A scalable double-barcode sequencing platform for characterization of dynamic protein-protein interactions

Ulrich Schlecht1,2,*, Zhimin Liu3,4,*, Jamie R. Blundell3,4,5, Robert P. St.Onge1,2 & Sasha F. Levy3,4

Several large-scale efforts have systematically catalogued protein-protein interactions (PPIs) of a cell in a single environment. However, little is known about how the protein interactome changes across environmental perturbations. Current technologies, which assay one PPI at a time, are too low throughput to make it practical to study protein interactome dynamics. Here, we develop a highly parallel protein-protein interaction sequencing (PPiSeq) platform that uses a novel double barcoding system in conjunction with the dihydrofolate reductase protein-fragment complementation assay in Saccharomyces cerevisiae. PPiSeq detects PPIs at a rate that is on par with current assays and, in contrast with current methods, quantitatively scores PPIs with enough accuracy and sensitivity to detect changes across environments. Both PPI scoring and the bulk of strain construction can be performed with cell pools, making the assay scalable and easily reproduced across environments. PPiSeq is therefore a powerful new tool for large-scale investigations of dynamic PPIs.
A genome-scale understanding of protein–protein interactions (PPIs) has been a long-standing goal of molecular and systems biology. Several large-scale efforts have used the model yeast *Saccharomyces cerevisiae* to systematically catalogue all PPIs in a single environment using the yeast two-hybrid (Y2H) or the protein fragment complementation assay (PCA) systems. These high-throughput screens take advantage of arrayed mating and differential colony growth on agar to identify PPIs. The resulting PPI networks have aided in defining the function of uncharacterized proteins, in characterizing protein complexes and modules that act in concert, and in identifying common functional network motifs. Despite these advances, we are left with a static view of a cell. Unknown is how the protein interactome changes as cells differentiate or react to stress, even though these are each fundamental biological processes. Also missing are PPIs that only occur in non-standard growth conditions.

Describing the global protein interactome dynamics, however, requires overcoming two main hurdles. First, higher throughput is required. *S. cerevisiae* has ~6,000 genes, so measuring all pairwise PPIs in one environment requires ~18 million strains to be assayed, each of which characterizes a single PPI. Observing how the protein interactome changes across perturbations would require re-assaying all ~18 million strains for each perturbation studied. Even with recent advances such as high density yeast arrays and automated mating and scoring, the time and reagent costs associated with performing screens across several environments make these studies impractical. Recently, a Y2H- and barcode-based pooled assay has been developed, which could potentially greatly increase the throughput of PPI detection. However, because this method relies on plasmid-borne constructs under exogenous promoters, and because Y2H reporter activation requires nuclear localization, this platform is not ideal for characterizing how PPIs change under natural physiological conditions.

Second, a graded PPI score is necessary to identify PPIs that change. In contrast with barcode-based Y2H, PCA uses genomically integrated fusion proteins under endogenous promoters with protein products interacting in their natural physiological setting, and therefore serves as a viable platform for detecting dynamic PPIs. Many different protein fragment designs have been used with PCA, however the highest-throughput version uses a split murine dihydrofolate reductase (mDHFR) that, once reconstructed, allows colony growth on agar containing the *S. cerevisiae* DHFR inhibitor methotrexate (MTX). These assays have historically used colony size thresholds to call whether or not a PPI exists. Graded scoring techniques that assign a fitness to each colony have been employed elsewhere, for example, in synthetic genetic interaction screens. However, it is unclear if these techniques are applicable to PCA. More recently, we and others have developed methods to detect dynamic PPIs via isolated growth of PCA strains in liquid multiwell plates followed by scoring of optical density growth curves, a method that is capable of accurately detecting at least fourfold differences in the number of interacting protein molecules over a broad range of protein abundances. However, the throughput of isolated liquid growth assays is inferior to colony-based assays, making their application to interactome-scale studies challenging.

A more scalable alternative is pooled growth of barcoded PCA strains followed by quantification to detect how changes in relative barcode abundance are altered across perturbations. However, this technique is limited in two ways. First, a uniquely barcoded strain must be generated for each PPI, making interactome-scale barcoded libraries impractical to construct (for example, 18 million barcoded strains). Second, changes in relative barcode abundance are detected by comparing to a control condition, as is typical for barcode-based assays. As we describe here, single time point relative barcode frequencies following growth are sensitive to the fitnesses of other cells in the pool, and therefore do not provide a robust score that can be compared across different pools or environments.

Here, we present a scalable and robust method to identify and quantitatively score dynamic PPIs that we call Protein–Protein interaction Sequencing (PPISeq). The PPISeq platform combines PCA, a new genomic double-barcoding technology, time-course barcode sequencing of competing cell pools, and an analytical framework to precisely call fitnesses from barcode lineage trajectories. We use these tools to examine the interactions between ~100 protein pairs at high replication and across five environments. In a permissive environment, we find that the ability for PPISeq to identify PPIs is on par with existing assays. In addition, PPISeq finds that a large fraction of PPIs change across environments, many of which could be validated by other PPI assays. Finally, we show that bulk mating of barcoded strains can generate libraries exceeding 10^9 double barcodes and therefore PPISeq could potentially be used to simultaneously assay large networks containing thousands of PPIs.

### Results

#### The PPISeq platform.

We developed a general interaction Sequencing platform (iSeq). Barcodes that are adjacent to a loxP recombination site are introduced at a common chromosomal location in closely related *MATa* and *MATα* haploids. Barcodes are placed on opposite sides of the loxP site in each mating type such that mating and Cre induction causes recombination between homologous chromosomes, resulting in a barcode-loxP-barcode configuration on one chromosome (Fig. 1 and Supplementary Fig. 1). This event is selected for by loxP recombination-induced reassembly of a split URA3 marker. A double barcode unambiguously identifies both parents of a cross in highly complex cell pools, with each barcode half being in close enough proximity to allow the pair to be sequenced together by short-read sequencing. Next, double barcode strains are grown in pools, relative double barcode frequencies are assayed at several times, and their trajectories are used in combination with a global maximum likelihood method to estimate the relative fitness of each strain. Here, we use iSeq in combination with the DHFR PCA system to construct a PPISeq platform (Fig. 1, Supplementary Fig. 1 and Supplementary Note 1).

**PPISeq is accurate and highly reproducible.** To test the reproducibility of PPISeq and compare it to existing PPI assays, we selected nine bait (mDHFR-F[1,2]) and nine prey (mDHFR-F[3]) PCA strains and added five different barcodes to each. Bait (*IMD3, HOM3, SHR3, PRS3, DBP2, TPO1, DST1, FTR1 and FMP45*) and prey (*IMD3, FPR1, SHR3, PRS3, DBP2, HXT1, PDR5, RBP9 and SNQ2*) PCA constructs were chosen to encompass a number of previously discovered PPIs. We also added five different barcodes to two control strains that do not contain an mDHFR. Haploid barcoded PCA strains were pairwise mated and pooled to generate a library of 2,500 double barcode (PPISeq) strains, with each of the 100 genotypes being represented by 25 unique double barcodes.

We next developed a pooled growth and bar-seq assay capable of robustly measuring the relative fitness of all strains in the pool. We expected that as low fitness PPISeq strains drop out of the population, the frequency trajectories of a higher fitness strain will begin to "bend" as its competition gets tougher (green lines, Fig. 2a). The dynamics of this competition depends on the abundances and relative fitnesses of all strains in the pool and...
Because of this, barcode frequencies at a single time point do not provide a constant measure of fitness across conditions. We therefore monitored relative barcode frequencies over several early time points. We grew the PPiSeq pool in triplicate in standard yeast media and the presence or absence of a low concentration of MTX for 12 generations in serial batch culture, diluting 1:8 every 24 h (3 generations, Supplementary Fig. 2). To allow for fitness measurements of all strains, we chose a low concentration of MTX (0.5 \( \mu \)g ml\(^{-1} \), 400-fold lower concentration than traditional PCA) where even strains lacking mDHFR will grow slowly. Double barcodes were sequenced at each dilution (every three generations). We found that reads representing putative PCR chimeras, double barcodes where each double barcode half stems from a different template, occurred at a low but predictable frequency (\( \sim 0.2\% \), Supplementary Note 2) and could confound our results. We therefore subtracted the expected number of PCR chimeras from each double barcode count and generated lineage trajectories with these corrected counts (Fig. 2a and Supplementary Data 1). In the absence of MTX, most PPiSeq strains do not change in frequency over time. However, in the presence of MTX, most strains are driven close to extinction by 12 generations, while others with higher fitness rise in frequency or have a slower decline.

**Figure 1 | Construction of a PPiSeq library.** Primers containing a random nucleotide barcode are inserted into a common genomic location of both MAT\(a\) and MAT\(a\) cells by homologous recombination, yielding large libraries of barcoded yeast cells. Clones from each library are picked at random and barcodes are identified by sequencing. Barcoded cells are mated to strains containing either a bait or prey PCA construct. Diploids are sporulated and haploids containing both a barcode and a PCA construct are selected. These haploids are mated to generate diploids that contain two barcodes and both bait and prey PCA constructs. Cre-induced loxP recombination brings the two barcodes to the same chromosome, and is selected for by reconstruction of a split URA3 selectable marker. The barcode-loxP-barcode sequence resides in an artificial intron and is not translated. Double barcodes mark which two PCA constructs are in each cell and are subsequently used as part of a sequencing-based pooled fitness assay to measure PPI scores.
fitness indicates that protein-mDHFR-fragment pairs within that strain interact to generate complete and functional mDHFR reporter proteins that, in turn, allow the strain to grow faster in the presence of low amounts of MTX.

To robustly calculate the fitness of each trajectory, we use a maximum likelihood strategy under a noise model that accounts for experimental errors (Supplementary Note 3)\(^22\). To make fitnesses comparable between replicates, or across different barcode pools or environments, we define a strain’s fitness relative to the control strain that lacks any mDHFR fragments, whose fitness is set to zero. We find that this procedure performs extremely well on simulated data with parameters similar to our pooled growth experiments (Pearson’s \(r = 0.996\), Fig. 2b and Supplementary Note 3 and 4), and across replicate growth experiments (Pearson’s \(r > 0.91\) between all MTX(+) replicates, Fig. 2c). Fitness estimates are generally more accurate for higher fitness strains (those putatively identifying a PPI) because these trajectories are unlikely to fall to low frequencies where counting noise of sequencing reads will be high.

We compared the fitness for each PPI across all \(\sim 75\) replicate estimates (\(\sim 25\) double barcodes per PPI, \(3\) replicate growth experiments) in the presence or absence of MTX (Fig. 3 and Methods). Standard errors on fitness are low (typically, s.e.m. < 0.05 in MTX(+)\(^+\)), Supplementary Fig. 3), with higher fitness PPIs having the lowest errors (s.e.m. < 0.02 in MTX(+)\(^+\) for PPIs with fitness > 0.07). We next compared the fitness values of each PPI against the fitness values of the control strains lacking mDHFR in both MTX(+)\(^+\) and MTX(−)\(^−\) conditions (Fig. 3 and Supplementary Data 2). As expected in MTX(−)\(^−\), almost none of the strains differ significantly in fitness from the control. The single exception is Prs3-F\([1,2]\):Fpr1-F\([3]\), which displayed a small but highly significant fitness advantage (fitness = 0.04, \(P\) value \(< 2 \times 10^{-6}\), Bonferroni corrected one-sided Student’s t-test) that

![Figure 2](https://www.nature.com/naturecommunications/)

**Figure 2** | Lineage tracking and fitness estimation of double barcodes. (a) The frequency trajectories of 2,500 double barcoded PCA strains in the absence or presence of 0.5 \(\mu\)g ml\(^−1\) methotrexate (MTX). Frequencies are assayed every three generations during serial batch growth. Colour indicates the estimated fitness relative to strains in the same pool that lack mDHFR fragments. (b) Performance of fitness estimates on simulated data. Pearson’s \(r = 0.996\). (c) Reproducibility of fitness estimates across growth replicates. Pearson’s \(r > 0.93\) in MTX.

![Figure 3](https://www.nature.com/naturecommunications/)

**Figure 3** | PPiSeq performance. Top: Relative fitnesses of each protein fragment pair grown in the absence (black) or presence (purple) of MTX. Each protein fragment pair is assayed with 25 unique double barcodes across \(3\) growth replicates for a total of \(\sim 75\) fitness estimates. Asterisks indicate the mean fitness of the protein fragment pair in MTX across all measurements and PPIs are ranked according to this fitness. Bottom purple: Heat map of the significance of the fitness difference between each protein fragment pair and control strains in the same pool that lack mDHFR fragments. \(P\) values are calculated using a Bonferroni-corrected Student’s t-test. Bottom grey: the number of times each protein-protein interaction has previously been cited. Biogrid is the sum of all forms of evidence: protein fragment complementation (PCA), yeast two-hybrid (YTH), pull down/mass spectroscopy (Pulldown), and low-throughput studies (Literature).
is perhaps due to an adaptive mutation that occurred in the parental PCA strain before barcoding. After removing the Prs3:Fpr1 strain from consideration, we find 11 significant PPIs in MTX(+), 10 that have been previously identified	extsuperscript{4,23}, and one that is new, Ftr1:Pdr5 (fitness = 0.10, \textit{P} value < 0.002, Supplementary Data 2). We validated Ftr1:Pdr5 by two additional assays. First, we tracked the optical density (OD600) of Ftr1:Pdr5 PPiSeq strains and the mDHFR(−) control strains grown in isolation in MTX(+) media	extsuperscript{17,18} and found that Ftr1:Pdr5 strains rise in optical density faster (P value < 2 × 10\textsuperscript{-11}, Student’s \textit{t}-test, Supplementary Fig. 4)	extsuperscript{18}. Second, we performed a less sensitive Renilla luciferase (Rluc) PCA	extsuperscript{24} and found that Ftr1:Pdr5 has a consistently higher (but not significant) luminescence when compared to control cells (P value = 0.24, Student’s \textit{t}-test). As discussed below, the Rluc PCA finds a significant Ftr1:Pdr5 interaction in an alternative environment (P value < 0.02 in 200 \textmu M copper sulfate, Student’s \textit{t}-test), strongly suggesting that our finding in this permissive environment is not a false positive.

Our PPiSeq assay missed five putative PPIs that had been discovered by traditional PCA	extsuperscript{4}. Three (Shr3:Hxt1, Tpo1:Snq2, and Fmp45:Pdr5) showed elevated but not significant fitness increases in MTX(+) (0.10, 0.08, and 0.06, respectively). As discussed below, PPiSeq does find all of these interactions to be significant in at least one perturbation environment, suggesting that these PPIs are sensitive to the environment and that environmental differences between PPiSeq and traditional PCA may impact their detection. The remaining two PPIs (Fmp45:Snq2 and Tpo1:Shr3) could not be detected by PPiSeq in any environment, but could be validated as being PPIs using isolated growth and optical density tracking over 32 h of growth (Supplementary Fig. 5). Notably, differences in optical density between Tpo1:Shr3 and control strains only began to appear around 25 h of growth, likely caused by a change in Tpo1 localization following the diauxic shift	extsuperscript{25,26}, suggesting that our current 24 h growth-bottleneck regime is not sensitive to PPIs that are specific to this later growth phase and that longer growth-bottleneck cycles may capture additional PPIs.

Overall, the ability of PPiSeq to detect PPIs appears to be on par with existing PPI assays; in this test set, PPiSeq discovered 10 of 15 PPIs that have been described by other assays, 1 new PPi validated here, and no additional spurious PPIs. When considering other environments, PPiSeq discovered 14 of 16 PPIs that have been discovered here or elsewhere. However, in contrast with previous high-throughput assays, detected PPIs span a reproducible range of positive fitnesses. Growth rate of PCA strains in MTX has previously been found to correlate with the number of functional mDHFR molecules per cell	extsuperscript{17,18}, suggesting that fitness differences in our assay are founded in differences in the abundance, localization, or binding of the interacting proteins.

PPiSeq detects dynamic PPIs. One advantage of using a pooled growth and bar-seq approach for detecting PPIs is that, once a barcoded PCA pool is constructed, it is trivial to re-test the entire interaction space across perturbations in order to detect PPIs that are dynamic. Here, we grew the pool of 2,500 PPiSeq strains in triplicate in MTX(−) and MTX(+) media supplemented with one of four additional perturbagens: 0.001% hydrogen peroxide (oxidative stress), 175 mM sodium chloride (high salt), 200 \textmu M copper sulfate (high copper), and 50 \textmu M of FK506, an inhibitor of calcineurin function in yeast. We calculated the fitness of each strain in each environment relative to the mDHFR(−) control strain using the maximum likelihood strategy described above. As expected, we found major fitness differences between strains within each MTX(+) environment, but not within the MTX(−) environments (Fig. 4, Supplementary Fig. 6, and Figure 4 | Dynamic PPIs. (a) Heatmap of PPIs across environments. All PPIs discovered here or elsewhere in standard conditions are shown. Colors are the fitness in each condition minus the fitness in the permissive condition. Cells are arranged by unsupervised hierarchical clustering. (b) PPI network plots of PPIs across five environments. Proteins that only interact with self are omitted. Colours are as in a. Edge width is proportional to the fitness and only significant edges are shown. PPIs expected under the standard condition are listed in a. (c) Barplots of the log ratio of the interaction score of a perturbation over the interaction score in the permissive environment as detected by three assays: PPiSeq (dark brown), split mDHFR clonal growth dynamics (light brown), and split Renilla luciferase luminescence (grey). Error bars are s.e.m. *, \textit{p}<0.05. Student’s \textit{t}-test against the permissive environment.
Supplementary Data 4). Surprisingly, 86% of detected PPIs significantly changed in fitness in at least one perturbation relative to the permissive environment (12 of 14, \( P < 0.05 \), Bonferroni corrected Student’s \( t \)-test, Supplementary Data 4) and 50% were undetectable by our assay in at least one environment (7 of 14, \( P > 0.05 \), Bonferroni corrected one-sided Student’s \( t \)-test, Supplementary Data 5). To validate these changes, we next selected 16 PPI-environment combinations where fitness was significantly different from the permissive environment, and assayed each by both optical density tracking and Rluc PCA (Fig. 4c). We found that 9 of 16 dynamic PPIs could be validated by at least one method.

A number of factors appear to underlie PPI changes across environments. One expected change is the interaction between the aspartate kinase Hom3 and the peptidyl-prolyl cis-trans isomerase Fpr1 in FK506, which has been previously found to physically disrupt this interaction27. Consistent with our previous work18, we find that the fitness of the Hom3:Fpr1 PPiSeq strain is diminished \( \sim 10\)-fold in FK506 (\( P < 10^{-59} \), Supplementary Data 3 and 4). Other dynamic PPIs appear to be due, at least in part, to changes in protein expression. For example, many of the changes that we detect—increased fitness of Tpo1:Pdr5 and Tpo1:Sno2 in FK506, Ftr1:Hxt1 and Ftr1:Pdr5 in copper, and Hxt1:Fmp45 in high salt—co-vary with changes in mRNA expression of one or both interacting partners that are reported in the literature18,28–31. Still other dynamic PPIs may be due to changes in protein localization. For example, fitness of the Tpo1:Pdr5 PPiSeq strain decreases in high salt (\( P < 10^{-11} \), even though both TPO1 and PDR5 have been shown to have high mRNA expression (4.7- and 2.7-fold, respectively)32). This contradiction appears to be resolved by the finding that Pdr5, but not Tpo1, becomes depleted from the plasma membrane in high salt (Supplementary Fig. 7, Supplementary Note 5)33.

PPiSeq is scalable. We have previously shown that at least 500,000 uniquely barcoded strains can be tracked in parallel in a single cell pool22. Furthermore, we found that for the majority of barcodes, errors in frequencies are consistent with counting noise stemming from finite read depths, rather than some other factor in the experimental protocol (Supplementary Note 4). Given exponentially declining sequencing costs34, it is therefore possible that several million double barcodes could be assayed in parallel. We reasoned that in order for our PPiSeq platform to reach these scales, two criteria must be met. First, PPiSeq must be capable of generating a large number of double barcode strains by pooled mating. Although it is technically possible to probe extremely large interaction spaces by pairwise mating in ordered arrays generating a large number of double barcode strains by pooled mating, we developed a protocol that mates \( \sim 10^{10} \) haploids on a standard agar plate, and then selects for diploid double barcode recombinants (Supplementary Methods). On the basis of experimental tests, we estimated the lower bounds of the frequency of mating (8%) and loxP recombination (2%) of this protocol, and predicted that at least \( 2 \times 10^7 \) (i.e. \( 10^9 \times 8.1\% \times 2.7\% \)) unique double barcoded diploids are generated per plate (Fig. 5a and Supplementary Note 6). Based on this performance, we estimate that double barcode library sizes exceeding \( 10^9 \) could be achieved by a single investigator (\( \sim 50 \) mating plates).

We next compared the initial double barcode frequency distribution of a large bulk mating (\( \sim 1 \) million double barcodes possible across 5 mating plates) to the smaller pairwise mating we used to generate the PPiSeq strains above (2,500 double barcodes possible), and found that the two protocols resulted in similar barcode frequency distributions (Fig. 5b). At an average sequencing depth of \( \sim 67 \) reads/barcode (Supplementary Note 7), bulk and pairwise mating protocols detect a similar number of double barcodes at low and moderate frequencies (\( > 98\% \) at \( > 1 \) read, \( > 95\% \) at \( > 10 \) reads), suggesting that even moderate read depths will be sufficient to characterize most double barcodes in the pool.

Discussion
We describe a highly parallel PPiSeq assay that is sensitive, accurate, and graded. Importantly, PPiSeq provides a quantitative score (fitness) for each PPI that is robust to changes in the environment or pool constituents. To probe larger interaction sets with PPiSeq, large barcode libraries must first be generated and mated to individual PCA strains. However, we have recently shown that libraries of thousands of barcoded yeast can be generated in parallel using our iSeq platform and Illumina sequencing35, and these barcodes can be easily added to PCA strains using the synthetic genetic array technology11,36. Furthermore, once barcoded PCA libraries have been generated, both double barcode library construction and fitness assays can be performed in large cell pools. PPiSeq is therefore a powerful new platform for large-scale investigations of dynamic PPIs.

The growth of each PCA strain is known to correlate with the number of reconstituted mDHFR reporter proteins per cell15,17,37, which, in turn could be influenced by several factors including the abundance of each interacting protein, the binding affinity, and the extent of co-localization of each binding pair. Protein abundances appear to have a large influence on fitness. For the 16 PPIs in our test set, fitness correlates reasonably well with the abundance of the least abundant interaction partner (Spearman’s \( \rho = 0.68 \), Supplementary Fig. 8)38. Additionally, many of the changes in fitness across environments that we detect
co-variation in mRNA expression of one or both interacting partners that are reported in the literature. However, other factors are likely to be important as well. For example, a recent proteome-wide screen found that nearly as many proteins change in localization as change in abundance when cells are exposed to hydroxyurea (28 and 40, respectively)\(^1\). In our test set, we find one example where a change in localization appears to be driving a PPI change (Tpo1:Pdr5). However, we do caution that an unbiased and systematic characterization of the factors underlying dynamic protein interaction will require combining PPiSeq with genome-scale mRNA abundance, protein abundance, and protein localization studies under the same conditions.

For cells treated with FK506, PPiSeq not only detects a change in the PPI target of the drug, Hom3:Fpr1, but also changes in the PPI target of the drug, Hom3:Fpr1, but also changes in PPiSeq with genome-scale mRNA abundance, protein abundance, and protein localization studies under the same conditions. The method of evidence: Protein Fragment Complementation (PCA)\(^4,18,44\), Yeast Two Hybrid (YTH)\(^1\),\(^2\), Affinity Pull-Down Assays (Pulldown)\(^5\), and other lower-throughput methods in the literature (Literature).

**Significance test for dynamic PPIs.** The fitness of each double barcode strain in each environment was determined as described in Supplementary Note 3. Fitnesses for a given PPI were compared across environments using a two-sided Student’s \(t\)-test Bonferroni corrected for 400 tests.

**PPi scoring by isolated growth optical density dynamics.** Haploid PCA strain\(^4\) were streaked from frozen stocks onto YPD to recover isolated colonies. MATa PCA strains harbouring BAIT-DHFR-F[1,2]-NatMX were mated one-to-one to MAT\(\alpha\) PREY-DHFR-F[3]-HphMX PCA strains in YPD liquid media. A control diploid strain that lacks DHFR was generated by mating a barcoded MATa hoc:NatMX strain with a barcoded MAT\(\alpha\) hoc:HphMX strain. Following 12 h of mating, cells were plated onto YPD + nourseothricin + hygromycin B agar and grown for 48 h at 30°C to select for diploids. One colony of each diploid was inoculated into YPD + nourseothricin + hygromycin B liquid media, grown for 12 h at 30°C, and then stored in 15% glycerol at −80°C. Cells from each strain were streaked from frozen stocks onto YPD and grown for 48 h at 30°C. Three isolated colonies of each strain were then streaked in sterilized water and counted. For each replicate, 6.4 × 10\(^5\) cells were inoculated into 150 μl of media in black-walled, clear-bottom 96-well plates (Nunc \#265301). Media was synthetic dextrose supplemented with standard concentrations of the amino acids histidine, leucine, and uracil, plus methotrexate (0.5 μg ml\(^−1\)) and one of the following perturbagens: DMSO (final at 0.5%), FK506 (final at 50 μM), hydrogen peroxide (final at 0.001%), sodium chloride (final at 175 mM), or copper sulfate (final at 200 μM). Plates were sealed with foil (Costar \# 6570) and shaken at 1,300 r.p.m. (DTS4, Elmi) at 30°C. 96-well plates were stored at 600 μM of DMSO. The reporter colour was recorded (F500, Tecan) at 0, 8, 10, 12, 14, 16, 18, 20, 22, 24, and 32 h. The area under the curve (AUC) was calculated as the sum of all OD readings before saturation (32h) for each strain in each environment. The relative fitness for a strain in a specific condition was quantified with following equation: AUC\(_\text{target strain}\)/AUC\(_\text{control strain}\)\(\times\)100.

**Luciferase protein fragment complementation assay.** To construct Renilla luciferase (Rluc) PCA strains, we replaced the DHFR fragments with Rluc PCA fragments in haploid DHFR PCA strains\(^6\) via homologous recombination. The Rluc-F[1]-NatMX homologous recombination cassette was PCR amplified from the pAG32-linker-Rluc-F[1]-NatMX plasmid\(^2\), and the Rluc-F[2]-HphMX cassette was PCR amplified from the pAG32-linker-Rluc-F[2]-HphMX plasmid\(^4\). We then replaced the two amplification products with the same sequence upstream of the stop codon. The forward primer (5\‘-GGGGTGGTGGGTACACTAGAG GCCG-3\‘) anneals to the linker sequence in pAG32-linker-Rluc-F[1]-NatMX, or pAG32-linker-Rluc-F[2]-HphMX. The reverse primer (5\‘-TTCGACACTGGATG -GGCGGTGGCGGATCAGGA CCG-3\‘) anneals to the 3\‘-end of the TEF terminator region of NatMX or HphMX. To increase the recombination efficiency for some genes, it was necessary to add an additional 40 bp to the forward primer that matches gene specific sequence upstream of the stop codon. In all cases, MATa PCA DHFR-F[1,2]-NatMX strains were transformed with the Rluc-F[2]-HphMX cassettes and MATa PCA DHFR-F[3]-HphMX strains were transformed with the Rluc-F[1]-NatMX cassettes. Transformants were selected by plating on YPD plus the appropriate antibiotic, and proper incorporation of the Rluc PCA cassette was validated by PCR. Next, MATa PCA strains harbouring BAIT-Rluc-F[1]-NatMX were mated one-to-one to MAT\(\alpha\) PREY-Rluc-F[2]-HphMX strains in YPD liquid media. Following 12 h of mating, cells were plated onto YPD + nourseothricin + hygromycin B agar and grown for 48 h at 30°C to select for diploids. One colony of each diploid was inoculated into YPD + nourseothricin + hygromycin B liquid media, grown for 12 h at 30°C, and then stored in 15% glycerol at −80°C. Cells from each strain were streaked from frozen stocks onto YPD and grown for 48 h at 30°C. Three isolated colonies of each strain were then streaked in sterilized water and counted. For each replicate, 6.4 × 10\(^5\) cells were inoculated into 150 μl of media in black-walled, clear-bottom 96-well plates (Nunc \#265301). Media was synthetic dextrose supplemented with standard concentrations of the amino acids histidine, leucine, and uracil, plus methotrexate (0.5 μg ml\(^−1\)) and one of the following perturbagens: DMSO (final at 0.5%), FK506 (final at 50 μM), hydroperoxide (final at 0.001%), sodium chloride (final at 175 mM), or copper sulfate (final at 200 μM). Plates were sealed with foil (Costar \# 6570) and shaken at 1,300 r.p.m. (DTS4, Elmi) at 30°C. The area under the curve (AUC) was calculated as the sum of all OD readings before saturation (32h) for each strain in each environment. The relative fitness for a strain in a specific condition was quantified with following equation: AUC\(_\text{target strain}\)/AUC\(_\text{control strain}\)\(\times\)100.

**Double barcode sequence analysis.** Barcode reads were processed with custom written software in Python and R as described\(^2\), with modifications. Briefly, sequences were parsed to isolate the two barcode regions (38 base pairs each), sorted by their multiplexing tags (see above), and removed if they failed to pass any of three quality filters: (1) The average Illumina quality score for both barcode regions must be greater than 30, (2) the first barcode must match the regular expression G.TA|GG.A|GGT.), and (3) the second barcode must match the regular expression D*\(\{4,7\}\)?AA/D*\(\{4,7\}\)?TT/D*\(\{4,7\}\)?TT/D*\(\{4,7\}\)?TGTA/G.TA/G.GA(GOT.) D*. A BLAST database containing all expected double barcode sequences and each restriction (76 bases each) was created and each read was blasted to size = 11, reward = 1, penalty = −2 against this database. Double barcode reads that blasted at an e < 10−28 (~2 mismatches) to an expected double barcode were summed to calculate an initial estimate of the read database for each double barcode in each condition.

**Comparisons to existing PPI studies.** Interaction data was downloaded from the Biogrid on 1 December 2013 (S. cerevisiae version 3.4.131)\(^23,43\). PPIs we sorted based on the form of evidence: Protein Fragment Complementation (PCA),\(^4,18,44\), Yeast Two Hybrid (YTH),\(^1\),\(^2\), Affinity Pull-Down Assays (Pulldown)\(^5\), and other lower-throughput methods in the literature (Literature).
Triplicate fresh colonies of each diploid Bluc PCA strain were grown in 5 ml synthetic dextrose media supplemented with standard concentrations of histidine, leucine, and uracil at 30 °C for 24 h, then diluted 1:3 to 5 ml of the same media supplemented DMSO (0.5%), FK506 (50 µM), hydrogen peroxide (0.001%), sodium chloride (175 mM), or copper sulfate (200 µM). Cells were grown for 24 h at 30 °C, diluted 1:3 again into fresh media containing the same supplement, and grown for another 30 h. Cells were counted. 1–2 × 107 cells were pelleted and resuspended in 180 µl phosphate-buffered saline (PBS), pH 7.2 containing 1 mM EDTA. Cells were transferred to white 96-well flat bottom plates (Greiner bio-one # 655075). The luciferase substrate, benzyl coelenterazine (NanoLight # 301), was diluted 1:10 from the stock (2 mM in absolute ethanol) using 1 × PBS, and 20 µl of diluted substrate was added to each sample (to a final concentration of 20 µM). A Centro LB 960 microplate luminometer (Berthold Technologies) was used to measure the Bluc PCA signal, which was integrated for 10 s. Changes in luminescence in response to a specific condition were calculated by the following equation: luminescencecondition/luminescenceDMSO.

**References**

1. Cesick, M. E., Kligendorf, N., Vidal, M. & Hill, D. E. Interactome: gateway into systems biology. *Hum. Mol. Genet.* 14, R171–R181 (2005).
2. Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* 98, 4569–4574 (2001).
3. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627 (2000).
4. Tarassov, K. et al. An in vivo map of the yeast protein interactome. *Science* 320, 1465–1470 (2008).
5. Yu, H. et al. High-quality binary protein interaction map of the yeast interactome network. *Science* 322, 104–110 (2008).
6. Stelzl, U. et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122, 957–968 (2005).
7. Spirin, V. & Mirny, L. A. Protein complexes and functional modules in molecular networks. *Proc. Natl Acad. Sci. USA* 100, 12123–12128 (2003).
8. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell’s functional organization. *Nat. Rev. Genet.* 5, 101–113 (2004).
9. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461 (2007).
10. Ideker, T. & Krogan, N. J. Differential network biology. *Mol. Syst. Biol.* 8, 656 (2012).
11. Baryshnikova, A. et al. in: Guide to Yeast Genetès: Functional Genomics, Proteomics, and Other Systems Analysis, chapter 7, 470–145–179 (Elsevier Inc., 2010).
12. Yachie, N. et al. Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* 12, 863–863 (2016).
13. Mering, V., Von, C., Krause, R., Snel, B., Cornell, M. & Oliver, S. G. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403 (2002).
14. Remy, I. & Michnick, S. W. Visualization of biochemical networks in living cells. *Proc. Natl Acad. Sci. USA* 98, 7678–7683 (2001).
15. Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I. & Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discov.* 6, 569–582 (2007).
16. Baryshnikova, A. et al. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 7, 1017–1024 (2010).
17. Freschi, L., Torres-Quiroz, F., Dubé, A. K. & Landry, C. R. qPCA: a scalable assay to measure the perturbation of protein–protein interactions in living cells. *Mol. Biosyst.* 9, 36–43 (2013).
18. Schlegel, U., Miranda, M., Suresh, S., Davis, R. W. & St Onge, R. P. Multiplex assay for condition-dependent changes in protein-protein interactions. *Proc. Natl Acad. Sci. USA* 109, 9213–9218 (2012).
19. Smith, A. M. et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res.* 19, 1836–1842 (2009).
20. Gresham, D. et al. System-level analysis of genes and functions affecting survival during nutrient starvation in *Saccharomyces cerevisiae*. *Genetics* 187, 309–317 (2011).
21. Gibney, P. A., Lu, C., Caudy, A. A., Hess, D. C. & Botstein, D. Yeast metabolic and signaling genes are required for heat-shock survival and have little overlap with the heat-induced genes. *Proc. Natl Acad. Sci. USA* 110, E4393–E4402 (2013).
22. Levy, S. E. et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519, 181–186 (2015).
23. Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539 (2006).
24. Malleshaiah, K. M., Shahrnaz, V., Swan, P. S. & Michnick, S. W. The scaffold protein Ste5 directly controls a switch-like mating decision in yeast. *Nature* 465, 101–105 (2010).
25. Albertsen, M., Bellahn, J., Kämmer, R. & Waffenschmidt, S. Localization and function of the yeast maltodextrin transporter Tpo1p. *J. Biol. Chem.* 278, 12820–12825 (2003).
26. Uemura, T., Taihachira, K., Tomitori, H., Kashiiwagi, K. & Igarashi, K. Characteristics of the polyanime transporter TPO1 and regulation of its activity and cellular localization by phosphorylation. *J. Biol. Chem.* 280, 9646–9652 (2005).
27. Arevalo-Rodriguez, M., Wu, X., Hanes, S. D. & Heitman, J. Prolyl isomerases in yeast. *Front. Biosci.* 9, 2420–2446 (2004).
28. Gross, C., Kelleher, M., Iyer, V. R., Brown, P. O. & Winge, D. R. Identification of the copper regulator in *Saccharomyces cerevisiae* by DNA microarrays. *J. Biol. Chem.* 275, 32310–32316 (2000).
29. Yasokawa, D. et al. Mechanisms of copper toxicity in *Saccharomyces cerevisiae* determined by microarray analysis. *Environ. Toxicol.* 23, 599–606 (2008).
30. Hirayama, T., Maeda, T., Saito, H. & Shinozaki, K. Cloning and characterization of seven cDNAs for hyperosmolarity-responsive (HOR) genes of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 249, 127–138 (1995).
31. Greatrix, B. W. & van Vuuren, H. J. H. Expression of the HXT13, HXT15 and HXT17 genes in *Saccharomyces cerevisiae* and stabilization of the HXT1 gene transcript by sugar-induced osmotic stress. *Curr. Genet.* 49, 205–217 (2006).
32. Vale, Y. & Rohnert, H. J. Transcription Expression in *Saccharomyces cerevisiae* at high salinity. *J. Biol. Chem.* 276, 15996–16007 (2001).
33. Szopinska, A., Degand, H., Hochstenbach, J.-F., Nader, J. & Morsomme, P. Rapid response of the yeast plasma membrane proteome to salt stress. *Mol. Cell Proteomics* 10, mcp.M111.009589 (2011).
34. Stephens, Z. D. et al. Big Data: astronomical or genomical? *PLoS Biol.* 13, e1002195 (2015).
35. Smith, J. D. et al. A method for high-throughput production of sequence-verified DNA libraries and strain collections. *Mol. Syst. Biol.* 13, 913 (2017).
36. Tong, A. H. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
37. Remy, I. & Michnick, S. W. Clonal selection and *in vivo* quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl Acad. Sci. USA* **96**, 5394–5399 (1999).
38. Ghaemmaghami, S., Huh, W. K., Bower, K. & Howson, R. W. Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
39. Chong, Y. T. et al. Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell* **161**, 1413–1424 (2015).
40. Chang, R. L., Xie, L., Xie, L., Bourne, P. E. & Palsson, B. Ø. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* **6**, e1000938 (2010).
41. Hellerstein, M. K. Exploiting complexity and the robustness of network architecture for drug discovery. *J. Pharmacol. Exp. Ther.* **325**, 1–9 (2008).
42. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682–690 (2008).
43. Breitkreutz, B. J. et al. The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* **36**, D637–D640 (2008).
44. Miller, J. P. et al. Large-scale identification of yeast integral membrane protein interactions. *Proc. Natl Acad. Sci. USA* **102**, 12123–12128 (2005).
45. Ho, Y. et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
46. Krogan, N. J. et al. High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* **13**, 225–239 (2004).
47. Krogan, N. et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
48. Babu, M. et al. Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. *Nature* **489**, 585–589 (2012).

**Acknowledgements**

We are grateful to Stephen Michnick for providing the pAG25-linker-Rluc F[1]-NatMx and PAG32-linker-Rluc F[2]-HphMX plasmids and for technical advice, to Xianan Liu, Richard Bennett, Gavin Sherlock and Mia Jaffe for technical advice, to Daniel Fisher for advice on the analysis, and to Rolf Sterniglaz for suggestions on the manuscript. This work was supported by grants from the US National Institutes of Health (R01HG008354 and U01HL127522 to S.F.L.), The Louis and Beatrice Laufer Center, and the New York State Center for Biotechnology.

**Author contributions**

S.F.L. wrote the manuscript. U.S., Z.L., J.R.B. and R.P.S.T. contributed to writing the manuscript. U.S., Z.L. and S.F.L generated experiments. Z.L., J.R.B. and S.F.L analysed data. S.F.L conceived of and developed the double barcode platform. S.F.L and U.S. generated the barcoded PPISeq strains. U.S. developed and performed the PPISeq assays. U.S. and Z.L. performed the sequencing. Z.L. and S.F.L. analysed the sequencing data. J.R.B. developed the fitness estimation method. Z.L. developed and performed validation assays. Z.L. performed the scalability assays. S.F.L. and Z.L. performed statistics and visualized the data. R.P.S.T. provided essential insight and advice.

**Additional information**

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article:** Schlecht, U. et al. A scalable double-barcode sequencing platform for characterization of dynamic protein-protein interactions. *Nat. Commun.* **8**, 15586 doi: 10.1038/ncomms15586 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/