

A pooling-deconvolution strategy for biological network elucidation

Fulai Jin¹, Tony Hazbun^{2,4}, Gregory A Michaud³, Michael Salcius³, Paul F Predki³, Stanley Fields² & Jing Huang¹

The generation of large-scale data sets is a fundamental requirement of systems biology. But despite recent advances, generation of such high-coverage data remains a major challenge. We developed a pooling-deconvolution strategy that can dramatically decrease the effort required. This strategy, pooling with imaginary tags followed by deconvolution (PI-deconvolution), allows the screening of 2^n probe proteins (baits) in $2 \times n$ pools, with n replicates for each bait. Deconvolution of baits with their binding partners (preys) can be achieved by reading the prey's profile from the $2 \times n$ experiments. We validated this strategy for protein-protein interaction mapping using both proteome microarrays and a yeast two-hybrid array, demonstrating that PI-deconvolution can be used to identify interactions accurately with fewer experiments and better coverage. We also show that PI-deconvolution can be used to identify protein-small molecule interactions inferred from profiling the yeast deletion collection. PI-deconvolution should be applicable to a wide range of library-against-library approaches and can also be used to optimize array designs.

Understanding protein function on a genome-wide scale is one of the central goals of biology¹. A fundamental task associated with this goal is the elucidation of cellular functional and interaction networks. Recently, large-scale protein-protein interaction experiments, using yeast two-hybrid screens²⁻⁷ or affinity purification⁸⁻¹⁰ have provided critical insights into protein function and biological network structure. In addition to the determination of protein interaction networks, systems biology will require the elucidation of other interaction (for example, protein-small molecule, protein-nucleic acid) and functional (for example, protein phosphorylation) networks. Although the need to generate these types of data sets is obvious, our current ability to do so is inadequate.

The challenges of generating protein-protein interaction data sets serve as a good model for what will be required to obtain other interaction and functional network data sets. Because the generation of a complete interaction network one protein pair at a time is labor- (and material-) intensive, simpler alternatives have been

sought. The creation of spatially addressable proteome-wide screening platforms (such as the array of yeast two-hybrid strains² and proteome microarrays^{11,12}), for instance, allows an entire subject library to be screened at once and hits to be decoded without the need for DNA or protein sequencing. But thousands of array screens are required to cover a yeast-sized interactome, an even more daunting undertaking when replicate screens are needed to improve accuracy and coverage.

To date, pools of 8 baits or more have been used to screen yeast two-hybrid arrays^{6,13}. In these approaches, secondary small-scale screens are used to deconvolute the hits. A major weakness of this procedure is that final coverage completely depends on the primary pooling screen, which is often performed only once. Conversely, a primary screen with reduced stringency may generate too many false positives, creating a huge burden on the secondary screens. Furthermore, small-scale secondary screens may not be possible for some platforms such as protein microarrays. Because genome- and proteome-wide arrays have the physical capacity to detect far more interactions than those of a typical single protein, an alternative pooling strategy that allows prey-bait deconvolution is possible. The PI-deconvolution strategy we describe here reduces the number of screens needed, and increases accuracy and coverage. In addition, we suggest a new method to optimize array design using the same principle.

RESULTS

Description of the strategy

In PI-deconvolution (**Fig. 1**), 2^n baits are distinguished by their assigned n -bit binary codes, which are text strings consisting of '+' and '-' symbols (**Fig. 1b**). The baits are assigned to n pairs of experiments (each pair containing one '+' and one '-' experiment) corresponding to the binary bits (**Fig. 1c**). In each experiment pair, half of all the 2^n baits will be loaded to the '+' experiment pool and the other half to the '-' pool. Any single bait is used once and only once in each pair of experiments. Whether a bait is used for the '+' or '-' experiment in a pair is determined by its symbol in the coding string at the corresponding bit (**Fig. 1b,c**). For example, bait 6 is

¹Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, and the Molecular Biology Institute, University of California, Los Angeles, California 90095, USA. ²Howard Hughes Medical Institute (HHMI), Departments of Genome Sciences and Medicine, University of Washington, Box 357730, Seattle, Washington 98195, USA. ³Protein Microarray Center, Invitrogen Life Technologies, 688 East Main Street, Branford, Connecticut 06405, USA. ⁴Present address: Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47907, USA. Correspondence should be addressed to J.H. (jinghuang@mednet.ucla.edu), S.F. (fields@u.washington.edu) or P.F.P. (paul.predki@invitrogen.com).

RECEIVED 27 OCTOBER 2005; ACCEPTED 23 JANUARY 2006; PUBLISHED ONLINE 17 FEBRUARY 2006; DOI:10.1038/NMETH859

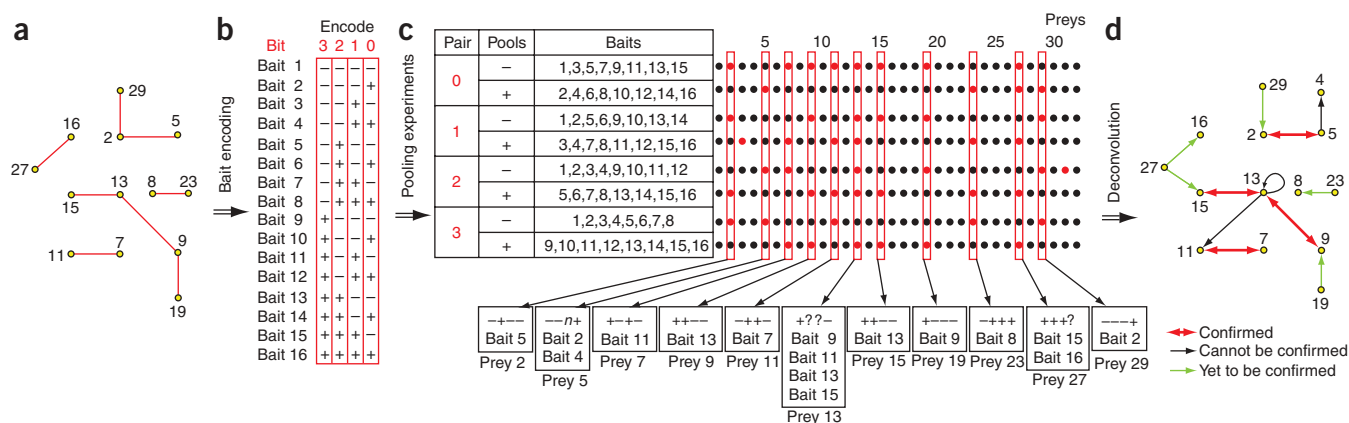


Figure 1 | Scheme for PI-deconvolution. **(a)** Graph representation of a hypothetical 32-protein network showing known interactions. Yellow filled circles, proteins (nodes); red lines, interactions (edges). For simplicity, only nodes and edges concerning proteins 1–16 are shown. **(b)** Proteins 1–16 are used as the first batch of baits to identify their preys. The total 32 proteins can be covered similarly with a second batch of experiments. We encode each bait with a 4-bit ‘+’ or ‘-’ string (imaginary coding tag); four bits are enough to uniquely encode 16 ($=2^4$) baits. Thus, n bits can encode 2^n distinct baits. **(c)** We prepare four pairs of bait pools numbered 3, 2, 1 and 0, corresponding to each of the 4 bits. Every pair contains a ‘+’ pool and a ‘-’ pool, each using 8 baits (half the batch size). Altogether, there will be 8 ($2n$) experiments (rows)—instead of 16 (2^n)—to identify all interacting preys. Each column represents the profile of a prey: positive signal, red, and negative signal, black. All valid preys (outlined columns) and their possible baits are listed. If a prey binds to only one bait in a batch, the prey should be detected only once in each pair of experiments. The profile of a prey denotes its readout in each of the experiment pairs. ‘+’ or ‘-’ at a given bit in the profile indicates that the prey is detected in the ‘+’ pool or the ‘-’ pool, respectively, of the corresponding experiment pair. Deconvolution is accomplished by matching a prey’s profile with the coding strings of the baits. We use degenerate profile ‘n’ or ‘?’ to indicate neither or both experiments in a pair give a positive call (such as prey 5 or prey 13). Preys with degenerate profiles can still be partially deconvoluted and further narrowing-down can be achieved by reciprocal confirmation. **(d)** A graphical representation of the result in **c**.

represented by string ‘- + - +’. At bit 2 (third digit from the right in the string), its symbol is ‘+’. Thus in pair 2, bait 6 is included in the ‘+’ experiment. This ‘+’ experiment also includes all the other baits with a ‘+’ sign in the column for bit 2. Each prey’s interacting bait(s) can be revealed by the prey’s profile in all the n pairs of experiments. For example, prey 2 binds to only bait 5 among baits 1–16. In **Figure 1c**, prey 2 is detected in four experiments, which are ‘-’ of pair 3, ‘+’ of pair 2, ‘-’ of pair 1 and ‘-’ of pair 0. Accordingly, prey 2 can be represented by the profile ‘- + - -’, denoting its readout in each of the four experiment pairs. As pair numbering corresponds to bit numbering in the tag of a bait, the prey’s profile can allow a direct track back to its own bait(s). In this case, the profile of prey 2 is identical to the bit tag for bait 5; thus, we know that prey 2 binds to bait 5. With this strategy, 2^n baits can be screened in $2 \times n$ arrays (**Table 1**). Pool size will be limited by the technical false negative and false positive rates (discussed below); the flexibility in setting different bit numbers (n) allows the strategy to be applied to different scenarios. Besides substantially decreasing the number of screens needed, a major advantage of PI-deconvolution is that all the baits are screened n times (**Table 1**), which allows cross-validation and thus improves both coverage and accuracy of data.

PI-deconvolution on proteome microarrays

We tested PI-deconvolution using yeast proteome microarrays^{11,12}, which contain 4,088 purified *Saccharomyces cerevisiae* proteins (as glutathione *S*-transferase fusions) immobilized on nitrocellulose-coated glass slides. For this purpose, we used 15 ($\sim 16 = 2^4$) V5 epitope-tagged bait proteins (**Fig. 2a**). By first probing the yeast proteome microarrays with each of these 15 proteins individually, we derived a small network of protein interactions (**Fig. 2a**). We consider this a ‘gold standard’ network, because all the interactions in the network have been reciprocally confirmed (**Fig. 2a**). We

prepared pools of 8 baits (**Supplementary Table 1** online) and used them to probe the microarrays. All interactions among the 15 proteins were detected and deconvoluted using PI-deconvolution with only eight proteome microarrays, and all hits were reproducibly detected four times (**Fig. 2a**). Furthermore, although there were several interacting protein pairs within the mixed pools of baits, they did not appear to affect detection or deconvolution. (Raw chip images are available online; <http://labs.pharmacology.ucla.edu/huanglab/webpub>.)

PI-deconvolution on yeast two-hybrid arrays

We demonstrated the utility of PI-deconvolution to screen a genome-wide two-hybrid array consisting of $\sim 6,000$ yeast strains, each designed to contain one of the $\sim 6,000$ *S. cerevisiae* open reading frames (ORFs) fused to the Gal4 activation domain². To this end, we used 16 two-hybrid bait strains that each express a full-length ORF fused to the Gal4 DNA-binding domain (**Fig. 2b**). Thirteen of these bait strains have previously been screened against

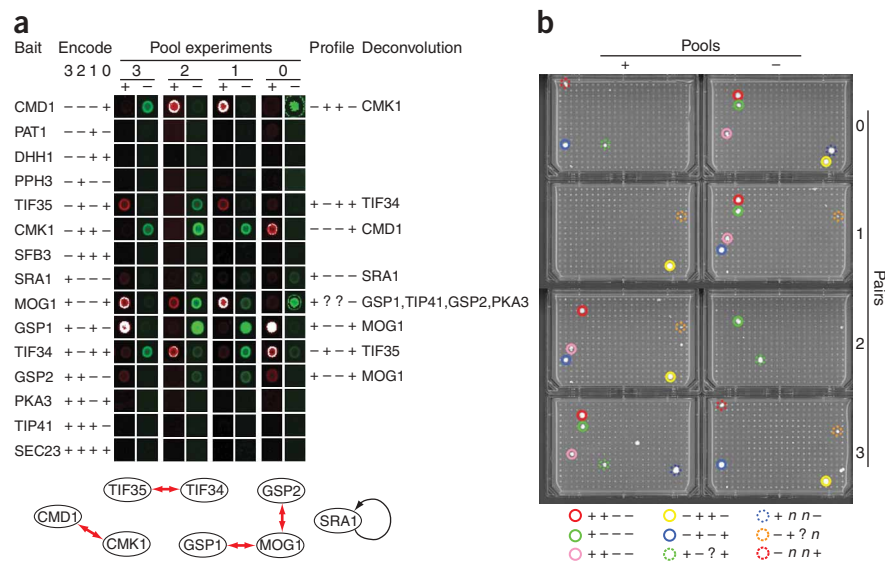
Table 1 | Number of probings required in PI-deconvolution

Bit number (n)	Batch size (2^n)	Pool size (2^{n-1})	Experiment number ($2n$)	Total number of experiments ($n \times N / 2^{n-1}$)	‘Depth of coverage’ ^a
4	16	8	8	$50.0\% \times N$	4
5	32	16	10	$31.3\% \times N$	5
6	64	32	12	$18.8\% \times N$	6
7	128	64	14	$10.9\% \times N$	7
8	256	128	16	$6.25\% \times N$	8
Conventional single-bait strategy				N	1

^N is the number of proteins in the genome.

^aNumber of times the whole genome is covered.

Figure 2 | PI-deconvolution applied to protein interaction mapping. **(a)** Yeast proteome microarray screening. Fifteen bait proteins are encoded as shown and eight bait pools are prepared accordingly (see also **Supplementary Table 1**). Each image column represents the result of a pooling screen, and each image row represents the same spot of the array. A positive signal indicates the presence of one or more binding proteins in the pool. Signals from '+' pools are false-colored red and '-' pools, green. For example, the prey spots representing CMD1 (top row) were positive when probed with the '+' pools of pairs 1 and 2 (red), and the '-' pools of pairs 0 and 3 (in green). The profile of CMD1 is thus read as '-++-', which equals the encoding tag for the bait CMK1. The results obtained by the PI-deconvolution analysis (using eight arrays) are identical to those obtained from single-bait probing (using 15 arrays). Only reciprocally confirmed interactions (red bidirectional arrows) and self interactions (black arrow) are shown (bottom). Detailed explanation of hit recognition is available in Methods. **(b)** Yeast two-hybrid array screening. Encoding and pooling schemes of 16 bait strains are shown in **Supplementary Table 2**. The whole library array consists of 16 plates with 384 strains each. Shown are images of one representative library plate screened with 16 baits using PI-deconvolution; each image is the result of a pooling screen with 8 baits.



the genome-wide array. Because of experimental variability, these single-bait screens required each bait to be screened in duplicate, resulting in a total of 32 screens for the 16 baits^{2,14}. For the PI-deconvolution format, the 16 bait strains were mixed into eight pools and screened against the two-hybrid array. In this procedure, two 8-bait pools in the same pair cover all of the 16 baits. Therefore, four pairs of PI-deconvolution screens represent four independent screens of all the 16 baits (**Supplementary Table 2** online). This protocol provides a considerable advantage over the individual bait procedure because it reduces the number of screens from 32 to 8, yet each bait is screened in quadruplicate. (Raw data are available online; http://depts.washington.edu/sfields/supplemental_data/pooled_2hybrid/.)

In the 13 single-bait screens, we observed 484 preys and defined them as two-hybrid positive colonies^{2,14}. Among these positive colonies, 125 arose twice out of the duplicate screens, and we termed them 'reproducible' positives; the other 359 arose once as either false positives, or true positives that owing to experimental variability did not yield reproducible results. Further testing is usually required to confirm or reject such nonreproducible hits (for a complete list of all single bait-screen hits, see **Supplementary**

Table 3 online). For the PI-deconvolution screens, each pair of pools identified 153–189 hits, and in total 343 positive colonies were identified (**Table 2**). The number of 'reproducible hits' between any two independent experiment-pairs ranged from 103 to 112. Conversely, as many as 155 positive colonies were reproducible in at least two experiment pairs (40–50% higher than for any two experiment pairs only; **Table 2**). This result suggests that, as expected, a higher level of repetition indeed improves coverage. We consider all 155 as reproducible positives from PI-deconvolution. A complete list of all PI-deconvolution hits and deconvolution results can be found in **Supplementary Table 3**.

Further evidence for improved coverage by PI-deconvolution is provided by comparing the PI-deconvolution data with the single-bait data. The 155 reproducible hits from the PI-deconvolution data set include 26 (out of 359; 7%) 'nonreproducible positives' found in the single-bait screens, in addition to recapitulating 71 (of 125; 57%) reproducible single-bait positives (**Table 3**). This result suggests that in the single-bait data set, ~30% (26 of 97) of the true positives were not found as reproducible hits. The remaining 58 PI-deconvolution reproducible positives are either previously unidentified interaction partners for these 13 baits or belong to the 3 baits not screened before. In contrast, the 188 PI-deconvolution nonreproducible hits contain only 2 (of 125; 1.6%) reproducible hits from the single bait data (**Table 3**), suggesting that the PI-deconvolution 155 reproducible hits might represent almost complete (saturated) coverage, subject to the detection sensitivity of the current system. The increased coverage is due to the high repetitions inherent in PI-deconvolution screening.

Of the 155 PI-deconvolution reproducible positives, 57 could be assigned to a single bait (**Supplementary Table 4** online), 34 to two baits and 51 to four possible baits (see below about further deconvolution of positives assigned to more than one bait). Of the 57 unambiguously deconvoluted preys, 56 belong to the 13 previous screened baits. Of these 56, 38 were previously classified as reproducible positives in single-bait screens; 11 were previously

Table 2 | Reproducibility over PI-deconvolution experiment pairs

Overlapped hits	Pair 0	Pair 1	Pair 2	Pair 3	All pairs
Pair 0		103	107	109	
Pair 1	103		110	110	
Pair 2	107	110		112	
Pair 3	109	110	112		
Total number of hits	153	189	179	180	343
Number of hits overlapping with at least one other pair	123	127	130	133	155
Number of hits having no overlap with other pairs	30	62	49	47	188

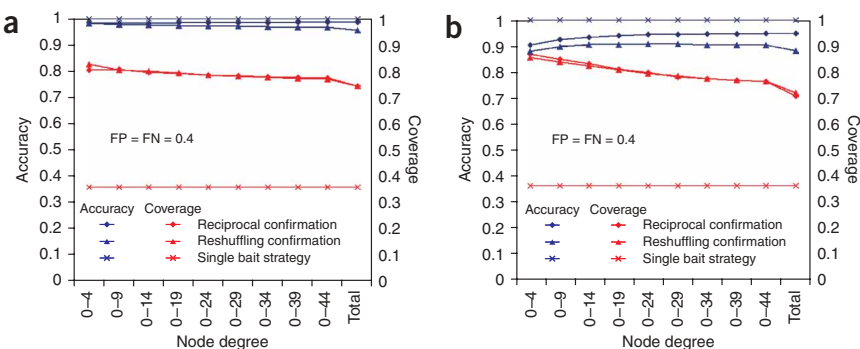


Figure 4 | PI-deconvolution simulated on yeast interactome (DIP). We assume that all interactions in DIP are correct. Only the interactions that are confirmed reciprocally or after bait-reshuffling are accepted. Both accuracy (fraction of true interactions among the detected ones; blue) and coverage (fraction of true interactions detected; red) are calculated for interactions between proteins with different node degrees (x axis). For example, blue and red curves with x axis value 0–24 showed accuracy and coverage of PI-deconvolution for interactions between nodes with degrees no more than 24. The curves with ‘x’ indicate the accuracy and coverage by duplicated single bait screening, where only reproducible hits are accepted. We assume 40% of the hits from each array experiment are false positives (FP), and each array experiment will lose 40% of true positives (FN). **(a)** PI-deconvolution simulated using BitNum = 5 ($n = 5$; 16 baits per pool) and Cut0cc = 2 (only hits that show up at least twice are accepted). **(b)** PI-deconvolution simulated using BitNum = 6 ($n = 6$; 32 baits per pool) and Cut0cc = 2.

reproducible hits were not found by the single-bait screens (Table 3). Thus, it does not appear that the single bait data set has a better coverage than the PI-deconvolution data set. These results suggest that intrinsic yeast two-hybrid variability, rather than loss of sensitivity owing to pooling, likely underlies the discrepancy between the two data sets.

PI-deconvolution in identifying drug-resistant mutants

The primary advantage of PI-deconvolution is that it increases screening efficiency by making better use of the physical capacity of whole-proteome platforms for parallel detection. We tested the PI-deconvolution approach on an assay independent of protein interaction mapping, namely the identification of yeast mutants resistant to specific drugs. The *S. cerevisiae* deletion collection is a set of ~4,500 strains, each deleted for one of the nonessential ORFs^{16,17}. We assayed 128 ($n = 7$) strains from this collection for fitness response to rapamycin, which targets the Tor proteins¹⁸, and wortmannin, which is a phosphatidylinositol 3-kinase inhibitor¹⁹. Using 14 ($= 2 \times 7$) pools each containing 64 ($= 2^6$) strains (see Supplementary Table 6 online), the PI-deconvolution approach deconvoluted the two strains previously known to be resistant to rapamycin (*fpr1Δ*) or to wortmannin (*ppg1Δ*; Fig. 3), a screening efficiency that is an order of magnitude higher than when single strains were used. Although higher efficiency could be obtained by setting a higher n value (that is, screening a larger pool) in PI-deconvolution, acceptable pool size is also determined by the sensitivity and background of the detection method (as is true for any pooling strategy). In addition, because pooled screening generally relies on the gain of a signal, drug hypersensitivity cannot be scored in a pooling screen using fitness as a readout.

In silico simulation on DIP

To predict the performance of PI-deconvolution in an actual large-scale network, we performed a series of *in silico* experiments to

simulate the performance of PI-deconvolution on the Database of Interacting Proteins (DIP)²⁰. DIP contains 4,716 proteins and 14,848 interactions, which is close to the estimation that a typical protein binds to 3–10 other proteins^{21,22}. The simulation shows that even with 16 or 32 baits per pool, both reciprocal and reshuffling methods can efficiently clarify the ambiguous profiles under a variety of experimental false positive and false negative conditions (Fig. 4). (For additional simulations, see Supplementary Note.) Notably, coverage after reciprocal or reshuffling confirmation is still considerably higher than for single-bait screens, especially for interactions between low-degree nodes (Fig. 4). Our simulation also shows that PI-deconvolution can be applied to random networks as well (see Supplementary Note).

PI-deconvolution for compressing arrays

PI-deconvolution is a robust alternative experimental design that will save time and resources in proteome-wide protein-

protein interaction mapping. Although we describe this strategy as a method to pool baits in a combinatorial fashion, the idea of PI-deconvolution can be also used to redesign prey arrays and maximize the efficiency of single-bait screens. For example, the yeast two-hybrid array consists of 16 plates, each containing 384 Gal4 activation domain fusion strains. Because we have demonstrated that 8-bait pools can be used to screen the prey array, it should also be possible to screen a single bait against pools of eight prey strains. Thus, the 16 plates can be compressed into 8 plates using the PI-deconvolution scheme (eight activation domain fusion strains per well). This compressed library can be maintained and screened against single baits, equivalent to screening the original 16-plate array in quadruplicate (total reduction to 12.5%).

DISCUSSION

Pooling and deconvolution designs have been of great interest to various fields^{23,24} (see also Supplementary Note). Although 8-bait or larger pools^{6,13} have been used for screening yeast two-hybrid arrays, two major concerns exist. First, a considerable number of true positives will be lost as false negatives, because only the hits that pass the initial pool screening (usually performed only once) will go to the secondary deconvolution screening. Second, because many false positives can pass the primary screen, the burden of secondary deconvolution screens can be huge. PI-deconvolution overcomes both of these limitations. We have demonstrated that, as a result of the built-in repetitions, PI-deconvolution improves coverage and accuracy simultaneously. Deconvolution is inherent in the PI-deconvolution method and involves direct ‘profile reading’ without the necessity of secondary screens. Most hits are at least partially deconvoluted (92% of the hits are narrowed down to at most 4 baits); further deconvolution can be achieved by pair-wise confirmation. Unlike methods requiring secondary screens, PI-deconvolution is generally applicable to both two-hybrid array and proteome microarray platforms.

Protein networks are best modeled as scale-free networks²⁵ in which the majority of nodes have only a few neighbors and a small number of nodes ('hubs') have many neighbors. As expected, the coverage of interactions between proteins of high connectivity is lower than that between proteins of low connectivity (coverage for highly connected nodes can be improved when the bit number is decreased; see **Supplementary Note**). PI-deconvolution is especially useful for mapping interactions of low-degree nodes, which account for the majority of nodes in protein networks (**Supplementary Note**) and require the largest number of experiments using traditional single-bait methods.

If the average degree (of connectivity) increases proportionally with number of nodes, PI-deconvolution will be able to cover a human interactome network with the same efficiency as for yeast (using same pool size). But as average degree appears to be conserved among different organisms^{21,22,26}, even fewer experiments may in fact be needed, because a larger pool of baits can be accommodated on a larger proteome array. We suggest that PI-deconvolution-guided community efforts will greatly accelerate interactome mapping from yeast to human. Likewise, PI-deconvolution should be amenable to increasing the throughput of high-content and/or high-dimensional screening/mapping projects^{27,28}.

Besides increased screening efficiency, a primary benefit of PI-deconvolution is better data accuracy because of a high level of repetition. Analogously, communication systems are made accurate through extensive redundancy to reduce noise²⁹ (see also <http://www.lucent.com/minds/infotheory/>), and this principle has been used in designing DNA microarrays³⁰. The similarity between PI-deconvolution and binary communication systems should allow the translation of error-correction methods in information and computer science to large-scale biological network analyses. For example, an extra pair of pool screening (for example, with bait reshuffling) can be performed to provide additional redundancy for error detection, analogous to a parity check in communication systems.

Finally, many topological motifs have been described for complex networks (including electronic circuits, the world-wide web, cells, the brain, ecological systems and social networks)³¹. Not all motifs, however, occur in biological networks (the same is true for nonbiological networks). For example, the 4-node feedback loop (square lattice) design, which is commonly found in electronic circuits, is not a motif in protein interaction networks³¹. It would be interesting to study to what extent the PI-deconvolution strategy can be robustly applied to each type of network and each type of subgraph.

In summary, we present a pooling and deconvolution strategy (PI-deconvolution) that is generally applicable to maximize screening efficiency in a wide variety of situations. The essence of PI-deconvolution is imaginary tagging (binary coding), combinatorial pooling and built-in deconvolution and cross-validation. The key advantages of PI-deconvolution include better accuracy, coverage and efficiency. PI-deconvolution is very flexible because it can be easily scaled up or down by setting different n values. In addition to protein interaction networks, PI-deconvolution can be useful for other library-against-library scenarios, as long as most probes in the query library have only a few targets in the subject library. The versatility lies in imaginary tagging, which is universally applicable regardless of the nature of the query (molecules, cells, organisms and others).

METHODS

Proteome microarray, yeast two-hybrid array and drug resistance screening. We probed the yeast proteome microarrays¹² (Protoarray Yeast Proteome Microarray; Invitrogen) with purified V5 epitope-tagged 'bait' proteins according to manufacturer's instructions. For pooling experiments, the final concentration of each probe (bait protein) was 5 ng/μl. To detect the bound bait proteins, we probed the arrays with an Alexa Fluor 647-labeled anti-V5 and acquired the array image using an Axon GenePix 4000B scanner. Yeast two-hybrid array screens were performed as described previously¹⁴. For drug resistance screening, the 128 yeast deletion strains containing one rapamycin-resistant strain and one wortmannin-resistant strain were chosen based on previous genome-wide, single strain fitness data^{32,33}. Pools of strains were tested in clear 96-well plates for fitness response to rapamycin (100 nM) and wortmannin (1 μM), using DMSO (drug carrier) as control.

Analysis of proteome microarray data. The amount of yeast protein for each ORF on the array is variable and approximately equivalent solution protein concentrations are available for each spot. Because there is the potential for proteins present in higher quantity on the array to skew data analysis by having higher signals, we developed an approach that identifies hits from parallel protein microarray experiments that considers the amount of protein in each spot. Suppose d_c is the fluorescence intensity read from spot X_c with concentration c , a subset of intensity data $D = \{d_i, i\}$ represents all spots with concentrations between $90\% \times c$ and $110\% \times c$ will be compared to d_c . Let $d_0 = \sum d_i / n_c$ (n_c is the number of spots in D), P value of testing if intensity of spot X_c is higher than d can be calculated as $1 - \Phi((d_c - d_0) / S_D)$, in which S_D is the standard deviation of dataset D and $\Phi(x)$ is the cumulative probability function of standard normal distribution. We will usually consider spot X_c as a positive hit if its P value is < 0.01 . Spot data representing the results of probing with the 15 bait proteins in both single and pooling screens are available in **Supplementary Table 7** online.

Simulation on the DIP network. We assume that all the interactions in the DIP graph are true interactions. In a simulated pooling experiment, a set of preys of a given set of baits will be returned. Experimental false-positive rate (FP) and false-negative rate (FN) are considered in the virtual screening procedures. Let N_{neighbor} be the number of neighbors of all the baits in a pool according to DIP, N_{tp} be the number of detected true positives in the neighborhood and N_{fp} be the number of false positives in the virtual screening procedure. A simulated screen will return a set of nodes containing N_{tp} random nodes from true positives and N_{fp} nodes from non-neighboring nodes. Knowing that $FN = (N_{\text{neighbor}} - N_{\text{tp}}) / N_{\text{neighbor}}$ and $FP = N_{\text{fp}} / (N_{\text{fp}} + N_{\text{tp}})$, we can get $N_{\text{tp}} = N_{\text{neighbor}} \times (1 - FN)$ and $N_{\text{fp}} = FP / (1 - FP) \times N_{\text{tp}}$. Accuracy (fraction of true interactions among detected interactions) and coverage (fraction of true interactions detected) are two criteria that we used to measure the performance of our strategy. Four parameters are simulated: bit number (BitNum), occurrence cutoff (CutOcc, the number of times a signal is detected to be considered a true positive), experimental false-positive rate (FP, fraction of false positive signals in all positive signals) and false-negative rate (FN, fraction of undetected true positives).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank H. Herschman, C. Miller, E. O'Shea, F. Fox, C. Stanyon and members of the Huang laboratory for critical readings and suggestions on the manuscript, the anonymous reviewer for introducing to us the idea of communication systems, Y. Du for assistance with drug screening and K. Scanlan for faithfully supporting our work. This research was partially supported by a Singleton Developmental Grant (to J.H.), University of California Systemwide Biotechnology Research & Education Program, Graduate Research and Education in Adaptive bioTechnology (GREAT) Training Grant 2005-268 (F.J. and J.H.), and a grant from the US National Center for Research Resources of the National Institutes of Health, P41 RR11823 (S.F.).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the Nature Methods website for details).

Published online at <http://www.nature.com/naturemethods/>
Reprints and permissions information is available online at
<http://npg.nature.com/reprintsandpermissions/>

- Phizicky, E., Bastiaens, P.I., Zhu, H., Snyder, M. & Fields, S. Protein analysis on a proteomic scale. *Nature* **422**, 208–215 (2003).
- Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ito, T. *et al.* Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA* **97**, 1143–1147 (2000).
- Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
- Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
- Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
- Rual, J.F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
- Gavin, A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
- Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
- Michaud, G.A. *et al.* Analyzing antibody specificity with whole proteome microarrays. *Nat. Biotechnol.* **21**, 1509–1512 (2003).
- Zhong, J., Zhang, H., Stanyon, C.A., Tromp, G. & Finley, R.L., Jr. A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res.* **13**, 2691–2699 (2003).
- Hazbun, T.R. *et al.* Assigning function to yeast proteins by integration of technologies. *Mol. Cell* **12**, 1353–1365 (2003).
- Wu, X., Hart, H., Cheng, C., Roach, P.J. & Tatchell, K. Characterization of Gac1p, a regulatory subunit of protein phosphatase type I involved in glycogen accumulation in *Saccharomyces cerevisiae*. *Mol. Genet. Genomics* **265**, 622–635 (2001).
- Winzler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Heitman, J., Movva, N.R. & Hall, M.N. Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science* **253**, 905–909 (1991).
- Carpenter, C.L. & Cantley, L.C. Phosphoinositide kinases. *Biochemistry* **29**, 11147–11156 (1990).
- Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32** (Database issue), D449–D451 (2004).
- Grigoriiev, A. On the number of protein-protein interactions in the yeast proteome. *Nucleic Acids Res.* **31**, 4157–4161 (2003).
- Bork, P. *et al.* Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.* **14**, 292–299 (2004).
- Janda, K.D. Tagged versus untagged libraries: methods for the generation and screening of combinatorial chemical libraries. *Proc. Natl. Acad. Sci. USA* **91**, 10779–10785 (1994).
- Agyare, F.D. *et al.* Mapping expressed sequence tag sites on yeast artificial chromosome clones of *Arabidopsis thaliana* DNA. *Genome Res.* **7**, 1–9 (1997).
- Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & Barabasi, A.L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
- Massoud, T.F. & Gambhir, S.S. Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev.* **17**, 545–580 (2003).
- Gray, P.A. *et al.* Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255–2257 (2004).
- Barry, J.R., Lee, E.A. & Messerschmitt, D.G. *Digital communication* 3rd edn. (Kluwer Academic Publishers, Boston, 2004).
- Khan, A.H., Ossadtchi, A., Leahy, R.M. & Smith, D.J. Error-correcting microarray design. *Genomics* **81**, 157–165 (2003).
- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Zewail, A. *et al.* Novel functions of the phosphatidylinositol metabolic pathway discovered by a chemical genomics screen with wortmannin. *Proc. Natl. Acad. Sci. USA* **100**, 3345–3350 (2003).
- Xie, M.W. *et al.* Insights into TOR function and rapamycin response: Chemical genomic profiling by using a high-density cell array method. *Proc. Natl. Acad. Sci. USA* **102**, 7215–7220 (2005).