union consists of interactions reported by Y2H, PCA, Biochemical activity, Affinity Capture-Luminescence, Reconstituted Complex, Co-crystal Structure, and FRET.

C. & D. Precision, recall, and MCC curve for BFG-PCA (C) and BFG-Y2H (D) on Yeast protein pairs.

BFG-PCA captured binary PPIs in the 26S proteasome with high resolution

We further investigated whether the quantitative PPI score from BFG-PCA correlates with interaction strength. We hypothesized that since all proteins are expressed from the same promoter, variation in signal intensity may depend on how proteins interact rather than on their expression level. We calculated the solvation free energy gain (ΔG) between subunits in the three-dimensional protein interface of crystal structure of the yeast 26S proteasome (Ding et al. 2019) (PDB: 6J2X). We observed a strong negative correlation between BFG-PCA PPI score

and ΔG (R = −0.58, p-value = 1.44 ×10^{-8}, Pearson correlation) (**Fig 4A**). Among the protein pairs with BFG-PCA PPI scores above the threshold to call interaction, 6 were unreported using binary PPI detection methods. We mapped these protein pairs on the crystal structure, and found that the interactions called by BFG-PCA are indeed neighboring subunits of the 26S proteasome (**Fig 4B**). These results suggest the potential of BFG-PCA to capture binary PPIs within protein complexes with high precision.

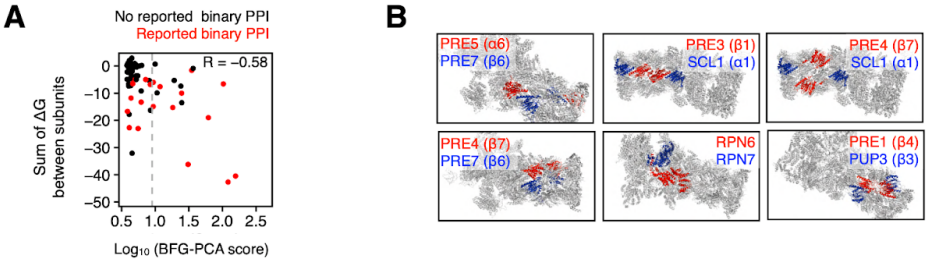

Figure 4. BFG-PCA PPI score agrees with ΔG between subunits within the yeast 26S proteasome.

A. Scatter plot of BFG-PCA score and solvation free energy gain (ΔG) upon formation of the interface between subunits. Grey dashed lines represent the threshold to call positives. Red represents PPIs in the BioGRID database reported by binary PPI detection methods. R represents Pearson correlation coefficient.

B. Detected positives with no previous binary PPI reports were mapped on the crystal structure (PDB:6J2X). Color for each indicated protein is shown within each image.

BFG-PCA compared to genomic integration-based DHFR-PCA

Compared to genome-based DHFR-PCA, our plasmid constructs differ in two key components. First, the standard DHFR-PCA detects PPIs among proteins under native expression levels, whereas plasmid-based DHFR-PCA and BFG-PCA express the gene under a constitutively active *ADH1* promoter. Second, while protein-linker-DHFR fusion in previous works generally used the linker sequence $(GGGS)_2$, our plasmid based linker sequence is NPAFLYKVVGGGSTS. To examine if these differences influence the detection of PPIs, we compared the interaction scores derived from DHFR-PCA with genomic integration from previous studies (Tarassov et al. 2008; Chrétien et al. 2018) with BFG-PCA. As expected from the results reported above, BFG-PCA detected a significant number of known binary interactions which were not captured by genome-based DHFR-PCA (**Appendix Fig S8A**). The expression levels of protein (Ho, Baryshnikova, and Brown 2018) with lower values within the pair (which may serve as a bottleneck for signal) was compared between each section of the scatter plot that define which PPI is detectable with each method (**Appendix Fig S8B**). The results showed that PPI negative pairs have significantly lower expression compared to that of positive pairs, which agrees with previous literature (Grigoriev 2001). Although no significant difference in protein expression was observed between BFG-PCA specific positives and Tarassov et al.'s positives, we noticed a case of lowly expressed proteins whose expression is detectable only by BFG-PCA, Gle1, interacting with Nup42. Gle1 and Nup42 are both subunits of the nuclear pore complex, and their PPI has been reported to interact by multiple methods in both low (Murphy and Wente 1996; Strahm 1999; Alber et al. 2007; Lin et al. 2018; Adams et al. 2018) and high-throughput (H. Yu et al. 2008), but has not been detected by genome-based DHFR-PCA. The higher expression level or the modified linker or both may allow a better detection of this PPI.

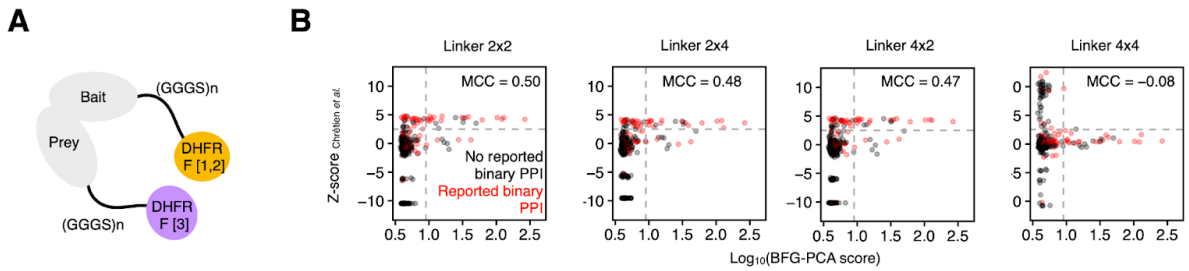

Figure EV5. BFG-PCA scores compared to DHFR-PCA with extended linkers.

A. Illustration of DHFR-PCA linker design. Chrétien et al. detected PPIs with linkers with $(GGGS)_2$ or $(GGGS)_4$ in both bait and prey orientations.

B. Scatter plot of BFG-PCA score and Z-score obtained in Chrétien et al. Grey dashed lines represent the threshold to call positives. Red represents PPIs in the BioGRID database reported by binary PPI detection methods.

We investigated whether variation in linker length or composition of polar amino acids affected the detected PPIs, we compared the BFG-PCA PPI scores to a previous effort of extending the linker sequence to $(GGGS)_4$ in genome-based DHFR-PCA (Chrétien et al. 2018), **Fig EV5A**). Out of the four combinations of linkers they have tested, we observed that the scores obtained by standard $(GGGS)_2$ for both DHFR F[1,2] and DHFR F[3] had the best agreement with BFG-PCA positives (MCC = 0.50), where having an extended linker for either DHFR F[1,2] and DHFR F[3] slightly decreases the agreement (MCC = 0.48 and 0.47), and having both linkers extended drastically decreases the agreement (MCC = −0.08, **Fig EV5B**).

Tagging orientation of DHFR fragments modifies the space of detectable PPIs

It was previously reported that the DHFR fragment position of fusion protein influences detection ability (MacDonald et al. 2006; Choi et al. 2019). In the context of the proteasome for instance, we indeed observed a stronger negative correlation between BFG-PCA PPI scores and the distance between pairs of C-termini ($R = -0.41$, p-value = 2.04×10^{-14}, Pearson correlation) than the N-termini ($R = -0.22$, p-value = 6.46×10^{-5}, Pearson correlation) (**Fig 5A,B**). We therefore also constructed BFG-PCA plasmids amenable for DHFR N-terminal tagging. We investigated if the number of detected PPIs can be increased by using a N-terminus fusion version of DHFR-PCA. We tested interactions for 41 bait-prey pairs by spot assay with the N-terminus fusion version of plasmid-based DHFR-PCA. As a result, the N-terminus DHFR-PCA captured 7 interactions which the C-terminus taggingBFG-PCA could not capture (**Fig 5C**). Since 5 out of the 7 were detected by BFG-Y2H (N-terminus tagging), these results suggest that part of the difference between PCA and Y2H comes from tagging orientation. Another captured interaction (Rpt5p and Rpt4p) had a BFG-PCA score slightly higher compared to other tested pairs but still below the threshold to call as positives. Since the distance from C/N-terminus of this pair was 47.841 Å and 9.692 Å on the crystal structure, we suspect that the DHFR reporter reconstitution needed for cell growth was not sufficient for C-terminus DHFR fusions, but adequate for N-terminus, showing its potential for further discoveries of binary PPIs.

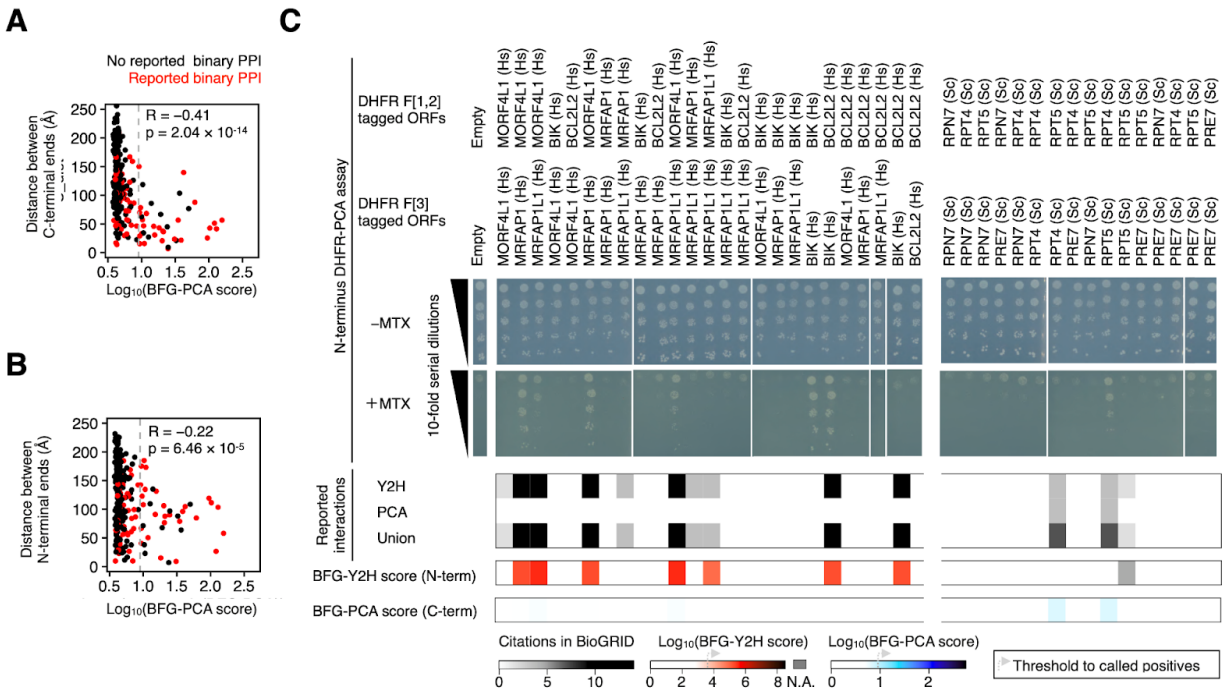

Figure 5. N-terminal DHFR-PCA detects BFG-Y2H specific positives.

A,B Scatter plot representation of BFG-PCA PPI score and distance between the most C-terminal (A) and N-terminal (B) residue of the subunits annotated within the Yeast 26S proteasome (PDB:6J2X). Red represents PPIs in the BioGRID database reported by binary PPI detection methods. Grey dashed lines represent the threshold to call positives. R represents Pearson correlation.

C Plasmid based N-terminus DHFR-PCA spot assay results for a subset of interactions screened in BFG-PCA and BFG-Y2H. Two matrices each having 5 Human proteins, and 4 Yeast proteins were tested. (Middle) Images from the spot assay. −MTX: Cell viability control without methotrexate. +MTX: Selection condition for PPI with 200 µg/mL methotrexate. Serial 10-fold dilution of cells starting at OD_{600nm} = 0.5 was plated. (Bottom) Heatmap representation of reported binary PPIs in the BioGRID database, BFG-Y2H score, and BFG-PCA score. Data for Rpt5-Rpt5 was not available for BFG-Y2H.

Conclusion

We developed a toolkit for plasmid based DHFR-PCA that exploits DNA barcode technologies for pooled screening (BFG-PCA). These tools are ready for systematic binary PPI mapping. We demonstrated the significance of BFG-PCA by screening >11,000 bait-prey pairs corresponding to 6,575 unique putative PPIs. We also performed a side-by-side comparison with BFG-Y2H for quality assessment of the method. Although it has been known that PPIs detected by DHFR-PCA and Y2H have very little overlap, no systematic comparison of the two methods has been done using the same expression promoters and the same analytical pipeline. Here, we showed that BFG-PCA and BFG-Y2H detect distinct sets of PPIs expressed from the same vector and promoter, confirming their complementarity for binary mapping. We note that BFG-PCA is significantly better at detecting yeast proteasome and nuclear pore complex related PPIs. Many reasons could explain these differences, for instance the localization of the fusion proteins. Y2H domain fused proteins are localized in the nucleus and need to have access to the chromatin and DNA to activate the expression of the selection marker, which may not always be possible for Proteasome and Nuclear Pore subunits.

Previous reports described DHFR-PCA as being able to rescue the growth of cells by having as little as 25 reconstituted complexes per cell (Remy and Michnick 1999). Since low gene expression of an interacting partner can limit the number of DHFR reporter reconstitution, plasmid-based DHFR-PCA can in theory be more sensitive to such protein pairs than genome integration based methods. However, we found no strong evidence of such increased sensitivity of BFG-PCA compared to previous genome-based DHFR-PCA datasets in the tested protein interaction space, with a few exceptions. However, we should take into account the targeted space (the proteasome and nuclear pore complexes) in this study. It is known for instance that the subunits in the proteasome are regulated at the post transcriptional level (Ascencio et al. 2021), which means that higher transcription levels from the *ADH1* promoter may not influence PPIs. Also, subunits of these two protein complexes may be more expressed already than many other proteins, leaving little room for signal improvement with this promoter. In order to comprehensively assess the sensitivity of BFG-PCA on low expressed proteins, further investigation will be needed.

The C-terminus fusion DHFR-PCA constructs used in the BFG-PCA screening in our work favoured protein pairs with closer C-terminal ends. By testing N-terminus fusion DHFR-PCA constructs, we have shown that we can detect PPIs which were not detected in C-terminus fusion DHFR-PCA. Previously, it has been reported that testing all possible fusion protein orientation (C-C, C-N, N-C, and N-N fusion of bait and prey) in a nano-luciferase complementation assay can capture as many PPIs as having multiple orthogonal assays (Choi et al. 2019). Both C-terminus and N-terminus plasmid based DHFR-PCA presented here can be used and one would be able to assay all 4 of the protein fusion combinations (C-C, C-N, N-C, and N-N), increasing the PPIs detected. BFG-PCA screenings for the four fusion combinations could be performed by the same BFG-PCA haploid yeast strains prepared for C-terminus and N-terminus BFG-PCA by simply mating them in desired combinations. This enables researchers to screen PPIs with the additional C-N and N-C combinations without additional cost to prepare barcoded yeast strains, which require investments in performing BFG screenings.

In summary, the newly developed plasmid-based pooled DHFR-PCA is a binary PPI detection method orthogonal to existing assays that can expand the interactome space to be targeted in yeast but also for any species for which it is possible to clone and expressed ORF in yeast.

Structured Methods

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Experimental Models		
Y8800 (Saccharomyces cerevisiae, MATa)	James et al., 1996	MATa leu2-3,112 trp1-901 his3-200 ura3-52 gal4D gal80D GAL2-ADE2 LYS2::GAL1-HIS3 MET2::GAL7-lacZ cyh2R
Y8930 (Saccharomyces cerevisiae, MATα)	James et al., 1996	MATα leu2-3,112 trp1-901 his3-200 ura3-52 gal4D gal80D GAL2-ADE2 LYS2::GAL1-HIS3 MET2::GAL7-lacZ cyh2R
RY1010 (Saccharomyces cerevisiae, MATa)	Yachie et al, 2016	MATa leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ PGAL2-ADE2 LYS2::PGAL1-HIS3 MET2::PGAL7-lacZ cyh2R can1Δ::PCMV-rtTA-KanMX4
RY1030 (Saccharomyces cerevisiae, MATα)	Yachie et al, 2016	MATα leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ PGAL2-ADE2 LYS2::PGAL1-HIS3 MET2::PGAL7-lacZ cyh2R can1Δ::TADH1-PtetO2-Cre-TCYC1-KanMX
YY3094 (Saccharomyces cerevisiae, MATa)	Marchant et al, 2019	MATa leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ LYS2::PGAL1-HIS3 MET2::PGAL7-lacZ cyh2R can1Δ::PCMV-rtTA-KanMX4
YY3095 (Saccharomyces cerevisiae, MATα)	Marchant et al, 2019	MATα leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ LYS2::PGAL1-HIS3 MET2::PGAL7-lacZ cyh2R can1Δ::TADH1-PtetO2-Cre-TCYC1-KanMX
Recombinant DNA		
pDEST-AD	Rural et al, 2005	N/A
pDEST-DB	Rural et al, 2005	N/A
pDN0501	Marchant et al, 2019	N/A
pDN0502	Marchant et al, 2019	N/A
pDEST-DHFR F[1,2]-C (TRP1)	Marchant et al, 2019	N/A
pDEST-DHFR F[3]-C (LEU2)	Marchant et al, 2019	N/A
pDEST-DHFR F[1,2]-N (TRP1)	This study	N/A
pDEST-DHFR F[1,2]-N (LEU2)	This study	N/A
pDEST-DHFR F[3]-N (LEU2)	This study	N/A
Barcoded pDEST-AD	This study	Table EV1.
Barcoded pDEST-DB	This study	Table EV1.
Barcoded pDN0509;pDEST-DHFR F[1,2]-C (TRP1)	This study	Table EV1.
Barcoded pDN0510;pDEST-DHFR F[3]-C (LEU2)	This study	Table EV1.

Oligonucleotides and sequence-based reagents		
PCR primers and gene fragments	Yachie et al, 2016 and this study.	Table EV2
Chemicals, enzymes and other reagents		
Yeast Nitrogen Base with ammonium sulfate, powder	MP Biomedicals	Cat #4027512
Yeast Nitrogen Base without ammonium sulfate, powder	MP Biomedicals	Cat #4027112
Amino acid dropout mix	Amberg, Burke, and Strathern, 2005	N/A
Doxycycline	Sigma-Aldrich	Cat #D9891
Methotrexate	Tokyo Chemical Industry	Cat #M1664
Gibson in vitro assembly mix	Gibson et al, 2009	N/A
Noble Agar	Sigma-Aldrich	Cat #A5431
Frozen-EZ Yeast Transformation II Kit	Zymo Research	Cat #T2001
Charge switch Yeast plasmid mini kit	Thermo Fisher Scientific	Cat #CS10203
Phusion® High-Fidelity PCR Master Mix with HF Buffer	NEB	Cat #M0531S
Software		
Python version 3.6.1	www.python.org	N/A
R version 4.0.4	https://www.r-project.org	N/A
BLASTn version 2.4.0	Altschul et al, 1990	N/A
PDBePISA version 1.52	Krissinel and Henrick, 2007	N/A
Pymol version 2.5.0	Schrödinger Inc.	N/A
Other		
Illumina MiSeq	Illumina	N/A
Qpix450	Molecular Device	N/A
Liquidator 96	RAININ	Cat #LIQ-96-200
Pre-Sterilized Breathable Sealing Film for cell growth	Corning	Cat #BF-400-S

539

Methods and Protocols

Plasmid construction

For the plasmid-based DHFR-PCA and BFG-PCA, we used Gateway cloning-compatible plasmid vectors which we previously constructed based on the Y2H plasmids pDEST-AD and pDEST-DB (Marchant et al. 2019). Plasmid-based DHFR-PCA vectors bearing the DHFR fragment domain on the N-terminus end of the protein (pDEST-DHFR F[1,2]-N and pDEST-DHFR F[3]-N) were constructed for this study. To generate pDEST-DHFR F[1,2]-N (LEU2) and pDEST-DHFR F[3]-N (LEU2), the DB domain of pDN0502 (LEU2) was replaced with DHFR fragments by ligation. The backbone fragment was prepared by restriction digestion of pDN0502 using HindIII and NotI and purified by size-selection on gel. The insert DHFR fragments DEY030 (DHFR F[1,2]) and DEY031 (DHFR F[3]) were ordered as gene fragments (TWIST biosciences). The fragments were amplified using primers DEY032 and DEY033. The insert fragments were purified on gel after digestion with HindIII and NotI. The pDN0502 backbone and each of the inserts were used for ligation to generate pDEST-DHFR F[1,2]-N (LEU2) and pDEST-DHFR F[3]-N (LEU2). To generate pDEST-DHFR F[1,2]-N (TRP1), pDN0501 and pDEST-DHFR F[1,2]-N (LEU2) were digested with I-CeuI and I-SceI to size select the backbone and insert, respectively. The two fragments were used for ligation to generate the pDEST-DHFR F[1,2]-N (TRP1) plasmid. After Gateway LR cloning of entry clones to these destination plasmids, the expression plasmids encode DHFR-linker-protein fusion protein with the following linker sequence GGGSTSTSLYKKVG. The plasmids were each confirmed for their correct construction by Sanger sequencing.

Construction of Y2H and DHFR plasmids and strains for PPI assay

Expression plasmids were generated by subcloning ORF regions of entry plasmids to destination vectors by Gateway LR reaction (Walhout et al. 2000). In detail, 10 ng each of entry plasmid and destination plasmid was mixed with Gateway LR clonase II (Invitrogen) in a 2 µL final volume and incubated at 25 °C for at least 16 hours. The entire volume of the enzymatic reaction was used to transform 25 µL of NEB5-alpha chemically competent *E.coli* cells, prepared as previously described (Swords 2003). The transformation was performed as in (Dreze et al. 2010) but with selection on LB+ampicillin plates followed by incubation overnight at 37°C instead of direct inoculation to liquid culture. The colonies were scraped and cultured in 5 mL LB+ampicillin for plasmid purification.

Yeast medium for Y2H and DHFR-PCA assays

Haploid and diploid strain cultures of Y2H and PCA samples were cultured in SC−Leu+Ade+His/SC−Trp+Ade+His and SC−Leu−Trp+Ade+His, respectively. Mating was performed in YPAD medium. For Y2H selection, the control condition was on SC−Leu−Trp+0.18 mg /mL Ade+8 mM His plates, and selection conditions was on SC−Leu−Trp-His+0.18 mg/mL Ade and SC−Leu−Trp−His+ 0.18 mg /mL Ade+ 1 mM amino-1,2,4-triazole (3-AT). Preparation was carried out as previously described (Dreze et al. 2010), and shown in **Appendix Note S2**. The amino acid dropout mix (DO mix) was prepared as previously described (Amberg, Burke, and Strathern 2005).

For DHFR-PCA and BFG-PCA screenings, methotrexate (MTX) (Tokyo Chemical Industry co., ltd.) was dissolved in 20 mL of DMSO according to the final concentration. The control condition was SC−Leu−Trp−Ade+8 mM His+2.0 % (v/v) DMSO, and default selection was SC−Leu−Trp−Ade+8 mM His+2.0 % (v/v) DMSO+ 200 μg/mL methotrexate. Preparation was carried out as previously described (Tarassov et al. 2008; Rochette et al. 2015), and shown in **Appendix Note S2**. The 10× DO mix solution was prepared by dissolving 15 g of the powder DO mix in deionized water, and filtered for sterilization.

DHFR-PCA spot and pintool assays

The purified pDHFR F[1,2]-ORF and pDHFR F[3] expression plasmids were used to transform stains YY3094 and YY3095, respectively. The pDHFR F[3]-ORF (bait) and pDHFR F[1,2]-ORF (prey) transformants were respectively selected on SC−Leu+Ade+His or SC−Trp+Ade+His plates at 30 °C for 48 hours. The resulting haploid bait and prey strains were pre-cultured individually in 96-well plates with 200 μL of media and incubated at 30 °C for 2 overnights. SC−Leu+Ade+His and SC−Trp+Ade+His media were used to culture bait and prey strains, respectively. The haploid strains were mated by spotting 5 μl each of the bait and prey culture for all protein pair combinations on YPAD plates and incubated overnight at 30 °C for mating. The mated samples were inoculated to 1 mL of SC−Leu−Trp+Ade+His liquid medium in a deep 96-well plate, and diploid cells were selected by incubation at 30 ˚C with 200 rpm agitation overnight. The resulting diploid culture was centrifuged at 500×g and resuspended in autoclaved Millipore quality H_2O twice and subjected to OD_{600} measurement. For selection, the samples were spotted at OD_{600} = 0.5 at a volume of 5 μl on 3.0% (w/v) agar plates of SC−Ade−Leu−Trp+2% (v/v) DMSO (−MTX), and SC−Ade−Leu−Trp+2% (v/v) DMSO + 200

602 μg/mL methotrexate (＋MTX). The selection plates were incubated for 72 hours at 30 °C for
603 growth scoring, and further incubated at 30 ˚C and observed every 24 hours.

604 The pintool assay was performed using the generated DHFR-PCA strains based on previously
605 described standard procedures (Rochette et al. 2015; Tarassov et al. 2008).

606 Barcode Fusion assay

607 Barcoded plasmids were transformed each into Y2H (Y8800 and Y8930), BFG-Y2H (RY1010
608 and RY1030), and BFG-PCA (YY3094 and YY3095) strains, mated, and selected for diploid as
609 above. The diploid samples were subjected to doxycycline induction after adjusting OD_{600nm} to
610 1.0 in 2.5mL of SC-Leu-Trp+Ade+10 μg/mL doxycycline, and incubated 30 °C in rotation for one
611 overnight until the OD_{600nm} reached 5.0. The samples were lysed as previously described
612 (Lõoke, Kristjuhan, and Kristjuhan 2011), and genotyping PCR was performed with conditions
613 as in Yachie *et al* (Yachie et al. 2016).

614 Selection of ORFs used in this study

615 Positive controls were picked based on known Y2H interactions reported in the BioGRID
616 database and retrieved from the CCSB human ORFeome resource (Rual et al. 2005). Nuclear
617 pore complex (NPC) and proteasome related proteins were searched in the Uniprot (UniProt
618 Consortium 2021) using keywords, “nuclear pore complex” and “proteasome”, respectively.
619 Among the list, we accessed clones available from the *S.cerevisiae* Movable ORF collection
620 (Gelperin et al. 2005), and quality controlled by Sanger sequencing using primer DEY034. The
621 complete list of selected ORFs are shown in **Table EV3**.

622 Generation of barcoded Y2H and DHFR-PCA destination plasmid libraries

623 In total, 1,867 barcoded Y2H destination plasmids (1,137 for pDEST-AD and 730 for
624 pDEST-DB) were generated as in (Yachie et al., 2016). Briefly, two PCR products each having a
625 random 25-bp flanked by lox sites and overlapping sequences were integrated into the SacI site
626 of pDEST-AD or pDEST-DB via *in vitro* DNA assembly (Gibson et al. 2009). This barcoded
627 destination vector pool was transformed into One Shot ccdB Survival 2 $T1^R$ Competent Cells
628 (Invitrogen) that were spread on 245 mm × 245 mm square LB+ampicillin plates and incubated
629 overnight at 37°C for colony isolation. Single colonies were picked by the QPix 450 robot
630 (Molecular Device) and arrayed into a 384-well format. Two Row Column Plate-PCRs

(RCP-PCRs) (Yachie et al. 2016) were performed to identify clonal samples with their barcode sequences (BC-RCP-PCR) and to check the integrity of loxP and lox2272 sequences (Lox-RCP-PCR). RCP-PCR samples were multiplexed with other libraries, and sequenced on an Illumina MiSeq (2×250 bp paired-end sequencing). Pair of barcodes that had less than 5 % abundance within the well were eliminated to cancel out sequencing errors. The quality criteria was set so that only wells containing a single pair of barcode sequences with the designed elements were used for downstream processes.

The barcoded DHFR-PCA destination plasmids were generated similarly using destination plasmids pDEST-DHFR F[1,2]-C (TRP1) and pDEST-DHFR F[3]-C (LEU2). In total, 1,483 barcoded DHFR-PCA destination plasmids (893 for pDEST-DHFR F[1,2]-C and 590 for pDEST-DHFR F[3]-C) were generated. The destination vectors were digested with PI-PspI, and the two PCR products with random barcodes were inserted via *in vitro* DNA assembly (Gibson et al. 2009). The PCR primers to generate the barcodes were altered from that of BFG-Y2H due to change in insert site. The primers used here are shown in **Table EV2**. The isolated bacterial colonies having barcoded destination vectors were prepared in the same manner as Y2H destination vectors. Two RCP-PCR were performed with the same design as in Y2H, but with minor modification in the primer used for Lox-RCP-PCR (**Table EV2**). The list of prepared barcodes are shown in **Table EV3**.

BFG-Y2H and BFG-PCA ready yeast strain generation

Barcoded expression plasmids with defined ORF-barcode associations were generated by one-by-one Gateway cloning. Similar to the non-barcoded expression vector preparation, ORF regions of entry plasmids were subclonded by Gateway LR cloning to a mix of 2 pre-assigned uniquely barcoded destination vectors. In detail, 10 ng of each entry plasmid and destination plasmid was mixed with Gateway LR clonase II (Invitrogen) in a 4 µL volume and incubated at 25 °C for at least 16 hours. Transformation of LR samples was performed in the same manner as non-barcoded samples. More than 5 colonies were scraped per sample to ensure representation of both barcodes, and cultured in 5 mL LB+ampicillin for plasmid purification. Purified plasmid was used to transform corresponding strains with appropriate selection medium. All prepared strains are listed in **Table EV3**.

BFG-Y2H and BFG-PCA screenings

Haploid bait and prey strains were cultured to saturation by incubating at 30 °C in a static manner for approximately 60 hours in a 96-well deep well plate sealed with a breathable seal (Corning, BF-400-S). Each well contained 1 mL of SC−Leu+Ade+His or SC−Trp+Ade+His liquid media depending on the plasmids. Strains were pooled (AD,DB, DHFR F[1,2], or DHFR F[3]) by mixing 1 mL of 2 OD_{600nm} equivalent cells for each strain. For mating, two groups of cell pools were mixed at equal amounts, and incubated at room temperature for 3 hours. After the incubation, the sample was spun down at 500×g for 4 minutes and then the cell pellet was spread on a YPAD plate. The plate was incubated at room temperature for 16 hours. The mated sample was scraped with autoclaved Millipore quality H_2O, and then washed twice by spinning down the sample at 500×g for 4 minutes, and resuspending in SC−Leu−Trp+Ade liquid medium. The diploid cells were selected at a starting OD_{600nm} of 1.0 in a 2 L flask containing 500 mL of SC−Leu−Trp+Ade liquid medium, and incubated for two overnights at 30 °C, 160 rpm. Fifty mL of the diploid selected culture sample was spun down, and washed twice with water. The screening was performed by plating an equivalent of 1 mL of 5.0 OD_{600nm} cells per plate on 15 plates for each selection condition tested. The number of samples plated for selection was determined by a Monte-Carlo simulation model described in Yachie *et al* to ensure 100% of the positive diploid strain having at least 100 cells passing to the following step. The selected samples were collected after 72 hours of incubation at 30 °C. The collected samples was subjected to doxycycline induction after adjusting OD_{600nm} to 1.0 in 25mL of SC-Leu-Trp+Ade+10 µg/mL doxycycline, and incubated 30 °C for one overnight until the OD_{600nm} reached 5.0. The DNA extraction and deep sequencing library preparation was performed according to procedures shown in Yachie *et al.* Deep sequencing libraries were multiplexed with other libraries, and sequenced by Illumina MiSeq (2×250 bp paired-end sequencing). Reads were demultiplexed and fused DNA barcodes were counted by alignment of primer sequences and DNA barcodes using BLASTn version 2.4.0 (Altschul et al. 1990) with the blastn-short option and an E-value threshold of 1e-10.

BFG-Y2H and BFG-PCA data normalization

Both data normalization for BFG-Y2H and BFG-PCA data was performed with custom Python scripts. For BFG-Y2H data normalization, the procedures followed the method previously described (Yachie et al. 2016). The detailed procedure for normalizing BFG-PCA data is described below.

For each condition and barcode fusion type (BC1-BC1 or BC2-BC2 fusion), the relative abundance of each diploid strain was estimated from the aggregated barcode count data. Note that a constant value of 1 was added to the barcode count of each strain to reduce noise for smaller values. For the non-selective (control) condition, all diploid strains are expected to grow, which results in high sequence complexity. Given that the deep sequencing depth is limited for the entire dynamic range of this complex pool, we first estimate the relative abundance computed as frequency for each $Bait_i$ or $Prey_j$ amongst all diploid strains Dip_{ij} as

$$F_i^{control} = \frac{\Sigma C_i^{control}}{\Sigma C_{ij}^{control}} \quad \text{and} \quad F_j^{control} = \frac{\Sigma C_j^{control}}{\Sigma C_{ij}^{control}}$$

where C is the sequencing read count within each condition or barcode fusion type, respectively. Since chances for mating of each haploid combination is dependent on the relative abundance of each haploid strain, the frequency of diploid Dip_{ij} (having $Bait_i$ and $Prey_j$) in non-selection condition ($F_{ij}^{control}$) can be estimated as

$$F_{ij}^{control} = F_i^{control} \times F_j^{control}$$

Relative growth of diploid in selection condition $F_{ij}^{selection}$ was directly computed from raw count data as

$$F_{ij}^{selection} = \frac{C_{ij}^{selection}}{\Sigma C_{ij}^{selection}}$$

due to the sparse nature of PPI positives.

Based on relative abundance on non-selective and selective conditions, enrichment signal s was computed as

$$s_{ij} = \frac{F_{ij}^{selection}}{F_{ij}^{control}}$$

where s represents a degree of growth enrichment in favor of the selective condition for each diploid strain.

Similar to the BFG-Y2H data, we observed different background levels of s for each haploid strain in BFG-PCA. We defined the background as the median of all s values for each haploid. The normalized score ds was computed for each diploid as

$$ds_{ij} = \frac{s_{ij}+1}{\{median(s_i)+1\} \times \{median(s_j)+1\}}$$

and subjected to PPI calling and analysis.

PPI analysis

PPI analysis was performed by first aggregating all PPI scores for each protein pair, combining replicates and both bait-prey orientations tested. For each protein pair, PPI scores were sorted, and various percentiles (1, 5, 10, 15, 20, … 90, 95, 99), average, and median values were calculated. We scored protein pairs based on each of the scoring methods (average, median, and each of the percentiles). Based on each scoring method, the protein pairs were sorted from highest to lowest, and subjected to computing Matthews Correlation Coefficient (MCC) (Matthews 1975) against the BioGRID database (version 4.4.198) (Stark et al. 2006) for quality assessment. We defined all PPIs reported in BioGRID by binary PPI detection methods (Y2H, PCA, Biochemical activity, Affinity Capture-Luminescence, Reconstituted Complex, Co-crystal Structure, and FRET) as positives, and categorized true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each rank threshold. The MCC for each rank threshold was computed as

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$$

, and we defined the optimal threshold to call positives when the MCC value scored the highest.

Crystal structure analysis

The crystal structure data of the yeast 26S proteasome (Ding et al. 2019) (PDB: 6J2X) was used to calculate solvent free energy between each subunit using PDBePISA (Krissinel and Henrick 2007). The solvent free energy values were summed when multiple protein chains were available for the subunits. Kendall rank correlation was used for the statistical test. Distances from the C/N-terminal ends of the subunits were computed using the `get_distance` function of Pymol version 2.5.0 (Schrödinger, Inc.). The closest residue to the terminal ends available on the crystal structure was used. We adopted the closest values among subunits by considering only the α and β rings closer to the lid particle. Pearson correlation was used to compute the coefficient.

Method comparison analysis

Comparison analysis of the detected PPIs were carried out against previous genomic integration based DHFR-PCA (Tarassov et al. 2008; Chrétien et al. 2018). For the protein pairs present in both BFG-PCA dataset and Tarassov *et al*, the best performing scoring method and average score of replicates were extracted from each dataset. Protein expression analysis was

749 performed using protein abundance data (Ho, Baryshnikova, and Brown 2018). For each protein
750 pair we considered the lowest expression of the pair because it is likely the limiting partner in
751 complex formation. Mann-Whitney U-test was used for statistical tests.

752 Resource availability

753 The DHFR-PCA plasmids for both C-terminus fusion and N-terminus fusion will be available at
754 Addgene. The barcoded BFG-Y2H and BFG-PCA destination plasmids will also be available at
755 Addgene.

756 Acknowledgements

757 CRL holds the Canada Research Chair in Cellular Systems and Synthetic Biology. NY holds the
758 Canada Research Chair in Synthetic Biology. This work was funded by Canadian Institutes of
759 Health Research Foundation grant 387697 (to CRL) and the Japan Society for the Promotion of
760 Science (JSPS) 18H02428, the Daiichi Sankyo Foundation of Life Science, the Ube Foundation,
761 and the Astellas Foundation for Research on Metabolic Disorders (to NY). DEY was supported
762 by a DC1 Fellowship from JSPS, a Graduate Fellowship for Young Leaders from the Sylff
763 association, TTCK Fellowship, Taikichiro Mori Memorial Research Grant, and the Yamagishi
764 Student Project Support Program of Keio University.

765 Author contributions

766 DEY and NY conceived and designed the study. NY and CRL designed and supervised the
767 study. DEY, HM, and AKD performed the high-density plate DHFR-PCA assay experiments with
768 computational analysis by PD and DA. DEY and PN performed the experiments to optimize the
769 BFG-PCA screening condition. DEY, FDR, and KM performed the BFG-PCA and BFG-Y2H
770 screenings with support from MS. DEY and YL performed the N-terminus DHFR-PCA assay
771 experiment. DEY performed the data analysis. DEY and CRL wrote the manuscript and all
772 authors revised and approved the manuscript.

773 Conflict of interest

774 None declared.

775

References

Adams, Rebecca L., Aaron C. Mason, Laura Glass, Aditi, and Susan R. Wente. 2018. "Correction to: Nup42 and IP6 Coordinate Gle1 Stimulation of Dbp5/DDX19B for mRNA Export in Yeast and Human Cells." *Traffic* 19 (8): 650.

Aebersold, Ruedi, Jeffrey N. Agar, I. Jonathan Amster, Mark S. Baker, Carolyn R. Bertozzi, Emily S. Boja, Catherine E. Costello, et al. 2018. "How Many Human Proteoforms Are There?" *Nature Chemical Biology* 14 (3): 206–14.

Alber, Frank, Svetlana Dokudovskaya, Liesbeth M. Veenhoff, Wenzhu Zhang, Julia Kipper, Damien Devos, Adisetyantari Suprapto, et al. 2007. "Determining the Architectures of Macromolecular Assemblies." *Nature* 450 (7170): 683–94.

Alberts, Bruce. 1998. "The Cell as a Collection Overview of Protein Machines: Preparing the next Generation of Molecular Biologists." *Cell* 92: 291–94.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology*. https://doi.org/10.1016/s0022-2836(05)80360-2.

Amberg, David C., Dan Burke, and Jeffrey N. Strathern. 2005. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. CSHL Press.

Arabidopsis Interactome Mapping Consortium. 2011. "Evidence for Network Evolution in an Arabidopsis Interactome Map." *Science* 333 (6042): 601–7.

Ascencio, Diana, Guillaume Diss, Isabelle Gagnon-Arsenault, Alexandre K. Dubé, Alexander DeLuna, and Christian R. Landry. 2021. "Expression Attenuation as a Mechanism of Robustness against Gene Duplication." *Proceedings of the National Academy of Sciences of the United States of America* 118 (6). https://doi.org/10.1073/pnas.2014345118.

Braun, Pascal, Murat Tasan, Matija Dreze, Miriam Barrios-Rodiles, Irma Lemmens, Haiyuan Yu, Julie M. Sahalie, et al. 2009. "An Experimentally Derived Confidence Score for Binary Protein-Protein Interactions." *Nature Methods* 6 (1): 91–97.

Bruce, James E. 2012. "In Vivo Protein Complex Topologies: Sights through a Cross-Linking Lens." *Proteomics* 12 (10): 1565–75.

Buntru, Alexander, Philipp Trepte, Konrad Klockmeier, Sigrid Schnoegl, and Erich E. Wanker. 2016. "Current Approaches Toward Quantitative Mapping of the Interactome." *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2016.00074.

Celaj, Albi, Ulrich Schlecht, Justin D. Smith, Weihong Xu, Sundari Suresh, Molly Miranda, Ana Maria Aparicio, et al. 2017. "Quantitative Analysis of Protein Interaction Network Dynamics in Yeast." *Molecular Systems Biology* 13 (7): 934.

Choi, Soon Gang, Julien Olivet, Patricia Cassonnet, Pierre-Olivier Vidalain, Katja Luck, Luke Lambourne, Kerstin Spirohn, et al. 2019. "Maximizing Binary Interactome Mapping with a Minimal Number of Assays." *Nature Communications* 10 (1): 3907.

Chrétien, Andrée-Ève, Isabelle Gagnon-Arsenault, Alexandre K. Dubé, Xavier Barbeau, Philippe C. Després, Claudine Lamothe, Anne-Marie Dion-Côté, Patrick Lagüe, and Christian R. Landry. 2018. "Extended Linkers Improve the Detection of Protein-Protein Interactions (PPIs) by Dihydrofolate Reductase Protein-Fragment Complementation Assay (DHFR PCA) in Living Cells." *Molecular & Cellular Proteomics*. https://doi.org/10.1074/mcp.a117.000385.

Corominas, Roser, Xinping Yang, Guan Ning Lin, Shuli Kang, Yun Shen, Lila Ghamsari, Martin Broly, et al. 2014. "Protein Interaction Network of Alternatively Spliced Isoforms from Brain Links Genetic Risk Factors for Autism." *Nature Communications* 5 (April): 3650.

Ding, Zhanyu, Cong Xu, Indrajit Sahu, Yifan Wang, Zhenglin Fu, Min Huang, Catherine C. L. Wong, Michael H. Glickman, and Yao Cong. 2019. "Structural Snapshots of 26S Proteasome Reveal Tetraubiquitin-Induced Conformations." *Molecular Cell* 73 (6): 1150–61.e6.

Dreze, Matija, Dario Monachello, Claire Lurin, Michael E. Cusick, David E. Hill, Marc Vidal, and Pascal Braun. 2010. “High-Quality Binary Interactome Mapping.” *Methods in Enzymology* 470 (March): 281–315.

Fields, S., and O. Song. 1989. “A Novel Genetic System to Detect Protein-Protein Interactions.” *Nature* 340 (6230): 245–46.

Gelperin, Daniel M., Michael A. White, Martha L. Wilkinson, Yoshiko Kon, Li A. Kung, Kevin J. Wise, Nelson Lopez-Hoyo, et al. 2005. “Biochemical and Genetic Analysis of the Yeast Proteome with a Movable ORF Collection.” *Genes & Development* 19 (23): 2816–26.

Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison 3rd, and Hamilton O. Smith. 2009. “Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases.” *Nature Methods* 6 (5): 343–45.

Gillet, Ludovic C., Pedro Navarro, Stephen Tate, Hannes Röst, Nathalie Selevsek, Lukas Reiter, Ron Bonner, and Ruedi Aebersold. 2012. “Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis.” *Molecular & Cellular Proteomics: MCP* 11 (6): O111.016717.

Go, Christopher D., James D. R. Knight, Archita Rajasekharan, Bhavisha Rathod, Geoffrey G. Hesketh, Kento T. Abe, Ji-Young Youn, et al. 2021. “A Proximity-Dependent Biotinylation Map of a Human Cell.” *Nature* 595 (7865): 120–24.

Grigoriev, A. 2001. “A Relationship between Gene Expression and Protein Interactions on the Proteome Scale: Analysis of the Bacteriophage T7 and the Yeast Saccharomyces Cerevisiae.” *Nucleic Acids Research* 29 (17): 3513–19.

Ho, Brandon, Anastasia Baryshnikova, and Grant W. Brown. 2018. “Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces Cerevisiae Proteome.” *Cell Systems* 6 (2): 192–205.e3.

Hu, Chang-Deng, Yurii Chinenov, and Tom K. Kerppola. 2002. “Visualization of Interactions among bZIP and Rel Family Proteins in Living Cells Using Bimolecular Fluorescence Complementation.” *Molecular Cell* 9 (4): 789–98.

Krissinel, Evgeny, and Kim Henrick. 2007. “Inference of Macromolecular Assemblies from Crystalline State.” *Journal of Molecular Biology* 372 (3): 774–97.

Kristensen, Anders R., Joerg Gsponer, and Leonard J. Foster. 2012. “A High-Throughput Approach for Measuring Temporal Changes in the Interactome.” *Nature Methods* 9 (9): 907–9.

Lin, Daniel H., Ana R. Correia, Sarah W. Cai, Ferdinand M. Huber, Claudia A. Jette, and André Hoelz. 2018. “Structural and Functional Analysis of mRNA Export Regulation by the Nuclear Pore Complex.” *Nature Communications* 9 (1): 2319.

Li, Siming, Christopher M. Armstrong, Nicolas Bertin, Hui Ge, Stuart Milstein, Mike Boxem, Pierre-Olivier Vidalain, et al. 2004. “A Map of the Interactome Network of the Metazoan C. Elegans.” *Science* 303 (5657): 540–43.

Liu, Zhimin, Darach Miller, Fangfei Li, Xianan Liu, and Sasha F. Levy. 2020. “A Large Accessory Protein Interactome Is Rewired across Environments.” *eLife* 9 (September). https://doi.org/10.7554/eLife.62365.

Lõoke, Marko, Kersti Kristjuhan, and Arnold Kristjuhan. 2011. “Extraction of Genomic DNA from Yeasts for PCR-Based Applications.” *BioTechniques* 50 (5): 325–28.

Luck, Katja, Dae-Kyum Kim, Luke Lambourne, Kerstin Spirohn, Bridget E. Begg, Wenting Bian, Ruth Brignall, et al. 2020. “A Reference Map of the Human Binary Protein Interactome.” *Nature* 580 (7803): 402–8.

MacDonald, Marnie L., Jane Lamerdin, Stephen Owens, Brigitte H. Keon, Graham K. Bilter, Zhidi Shang, Zhengping Huang, et al. 2006. “Identifying off-Target Effects and Hidden Phenotypes of Drugs in Human Cells.” *Nature Chemical Biology*. https://doi.org/10.1038/nchembio790.

Marchant, Axelle, Angel F. Cisneros, Alexandre K. Dubé, Isabelle Gagnon-Arsenault, Diana

Ascencio, Honey Jain, Simon Aubé, et al. 2019. “The Role of Structural Pleiotropy and Regulatory Evolution in the Retention of Heteromers of Paralogs.” *eLife* 8 (August). https://doi.org/10.7554/eLife.46754.

Matthews, B. W. 1975. “Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme.” *Biochimica et Biophysica Acta* 405 (2): 442–51.

Murphy, R., and S. R. Wente. 1996. “An RNA-Export Mediator with an Essential Nuclear Export Signal.” *Nature* 383 (6598): 357–60.

Rajagopala, Seesandra V., Patricia Sikorski, Ashwani Kumar, Roberto Mosca, James Vlasblom, Roland Arnold, Jonathan Franca-Koh, et al. 2014. “The Binary Protein-Protein Interaction Landscape of Escherichia Coli.” *Nature Biotechnology* 32 (3): 285–90.

Remy, I., and S. W. Michnick. 1999. “Clonal Selection and in Vivo Quantitation of Protein Interactions with Protein-Fragment Complementation Assays.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (10): 5394–99.

Ren, Ling, Daryl Emery, Barbara Kaboord, Edith Chang, and M. Walid Qoronfleh. 2003. “Improved Immunomatrix Methods to Detect Protein:protein Interactions.” *Journal of Biochemical and Biophysical Methods*. https://doi.org/10.1016/s0165-022x(03)00105-2.

Rhee, Hyun-Woo, Peng Zou, Namrata D. Udeshi, Jeffrey D. Martell, Vamsi K. Mootha, Steven A. Carr, and Alice Y. Ting. 2013. “Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging.” *Science* 339 (6125): 1328–31.

Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Séraphin. 1999. “A Generic Protein Purification Method for Protein Complex Characterization and Proteome Exploration.” *Nature Biotechnology* 17 (10): 1030–32.

Rochette, Samuel, Guillaume Diss, Marie Filteau, Jean-Baptiste Leducq, Alexandre K. Dubé, and Christian R. Landry. 2015. “Genome-Wide Protein-Protein Interaction Screening by Protein-Fragment Complementation Assay (PCA) in Living Cells.” *Journal of Visualized Experiments: JoVE*, no. 97 (March). https://doi.org/10.3791/52255.

Rolland, Thomas, Murat Taşan, Benoit Charloteaux, Samuel J. Pevzner, Quan Zhong, Nidhi Sahni, Song Yi, et al. 2014. “A Proteome-Scale Map of the Human Interactome Network.” *Cell* 159 (5): 1212–26.

Roux, Kyle J., Dae In Kim, Manfred Raida, and Brian Burke. 2012. “A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells.” *The Journal of Cell Biology* 196 (6): 801–10.

Rual, Jean-François, Kavitha Venkatesan, Tong Hao, Tomoko Hirozane-Kishikawa, Amélie Dricot, Ning Li, Gabriel F. Berriz, et al. 2005. “Towards a Proteome-Scale Map of the Human Protein-Protein Interaction Network.” *Nature* 437 (7062): 1173–78.

Sahni, Nidhi, Song Yi, Quan Zhong, Noor Jailkhani, Benoit Charloteaux, Michael E. Cusick, and Marc Vidal. 2013. “Edgotype: A Fundamental Link between Genotype and Phenotype.” *Current Opinion in Genetics & Development* 23 (6): 649–57.

Salas, Daniela, R. Greg Stacey, Mopelola Akinlaja, and Leonard J. Foster. 2020. “Next-Generation Interactomics: Considerations for the Use of Co-Elution to Measure Protein Interaction Networks.” *Molecular & Cellular Proteomics: MCP* 19 (1): 1–10.

Schlecht, Ulrich, Zhimin Liu, Jamie R. Blundell, Robert P. St Onge, and Sasha F. Levy. 2017. “A Scalable Double-Barcode Sequencing Platform for Characterization of Dynamic Protein-Protein Interactions.” *Nature Communications* 8 (May): 15586.

Schlecht, Ulrich, Molly Miranda, Sundari Suresh, Ronald W. Davis, and Robert P. St Onge. 2012. “Multiplex Assay for Condition-Dependent Changes in Protein-Protein Interactions.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (23): 9213–18.

Sinz, Andrea. 2010. “Investigation of Protein-Protein Interactions in Living Cells by Chemical Crosslinking and Mass Spectrometry.” *Analytical and Bioanalytical Chemistry* 397 (8): 3433–40.

927 Smith, Lloyd M., Neil L. Kelleher, and Consortium for Top Down Proteomics. 2013. “Proteoform:
928 A Single Term Describing Protein Complexity.” *Nature Methods* 10 (3): 186–87.
929 Stark, Chris, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and
930 Mike Tyers. 2006. “BioGRID: A General Repository for Interaction Datasets.” *Nucleic Acids*
931 *Research*. https://doi.org/10.1093/nar/gkj109.
932 Strahm, Y. 1999. “The RNA Export Factor Gle1p Is Located on the Cytoplasmic Fibrils of the
933 NPC and Physically Interacts with the FG-Nucleoporin Rip1p, the DEAD-Box Protein
934 Rat8p/Dbp5p and a New Protein Ymr255p.” *The EMBO Journal*.
935 https://doi.org/10.1093/emboj/18.20.5761.
936 Swords, Edward W. 2003. “Chemical Transformation of E. Coli.” *Methods Mol Bio* 235: 49–53.
937 Tarassov, Kirill, Vincent Messier, Christian R. Landry, Stevo Radinovic, Mercedes M. Serna
938 Molina, Igor Shames, Yelena Malitskaya, Jackie Vogel, Howard Bussey, and Stephen W.
939 Michnick. 2008. “An in Vivo Map of the Yeast Protein Interactome.” *Science* 320 (5882):
940 1465–70.
941 Trigg, Shelly A., Renee M. Garza, Andrew MacWilliams, Joseph R. Nery, Anna Bartlett, Rosa
942 Castanon, Adeline Goubil, et al. 2017. “CrY2H-Seq: A Massively Multiplexed Assay for
943 Deep-Coverage Interactome Mapping.” *Nature Methods* 14 (8): 819–25.
944 UniProt Consortium. 2021. “UniProt: The Universal Protein Knowledgebase in 2021.” *Nucleic*
945 *Acids Research* 49 (D1): D480–89.
946 Venkatesan, Kavitha, Jean-François Rual, Alexei Vazquez, Ulrich Stelzl, Irma Lemmens,
947 Tomoko Hirozane-Kishikawa, Tong Hao, et al. 2009. “An Empirical Framework for Binary
948 Interactome Mapping.” *Nature Methods* 6 (1): 83–90.
949 Vidal, Marc, Michael E. Cusick, and Albert-László Barabási. 2011. “Interactome Networks and
950 Human Disease.” *Cell* 144 (6): 986–98.
951 Walhout, Albertha J. M., Gary F. Temple, Michael A. Brasch, James L. Hartley, Monique A.
952 Lorson, Sander van den Heuvel, and Marc Vidal. 2000. “[34] GATEWAY Recombinational
953 Cloning: Application to the Cloning of Large Numbers of Open Reading Frames or
954 ORFeomes.” *Methods in Enzymology*. https://doi.org/10.1016/s0076-6879(00)28419-x.
955 Yachie, Nozomu, Evangelia Petsalaki, Joseph C. Mellor, Jochen Weile, Yves Jacob, Marta
956 Verby, Sedide B. Ozturk, et al. 2016. “Pooled-Matrix Protein Interaction Screens Using
957 Barcode Fusion Genetics.” *Molecular Systems Biology* 12 (4): 863.
958 Yang, Fang, Yingying Lei, Meiling Zhou, Qili Yao, Yichao Han, Xiang Wu, Wanshun Zhong, et al.
959 2018. “Development and Application of a Recombination-Based Library versus Library
960 High- Throughput Yeast Two-Hybrid (RLL-Y2H) Screening System.” *Nucleic Acids*
961 *Research* 46 (3): e17.
962 Yang, Jae-Seong, Mireia Garriga-Canut, Nele Link, Carlo Carolis, Katrina Broadbent, Violeta
963 Beltran-Sastre, Luis Serrano, and Sebastian P. Maurer. 2018. “Rec-YnH Enables
964 Simultaneous Many-by-Many Detection of Direct Protein-Protein and Protein-RNA
965 Interactions.” *Nature Communications* 9 (1): 3747.
966 Yu, Clinton, and Lan Huang. 2018. “Cross-Linking Mass Spectrometry: An Emerging
967 Technology for Interactomics and Structural Biology.” *Analytical Chemistry* 90 (1): 144–65.
968 Yu, Haiyuan, Pascal Braun, Muhammed A. Yildirim, Irma Lemmens, Kavitha Venkatesan, Julie
969 Sahalie, Tomoko Hirozane-Kishikawa, et al. 2008. “High-Quality Binary Protein Interaction
970 Map of the Yeast Interactome Network.” *Science* 322 (5898): 104–10.
971 Yu, Haiyuan, Leah Tardivo, Stanley Tam, Evan Weiner, Fana Gebreab, Changyu Fan, Nenad
972 Svrzikapa, et al. 2011. “Next-Generation Sequencing to Generate Interactome Datasets.”
973 *Nature Methods* 8 (6): 478–80.

974