An experimentally derived confidence score for binary protein-protein interactions

Pascal Braun[1,2,8], Murat Tasan[3,8], Matija Dreze[1,2,4,8], Miriam Barrios-Rodiles[5], Irma Lemmens[6], Haiyuan Yu[1,2], Julie M Sahalie[1,2], Ryan R Murray[1,2], Luba Roncari[5], Anne-Sophie de Smet[6], Kavitha Venkatesan[1,2,7], Jean-François Rual[1,2,7], Jean Vandenhaute[4], Michael E Cusick[1,2], Tony Pawson[5], David E Hill[1,2], Jan Tavernier[6], Jeffrey L Wrana[5], Frederick P Roth[1,3] & Marc Vidal[1,2]

Information on protein-protein interactions is of central importance for many areas of biomedical research. At present no method exists to systematically and experimentally assess the quality of individual interactions reported in interaction mapping experiments. To provide a standardized confidence-scoring method that can be applied to tens of thousands of protein interactions, we have developed an interaction tool kit consisting of four complementary, high-throughput protein interaction assays. We benchmarked these assays against positive and random reference sets consisting of well documented pairs of interacting human proteins and randomly chosen protein pairs, respectively. A logistic regression model was trained using the data from these reference sets to combine the assay outputs and calculate the probability that any newly identified interaction pair is a true biophysical interaction once it has been tested in the tool kit. This general approach will allow a systematic and empirical assignment of confidence scores to all individual protein-protein interactions in interactome networks.

Physical protein-protein interactions are an elementary constituent of biological systems, and discovering interaction networks is a principal goal in systems biology. There are two complementary branches of protein interaction analysis. Analysis of protein complexes using affinity purification followed by mass spectrometry (AP/MS) identifies directly and indirectly associated proteins. High-throughput yeast two-hybrid (Y2H) analyses identifies direct, binary interactions; other methods for identifying direct interactions have recently yielded smaller data sets[1].

As binary interaction mapping has grown, data set quality has, rightly, been scrutinized. A first study that compared several interaction data sets to a gold standard of high-quality protein complexes suggested that high-throughput Y2H data have poor quality[2]. A more recent analysis showed that protein complexes are inappropriate for evaluating Y2H data and that high-throughput Y2H data are of high quality when compared against a gold standard of directly interacting proteins[3].

Given the importance of protein interactions and the demand for better and more comprehensive maps, standardized experimental methods for quality control are crucial. These are particularly important for determining all the direct, physical interactions between human proteins in the context of a human interactome project as they will enable the scientific community to evaluate assay implementations under a universally interpretable quality standard.

Methods for quality control can be categorized according to the evidence they analyze and according to whether they evaluate the quality of the data set as a whole or the quality of individual interactions (**Supplementary Fig. 1** online). Early quality assessments used the strength of a correlation with indirect secondary data, such as coexpression or functional annotation, mostly to measure the overall data set quality[2,4–6]. However, knowledge about biological roles of proteins is limited for most proteins. Furthermore, expression of interacting proteins need not be correlated over many conditions, and, conversely protein pairs with correlated expression patterns do not necessarily physically interact. Although a correlation with secondary data increases confidence in interaction data, no conclusion can be drawn from the absence of such correlations. This is particularly true for high-throughput data that do not have a sociological investigation bias and are more likely to contain unexpected connections. For rigorous quality assessment, it is important to use protein interaction evidence.

Several approaches assign confidence scores based on experimental data. The frequency with which interactions are found in

[1]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. [2]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, Massachusetts 02115, USA. [4]Facultés Universitaires Notre-Dame de la Paix, 61 Rue de Bruxelles, 5000 Namur, Belgium. [5]Centre for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto Ontario M5G 1X5. [6]Department of Medical Protein Research, Flanders Institute for Biotechnology, and Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium. [7]Present addresses: Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue Cambridge MA 02139 (K.V.) and Harvard Medical School, Department of Cell Biology, 240 Longwood Avenue, Boston, Massachusetts 02115, USA (J.-F.R.). [8]These authors contributed equally to this work. Correspondence should be addressed to M.V. (marc_vidal@dfci.harvard.edu), P.B. (pascal_braun@dfci.harvard.edu), J.L.W. (wrana@lunenfeld.ca), F.P.R. (fritz_roth@hms.harvard.edu) or J.T. (jan.tavernier@ugent.be).

RECEIVED 4 JUNE; ACCEPTED 17 NOVEMBER; PUBLISHED ONLINE 7 DECEMBER 2008; DOI:10.1038/NMETH.1281

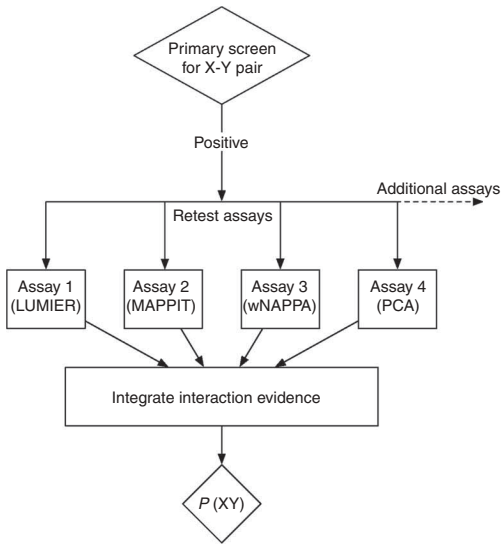

Figure 1 | Strategy for deriving a confidence score for individual protein-protein interactions after high-throughput screening using data from several complementary follow-up interaction assays. After initial screening using a high-throughput method—for example, Y2H—all positives are evaluated using secondary tool kit assays that have been benchmarked by the positive reference set (PRS) and random reference set (RRS). The resulting raw data are integrated using a model trained on the PRS and RRS calibration data to derive a confidence based on the experimental support for each interaction (see text for details).

high-throughput data sets has been used to inform error models for the combined data[7]. This approach does not provide a quality assessment for individual interactions, however. To prioritize individual interactions several groups have calculated a confidence score on the basis of their experimental data—for example, how often a given interaction was recovered in Y2H screens[8,9], how reliably a protein was identified in an AP/MS experiment[10,11] or combinations of such primary experimental data with secondary data[12]. These scores are strongly influenced by biases of the particular experimental setup and are therefore interpretable only within this context. A more universal confidence scoring approach is therefore desirable.

Previously, we and others experimentally assessed data set quality by testing a subset of newly identified interactions in an orthogonal interaction assay[3,13,14]. This experimental approach confirmed only a fraction of all interactions. Because every method has inherent limitations and because performance depends on stringency of the implementation, unconfirmed interactions may be merely false negatives of the second assay.

To individually confirm most 'true' biophysical interactions found in a high-throughput screen, we pursued a strategy of retesting every candidate interaction in a panel of several interaction assays. If the assay performances are benchmarked against a common reference, the data from these retesting experiments can be quantitatively integrated into a confidence score (**Fig. 1**). In future interaction mapping experiments, it will then be possible to report a probability for every protein interaction, along with the experimental evidence and the underlying benchmarking data for every assay.

RESULTS

Compiling an interaction assay tool kit

To establish a confidence-scoring methodology we first compiled and characterized a panel of protein interaction assays implemented in high-throughput format. The assays are based on complementary principles and expression systems (**Fig. 2**): Y2H[15,16], mammalian protein-protein interaction trap (MAPPIT)[17], luminescence-based mammalian interactome (LUMIER)[1], yellow fluorescent protein (YFP) protein complementation assay (PCA)[18] and a modified version of the nucleic acid programmable protein array (wNAPPA)[19]. Because the assays are compatible with recombinational cloning and performed in 96-well format, hundreds of thousands of interactions identified in high-throughput screens can be tested at a reasonable cost (**Supplementary Data** online).

Benchmarking tool kit assays using reference sets

To develop a confidence score, we characterized assay performance using a positive reference set (PRS) and a random reference set (RRS) for protein interactions[20]. Our first version of a human PRS (hsPRS-v1) contained 92 interacting human protein pairs for which we found more than one peer-reviewed publication in manually curated databases[21–25] (details in Methods and in **Supplementary Tables 1** and **2** online). Apart from verifying the curation reports[26] and ensuring open reading frame (ORF) availability in the human ORFeome version 1.1 (ref. 27), we applied no filters, so interactions between membrane proteins, interactions between ligand-receptor pairs, and interactions dependent on post-translational modifications (PTMs) were all included. HsPRS-v1 thus constituted a reasonable representation of well established human binary interactions. For our first version of a human RRS (hsRRS-v1), 92 protein pairs were chosen randomly from human ORFeome 1.1 (out of 10^8 possible pairwise combinations within ORFeome1.1) after removing all previously described interacting pairs[21–25]. Because there is no available gold standard for non-interacting proteins and because randomly chosen protein pairs are unlikely a priori to interact, our RRS serves as a negative control set. Alternative approaches for choosing negative training examples are possible, but they introduce unacceptable biases[28,29].

We tested all pairs in the reference sets by tool kit assays evaluating the effect of assay stringency on the detection of PRS and RRS pairs (**Supplementary Fig. 2** online). The use of 184 controls, as opposed to the small numbers usually used to characterize interaction assays, increases robustness. The receiver-operating characteristics (ROC) curves of the four tool kit assays, illustrating the tradeoff between true and false positive rates as a function of stringency, are shown in **Figure 3a**. For the analysis of assay performance, we used a threshold that maximized detection of interactions in PRS pairs while maintaining a low number of positive-scoring RRS pairs.

For Y2H, we measured activation of one or more reporter genes using both high- and low-copy plasmids to alter the expression levels of DNA-binding-domain and activation-domain fusion proteins (DB-X and AD-Y, respectively)[15,30]. Most implementations involved the Y8800/8930 yeast strains (Y strain) now used for screening at CCSB[3], and we included one implementation with the MaV103/203 (ref. 30) strains (MaV strain) for comparison[20]. The standard implementation for high-throughput Y2H screening at CCSB (low copy plasmids, both reporters activated[3]) detected 16% of hsPRS-v1 pairs and no hsRRS-v1 pairs, whereas relaxation of the

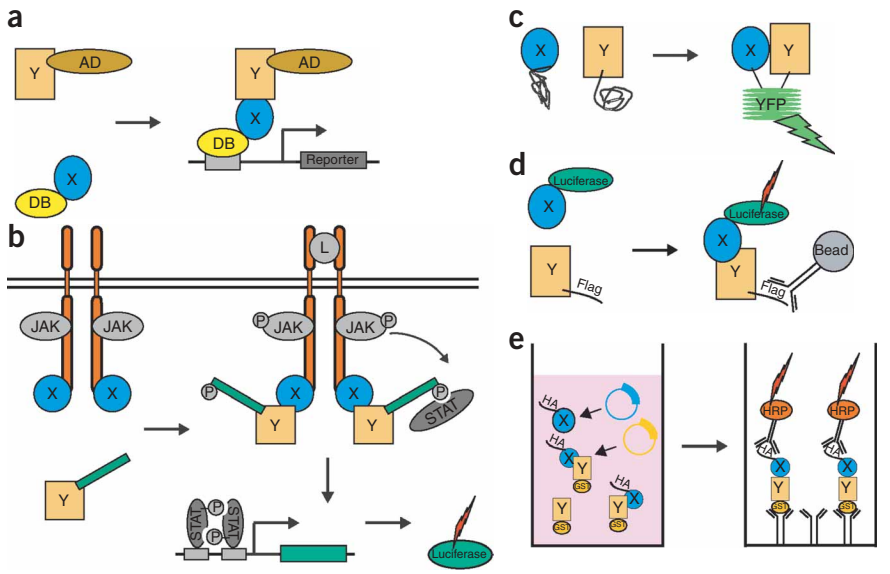

Figure 2 | Schematic descriptions of complementary tool kit assays for binary protein interaction. (**a**) Y2H. The Gal4 transcription factor is reconstituted by interaction between bait (X) fused to the Gal4 DNA-binding domain (DB) and prey (Y) fused to the Gal4 activation domain (AD), to activate one or more independent reporter genes in *S. cerevisiae*. (**b**) MAPPIT. A bait protein is fused to a hybrid erythropoietin–leptin receptor and the prey is fused to gp130. Upon stimulation with erythropoietin, JAK2 molecules transphosphorylate each other, and if bait and prey interact, the activated JAKs will phosphorylate gp130, which in turn recruits and subsequently activates STAT3, which then activates transcription of a reporter. (**c**) PCA. Two YFP fragments fused to bait and prey proteins reconstitute a fluorescent protein if brought in close proximity by two proteins that physically interact. (**d**) LUMIER. A luciferase-tagged bait (X) is expressed together with a Flag-tagged prey (Y) in HEK-293T cells. The association between these proteins is determined by coimmunoprecipitation using an antibody to Flag and the presence of the bait is detected through its luciferase activity. (**e**) wNAPPA. Plasmids encoding GST–bait and HA–prey are mixed in a coupled transcription-translation reticulocyte lysate (purple shading) to express protein. Subsequently, the bait–GST is captured on the bottom of 96-well plate coated with antibodies to GST. If the proteins are interacting, the HA–prey fusion protein can be immunologically detected.

scoring criteria (either reporter activated) led to an increase in assay sensitivity to 25% while not affecting RRS detection (**Fig. 3b**). Raising protein levels increased Y2H assay sensitivity to >40% at the expense of detecting four hsRRS-v1 pairs. The profile of individual interactions detected differed between Y2H implementations (**Supplementary Fig. 3** online). When the same search space of the human interactome was interrogated using two different yeast strains, many interactions were detected by only one of the two implementations. Hence, different implementations of the same technology can lead to different results, which may partially explain why data sets acquired by different groups show limited overlap.

LUMIER

The LUMIER pull-down assay (**Fig. 2d**) had the highest assay sensitivity, detecting 36% of hsPRS-v1 pairs and four hsRRS-v1 pairs at the chosen scoring threshold (**Fig. 4a**). LUMIER recovered one phosphorylation-dependent interaction (**Fig. 4b**). Post-translational modification–dependent interactions, such as that between SMAD1 and SMAD4, are likely to be detected only after proper activation of specific signaling cascades[1].

MAPPIT

MAPPIT generates a ligand-dependent semiquantitative luciferase readout in mammalian cells without enrichment steps (**Fig. 2b**). The scoring conditions detected 33% of hsPRS-v1 pairs and two positive hsRRS-v1 pairs (**Fig. 4a**). Like Y2H and PCA, MAPPIT detected two phosphorylation-dependent interactions (**Fig. 4b**). The detection of other PTM-dependent interactions may require additional stimulation or use of the heteromeric MAPPIT variant[31].

Y2H

The Y2H system (**Fig. 2a**) had an assay sensitivity of 25%, and no hsRRS-v1 pairs were detected (**Fig. 4a**). Contrary to common perception, the detection of interactions involving nuclear proteins was not greater in Y2H than in other assays (**Supplementary Fig. 4a** online). We also detected phosphorylation-dependent interactions (PTK2-SRC, SMAD1-SMAD4) (**Fig. 4b**), so Y2H, like the other assays, can detect a subset of PTM-dependent interactions.

PCA

The YFP reconstitution of the PCA assay (**Fig. 2c**) gave such strong signals that stringent criteria had to be applied to decrease the number of hsRRS-v1 interactions that scored as positive. High signal strength may be due to the irreversible nature of YFP refolding. After optimization, PCA had an assay sensitivity of 23% with two positive-scoring RRS pairs (**Fig. 4a**).

wNAPPA

The completely *in vitro* wNAPPA (**Fig. 2e**) had an assay sensitivity of 21% under conditions in which three hsRRS-v1 pairs were detected (**Fig. 4a**). Notably, 7 out of 16 interactions between two membrane proteins scored positive (43%) (**Supplementary Fig. 4a**), in contrast to none of the seven hsRRS-v1 membrane-protein pairs.

Detection of hsPRS-v1 interaction pairs

The performance of each assay on PRS and RRS reveals the ability to report true protein interactions under the respective assay conditions (**Fig. 4a**). This approach provides a method for benchmarking different assays on a standardized set of controls. Unexpectedly, assay sensitivity of all methods fell in a similar range, ~20–35%.

A total of 55 out of the 92 (59%) hsPRS-v1 interactions were detected by at least one of the tool kit assays. Notably, the five complementary assays combined failed to detect ~40% of hsPRS-v1 (**Fig. 4b**). We investigated possible systematic reasons for this finding (**Supplementary Data** online). A combination of different

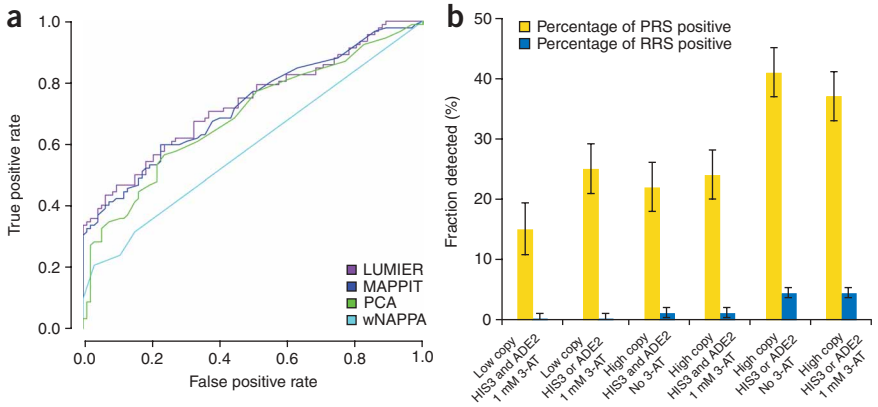

Figure 3 | Evaluation of assay performance at different stringencies using hsPRS-v1 and hsRRS-v1. (**a**) The receiver operating characteristics (ROC) curve shows the tradeoff between true and false positive rate at different stringencies of the tool kit assays. For different applications, different thresholds may be used. (**b**) Y2H assay performance in different scoring protocols (one versus two reporters) and different bait and prey expression levels (high-copy (2μ) versus low-copy (*CEN*) plasmid), studied in Y strain. 3-Amino-1,2,4-triazole (3-AT) is a competitive inhibitor of HIS3 and was included to reduce background in one set of experiments.

reasons explained detection failure, including PTM dependence, steric geometry of fusion proteins (**Supplementary Fig. 5b** online) and refractoriness to detection on the part of certain protein families, such as extracellular proteins (**Supplementary Fig. 4a**). The use of different sequence isoforms probably also has an effect. The method originally used to describe each hsPRS-v1 interaction had no detectable influence on the insensitivity (**Supplementary Table 1** and data not shown).

Detection of RRS protein pairs

In all assays except Y2H, the threshold values chosen led to several hsRRS-v1 pairs scoring positive. Such small numbers of detected RRS pairs in these experiments cannot be extrapolated directly to false-positive rates expected in large-scale screens[1]. Large-scale screens would need to be carried out at greater stringency than that used in these confirmatory assays, but this may come at the expense of detecting fewer real interactions.

Several hsRRS-v1 pairs were positive in one (seven pairs) or in two (ITPA/WDR62 and MCCC1/GALK1) different assays (**Fig. 4b**). The protein pairs were picked randomly, and because current knowledge about interaction networks is incomplete, these pairs may be real yet previously unknown interactors. However, it is unlikely that all nine hsRRS-v1 pairs detected (10%) fall into this category. Some interactions may be 'pseudointeractions'—that is, biophysically valid interactions that never occur *in vivo* because the involved proteins are separated spatially or temporally. Artifactual interactions may occur at a given frequency as a consequence of the particular conditions of the respective assay (expression levels plus fusion tags plus mechanics of the assay). Although interactions detected by only a single assay should be interpreted with caution, one-quarter (22) of the well documented hsPRS-v1 were detected by just a single assay. As such, protein pairs detected by a single method cannot be dismissed outright. Such interactions emphasize the need to integrate the outcome of several assays into a quantitative confidence score.

Integrative analysis of assay outcome

The PRS/RRS characterization quantitatively benchmarks protein interaction assays against a standardized reference and thus measures the reliability of a positive result in each assay. The tool kit assays in conjunction with the PRS/RRS benchmarking data can therefore be used for confidence scoring, with all interactions identified in large-scale primary screens retested by the tool kit. The confidence assigned to each interaction may then be adjusted according to the outcome of these retests (**Fig. 1**).

We modeled the probability that a given protein pair is truly interacting using a bayesian framework. Traditionally, the continuously valued raw data of each assay are converted into a binary

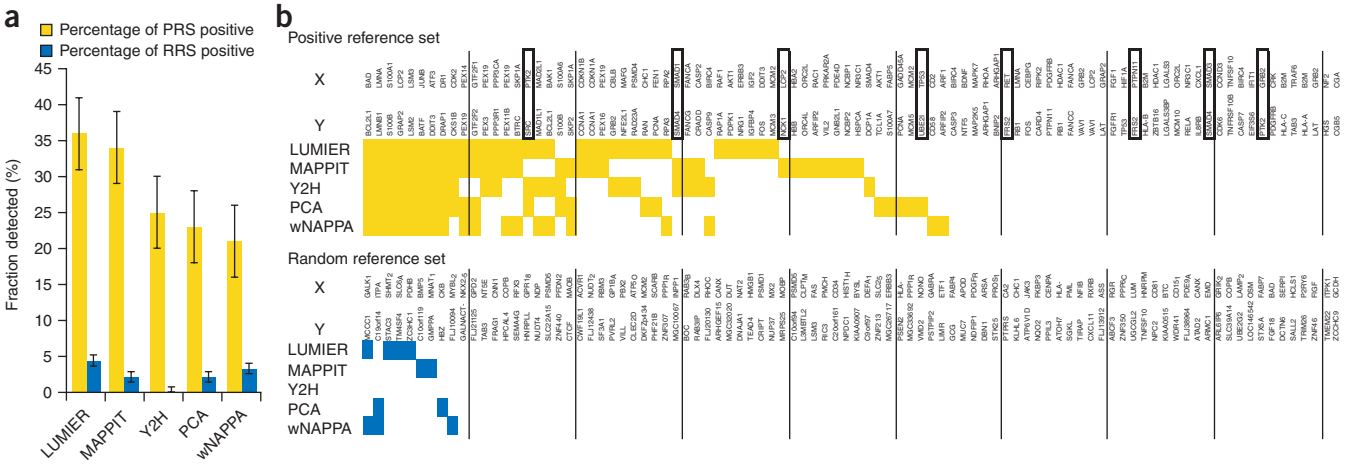

Figure 4 | Performance of assays against positive and random reference sets PRS and RRS. (**a**) Quantification of assay sensitivity and specificity, with s.e.m., using hsPRS-v1 and hsRRS-v1. (**b**) Detection of individual hsPRS-v1 and hsRRS-v1 pairs by the tool kit assays. Top panel: detected hsPRS-v1 pairs are indicated by yellow squares. Bottom: detected hsRRS-v1 pairs are indicated by blue squares. Phosphorylation-dependent interactions are boxed. Thresholds used for the assays can be found in Methods.

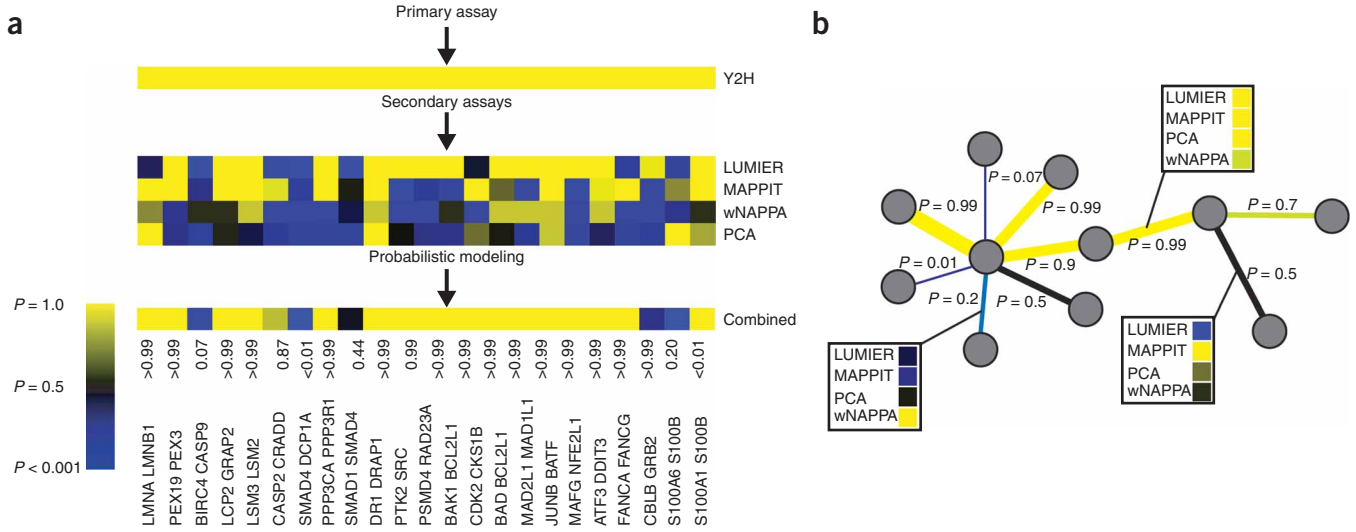

Figure 5 | Application of the integrated confidence score. (**a**) Application of the confidence scoring scheme to the Y2H-positive pairs from hsPRS-v1. Probabilities for pairs within each assay (middle panel) are computed using leave-one-out cross-validation (LOO-CV) and a single-assay logistic regression model, trained identically to the combined-assay regression model. (**b**) Every interaction reported in future protein-protein interaction mapping experiments can be assigned a confidence score based on the tool kit assay data for each individual protein pair. In the hypothetical network shown, confidence scores for each interaction are based on experimental data (boxes) and represented by the width and color of the edge.

('yes' or 'no') call for each protein pair, but much information is lost in this way and there is the potential to 'over-fit' the thresholds used to make a binary call. We instead used normalized raw data (for LUMIER, MAPPIT, wNAPPA and PCA) across hsPRS-v1 and hsRRS-v1 to train a logistic regression model that computes a probabilistic confidence score for each potential interaction. This confidence scoring method will not be applied to randomly chosen protein pairs but to candidate interactions identified in, for example, high-throughput Y2H screens. We used the 23 Y2H-positive hsPRS-v1 pairs as positive examples to train the regression model. Our approach is not limited to evaluating Y2H interactions, however, and can easily be adapted to other screening methods. To account for training example bias (the ratio of positive to negative samples), we used a Bayesian correction incorporating the prior probability that the interactions are 'true' before considering additional evidence from the tool kit assays. We used the 78% precision previously estimated for high-throughput Y2H interactions in the CCSB-HI1 human interactome data set (details in **Supplementary Methods** online)[14].

We evaluated predictive performance of the integrated model, i.e., sensitivity vs. specificity analysis, based on relative ranks using leave-one-out cross-validation (LOO-CV) (**Supplementary Fig. 6** online). The beta parameter for each assay indicates the contribution of this assay to the final score. The large values for LUMIER and MAPPIT indicate that these two assays provide most information to the final confidence score (Table A in **Supplementary Methods**). Thus, the beta parameters can be used to optimize the tool kit for greatest value—that is, most complementary assay implementations—which would lead to similar beta-parameters for the different assays. Features that could be optimized include eliminating less informative assays, replacing one assay with a more informative assay, or changing details of an existing assay implementation, i.e., use of different constructs.

The most important feature of our integrative analysis is that it calculates a confidence score for each candidate interaction, based on the results of a consistently benchmarked panel of tool kit assays. We demonstrated the concept by calculating the LOO-CV confidence score for the Y2H positive hsPRS-v1 pairs (**Fig. 5a**). In future interactome maps, such confidence scores can be provided along with the underlying experimental evidence for every reported interaction (**Fig. 5b**).

DISCUSSION

Our confidence scoring method crucially uses a PRS and RRS to benchmark interaction assays. The reference sets enabled standardized calibration of different interaction assays and implementations. This calibration can be expanded to other binary interaction assays and facilitates optimization and interpretation of protein interaction experiments. Establishment and widespread use of more reference sets would enhance the ability to compare and interpret data acquired in different studies.

These first version reference sets suffer from small size, but growing knowledge about protein interactions and increased ORF availability will enable construction of larger and more representative PRS and RRS versions. One potential bias in PRSs could be toward 'better-behaved' interactions that are more easily detected by different assays. Such a bias could lead to artificially low confidence scores. We did not find that interactions supported by multiple methods in the curation were more easily detected by the tool kit. Even after retesting in several assays, three Y2H-positive hsPRS-v1 interactions received a low confidence score (**Fig. 5a**). Even among low-scoring interactions valid pairs can be found, so the experimentally determined scoring parameters may be conservative.

In contrast to conventional assumptions, under our assay conditions Y2H (i) was not more prone to yield false positives than other binary assays, (ii) did not favor detection of interactions among nuclear proteins and (iii) was not blind to PTM-dependent interactions. Mammalian cell–based assays will nonetheless be better suited for focused studies of context-dependent interactions

and may provide time-resolved information[1,18]. Increasing protein expression in Y2H considerably increased the number of detected hsPRS-v1 pairs, albeit at the expense of a greater number of positive-scoring hsRRS-v1 pairs. Once a confidence scoring scheme becomes established, primary screening stringency can be relaxed, boosting the number of interactions detected while reporting a *P*-value for all published interactions and thus ensuring their high quality.

Our linear regression model calculates, on the basis of the tool kit data, a quantitative confidence score for any interaction pair. This model can be adapted to other assays and implementations and to different primary screening methods, so long as assays are benchmarked using PRS and RRS and the false discovery rate for screening methods is estimated. Thus, this method for assessing confidence in protein-protein interactions is universal.

A standardized and transferable approach for calculating confidence scores will affect interaction mapping in several ways. The ability to discriminate between 'exceedingly likely' and 'possible' interactions will be valuable for triaging follow-up studies. Quantitative scores assessing confidence in protein interaction may be readily integrated with other types of genomic data to predict functional relationships. Using these scores, the scientific community can set quality standards and compare data sets generated by different protein interaction mapping methods.

METHODS

Construction of the first version of a human positive reference set (hsPRSv1). To measure the performance of the interaction tool kit assays, we sought to compile a diverse panel of very well documented interactions from the literature as a positive reference set. Within the space of approximately 7,000 × 7,000 ORFs defined by the human ORFeome 1.1 collection available at that time, five manually curated databases (MINT[21], HPRD[22], BIND[23], MIPS[24] and DIP[25]) reported 4,067 direct binary interactions. To ensure high-quality interaction data, we required two publications describing each interaction, obtaining 275 interaction pairs. After elimination of homodimers, we were left with 188 interactions, which we manually recurated by reading all publications reported by the databases as supporting these interactions. One unexpected result of this recuration was that 38% of 'curation units' (one experiment supporting one interaction reported by any one database) were wrong, thus validating the insistence on multiple papers and multiple databases supporting each interaction. Of the 111 interactions supported by >1 verified publication, 92 were among full-length proteins, thus constituting hsPRS v1. All interaction pairs and a standardized description of the methods by which the interactions have been described can be found in **Supplementary Table 1**.

Y2H. The Y2H assay was done using a modified retesting protocol[3]. Haploid yeast expressing individual AD and DB constructs were mated, diploids were selected, and HIS^+ and ADE^+ phenotypes were scored after incubation for 4 d at 30 °C. The same protocol was used for the high-copy 2-micron (2μ) Y2H vectors (pVV212 (DB) and pVV213 (AD)[30]). MaV-strain experiments were performed as described[20]. The genotype of the Y strains is *GAL2::ADE2 LYS2::GAL1-HIS3 met2::GAL7-lacZ cyhR ura3-52 leu2-3 trp1-901 his3D200 gal4D gal80D*. The genotype of the MaV strains is *SPAL10::Ura3 GAL1::LacZ GAL1::HIS3@LYS2 can1R cyhR ura3-52 leu2-3 trp1-901 his3D200 ade2-101 gal4D gal80D*. For the experiments in **Figure 4**, CYH-sensitive activation of either HIS3 or ADE2 in either tested configuration (**Supplementary Fig. 5**) was used as a criterion for positive interactions using *CEN* plasmids in the Y strain. A detailed step-by-step protocol can be found in **Supplementary Protocol 1** online.

MAPPIT. MAPPIT experiments were performed as described[32] with minor changes. Briefly, HEK-293T cells were transfected in 96-well plates. Twenty-four hours after transfection, cells were stimulated with ligand (erythropoietin (R&D Systems)) or left untreated for an additional 24 h, followed by measurement of luciferase activity in triplicate. All interactions were tested in two configurations (**Supplementary Fig. 5**) and in two independent trials. For the data in **Figure 4**, pairs were scored positive if the luciferase activity was ≥12 times that obtained with both the irrelevant bait and the irrelevant prey in either of the two tested configurations. A detailed step-by-step protocol can be found in **Supplementary Protocol 2** online.

LUMIER. LUMIER assays were performed as described[1]. All configurations were tested in duplicate. HEK-293T cells were transfected in 96-well plates. Forty-eight hours after transfection, cells were lysed and 70% of lysate was processed for immunoprecipitation. Expression of *Renilla* luciferase (RLUC)-tagged baits was measured in 10% of the cell lysates. LUMIER intensity ratio (LIR) values were obtained as described[1] for the immunoprecipitates (LIR-IP) and calculated similarly for the total lysates (LIR-TOT). Normalized LIR (NLIR) was calculated as the ratio LIR-IP/LIR-TOT. For each tested pair, the corresponding control NLIR value, obtained with *Renilla* luciferase without fusion partner, was subtracted from the experimental NLIR value obtained from testing the bait and prey fusion proteins. These final NLIR values were used to generate the receiver-operating characteristics curve (**Fig. 3a**). The data in **Figure 4** were calculated by requiring an average NLIR score of ≥33.2 in either of the tested configurations (**Supplementary Fig. 5**). A detailed step-by-step protocol can be found in **Supplementary Protocol 3** online.

PCA. Baits were fused to the F1 fragment and preys to F2 fragments of YFP using Gateway LR reactions. Both plasmids were transfected in 96-well format into CHO-K1 cells. CFP-encoding plasmid was included to identify transfected cells. Eighteen hours after transfection, cells were washed, trypsinized at ~20–25 °C and analyzed using fluorescence-activated cell sorting. For **Figure 4**, protein pairs were scored as positive when, in any tested configuration (**Supplementary Fig. 5**), at least 30% of transfected cells in a well showed YFP signal above background and the YFP/CFP ratio was at least twice as high as the ratio of the average YFP signal to the average CFP ratio on that plate. A detailed step-by-step protocol can be found in **Supplementary Protocol 4** online.

wNAPPA. The protocol used was modified from ref. 19. Bait and prey fusion proteins were expressed in coupled transcription-translation mix. After protein expression, glutathione S-transferase (GST)-tagged bait proteins were captured at the bottom of a 96-well plate coated with antibody to GST. Interactions were

detected using antibody to hemagglutinin (anti-HA ascites) using standard immunochemical protocols. Signal was visualized using chemiluminescence. Signal was manually assigned a score between 0 and 5, and interactions that scored ≥ 2 in either configuration were scored positive (**Supplementary Fig. 5**). A detailed step-by-step protocol can be found in **Supplementary Protocol 5** online.

Calculation of confidence score. A complete and detailed description of the statistical methods can be found in **Supplementary Methods**.

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

ACKNOWLEDGMENTS

We thank S. Michnick and N. Ramachandran for reagents and technical help for the PCA and wNAPPA assays, respectively. We thank A. Datti, T. Sun and F. Vizeacoumar from the SMART Robotics Facility at the Samuel Lunenfeld Research Institute for help with the automated version of LUMIER assay. We thank all members of the Vidal, Tavernier, Roth, and Wrana laboratories for helpful discussions, Agencourt Biosciences for sequencing assistance, and A. Bird and D. Maher for administrative assistance. This work was supported by contributions from the W.M. Keck Foundation awarded to M.V., F.P.R. and D.E.H.; by the Ellison Foundation awarded to M.V.; by Institute Sponsored Research funds from the Dana-Farber Cancer Institute Strategic Initiative awarded to M.V. and CCSB; by US National Institutes of Health grants 5P50HG004233 and 2R01HG001715 awarded to M.V., F.P.R. and D.E.H., R01 ES015728 awarded to M.V., 5U54CA112952 awarded to J. Nevins (M.V. subcontract), 5U01CA105423 awarded to S.H. Orkin (M.V. project), R01 HG003224 awarded to F.P.R. and F32 HG004098 awarded to M.T.; by a University of Ghent grant GOA12051401 and the Fonds Wetenschappelijk Onderzoek–Vlanderen (FWO-V) G.0031.06 awarded to J.T., by a postdoctoral fellowship from the FWO-V awarded to I.L.; and by a grant from Genome Canada and funds from the Ontario Genomics Institute awarded to J.L.W. M.V. is a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS, French Community of Belgium).

AUTHOR CONTRIBUTIONS

P.B., M.T. and M.D. coordinated experiments and data analysis. P.B., M.D., J.M.S., J.-F.R., R.R.M. and H.Y. performed high-throughput Gateway cloning. P.B., H.Y. and J.M.S. implemented, developed and analyzed wNAPPA and PCA experiments. J.-F.R., K.V. and M.E.C. established PRSv1.0 and RRS reference sets. I.L., A.-S. de S., J.T. and K.V. coordinated, performed and analyzed MAPPIT experiments. M.B.-R., L.R. and J.L.W. coordinated, performed and analyzed LUMIER experiments. M.T. and F.P.R. developed the regression model. M.V. conceived the project. M.V., T.P., J.L.W. and D.E.H. developed the concepts underlying the overall strategy. D.E.H., F.P.R. and M.V. co-directed the project.

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

1. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307**, 1621–1625 (2005).
2. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002).
3. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
4. Ge, H., Liu, Z., Church, G.M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* **29**, 482–486 (2001).
5. Ramani, A.K., Bunescu, R.C., Mooney, R.J. & Marcotte, E.M. Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol.* **6**, R40 (2005).
6. Ramani, A.K. *et al.* A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol. Syst. Biol.* **4**, 180 (2008).
7. Chiang, T. *et al.* Coverage and error models of protein-protein interaction data by directed graph analysis. *Genome Biol.* **8**, R186 (2007).
8. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569–4574 (2001).
9. Sato, S. *et al.* A large-scale protein protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* **14**, 207–216 (2007).
10. Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
11. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
12. Han, J.D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
13. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
14. Rual, J.F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
15. Vidalain, P.O. *et al.* Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* **32**, 363–370 (2004).
16. Walhout, A.J. & Vidal, M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**, 297–306 (2001).
17. Eyckerman, S. *et al.* Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.* **3**, 1114–1119 (2001).
18. Nyfeler, B., Michnick, S.W. & Hauri, H.P. Capturing protein interactions in the secretory pathway of living cells. *Proc. Natl. Acad. Sci. USA* **102**, 6350–6355 (2005).
19. Ramachandran, N. *et al.* Next-generation high-density self-assembling functional protein arrays. *Nat. Methods* **5**, 535–538 (2008).
20. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* advance online publication, doi:10.1038/nmeth.1280 (7 December 2008).
21. Bader, G.D., Betel, D. & Hogue, C.W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**, 248–250 (2003).
22. Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Res.* **35**, D572–D574 (2007).
23. Mishra, G.R. *et al.* Human protein reference database–2006 update. *Nucleic Acids Res.* **34**, D411–D414 (2006).
24. Pagel, P. *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832–834 (2005).
25. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32**, D449–D451 (2004).
26. Cusick, M.E. *et al.* Literature-curated protein interaction datasets. *Nat. Methods* (in the press).
27. Rual, J.F. *et al.* Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14**, 2128–2135 (2004).
28. Ben-Hur, A. & Noble, W.S. Choosing negative examples for the prediction of protein-protein interactions. *BMC Bioinformatics* **7** (suppl. 1), S2 (2006).
29. Qi, Y., Bar-Joseph, Z. & Klein-Seetharaman, J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins* **63**, 490–500 (2006).
30. Vidal, M. *et al.* Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 10315–10320 (1996).
31. Lemmens, I. *et al.* Heteromeric MAPPIT: a novel strategy to study modification-dependent protein-protein interactions in mammalian cells. *Nucleic Acids Res.* **31**, e75 (2003).
32. Lemmens, I., Lievens, S., Eyckerman, S. & Tavernier, J. Reverse MAPPIT detects disruptors of protein-protein interactions in human cells. *Nat. Protoc.* **1**, 92–97 (2006).