SAINT: probabilistic scoring of affinity purification–mass spectrometry data

Hyungwon Choi[1], Brett Larsen[2], Zhen-Yuan Lin[2], Ashton Breitkreutz[2], Dattatreya Mellacheruvu[1], Damian Fermin[1], Zhaohui S Qin[3,8], Mike Tyers[2,4–6], Anne-Claude Gingras[2,4] & Alexey I Nesvizhskii[1,7]

We present 'significance analysis of interactome' (SAINT), a computational tool that assigns confidence scores to protein-protein interaction data generated using affinity purification–mass spectrometry (AP-MS). The method uses label-free quantitative data and constructs separate distributions for true and false interactions to derive the probability of a bona fide protein-protein interaction. We show that SAINT is applicable to data of different scales and protein connectivity and allows transparent analysis of AP-MS data.

The analysis of protein complexes and protein interaction networks is very important for biological research. A combination of affinity purification and mass spectrometry (AP-MS) has been increasingly used for both small-scale and large-scale analysis of protein complexes and interaction networks[1–4]. However, the development of computational tools for the processing of AP-MS data has not kept pace with improvements in experimental approaches. In addition to the general challenge of false positive protein identifications in mass spectrometry–based proteomic data[5], unfiltered AP-MS datasets contain many nonspecifically binding proteins; filtering these contaminants is the foremost computational challenge.

Whereas early methods filtered the noise using binary data (presence or absence of a protein), newer methods take into account quantitative information embedded in the mass spectrometric data (for example, label-free quantification, such as spectral counts). One such method converts the normalized spectral abundance factor (NSAF) into the posterior probability of a true interaction between a bait-prey pair using simple heuristics, which we term PP-NSAF hereafter[6]. Another method, named CompPASS, computes scores that adjust observed spectral counts relative to the reproducibility of detection across biological replicates and to the frequency of observing prey proteins in purifications of different baits[7]. Although both approaches can effectively analyze the datasets for which they had been developed, these scores are an empirical transformation of spectral counts without a probability model that can be used to estimate the measurement errors in the data in a transparent manner.

We have recently introduced an advanced approach for statistical analysis of interaction data from AP-MS experiments using label-free quantification, which we termed significance analysis of interactome (SAINT)[8]. As PP-NSAF and CompPASS, we had designed our original SAINT approach to analyze a specific dataset, the yeast kinase and phosphatase interactome. Expanding on this method, here we present a generalized SAINT framework that can be used to compute interaction probabilities in a variety of datasets. The method incorporates negative controls that are commonly generated as a part of the experimental study but can also be applied to large datasets in the absence of such data. We illustrate the methodology and its advantages through the analysis of datasets of different sizes and network density: from a large, sparsely connected network involving human deubiquitinating enzymes to a smaller, highly interconnected network for chromatin remodeling proteins and even to the analysis of a single bait, the protein CDC23.

The aim of SAINT is to convert the label-free quantification (spectral count X_{ij}) for a prey protein i identified in a purification of bait j into the probability of a true interaction between the two proteins, $P(\text{true} \mid X_{ij})$. The spectral counts for each prey-bait pair are modeled with a mixture distribution of two components representing true and false interactions. Note that these distributions are specific to each bait-prey pair. The parameters for true and false distributions, $P(X_{ij} \mid \text{true})$ and $P(X_{ij} \mid \text{false})$, and the prior probability π_T of true interactions in the dataset, are inferred from the spectral counts for all interactions that involve prey i and bait j. SAINT normalizes spectral counts to the length of the proteins and to the total number of spectra in the purification.

In addition to the experimental data for bait proteins, AP-MS data often contain negative controls (**Fig. 1a**). When these are available, SAINT estimates the spectral count distribution for false interactions directly from the negative controls, which makes the modeling approach semisupervised (Online Methods). SAINT modeling can also be performed without negative control data, so long as a sufficient number of independent baits are profiled and provided that these baits are not densely interconnected. In this case (**Fig. 1b**), a prey detected in the purification of a bait is

[1]Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA. [2]Centre for Systems Biology, Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada. [3]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. [4]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [5]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK. [6]Centre for Systems Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. [7]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [8]Present address: Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA. Correspondence should be addressed to A.C.G. (gingras@lunenfeld.ca) or A.I.N. (nesvi@med.umich.edu).

RECEIVED 28 JUNE; ACCEPTED 9 NOVEMBER; PUBLISHED ONLINE 5 DECEMBER 2010; DOI:10.1038/NMETH.1541

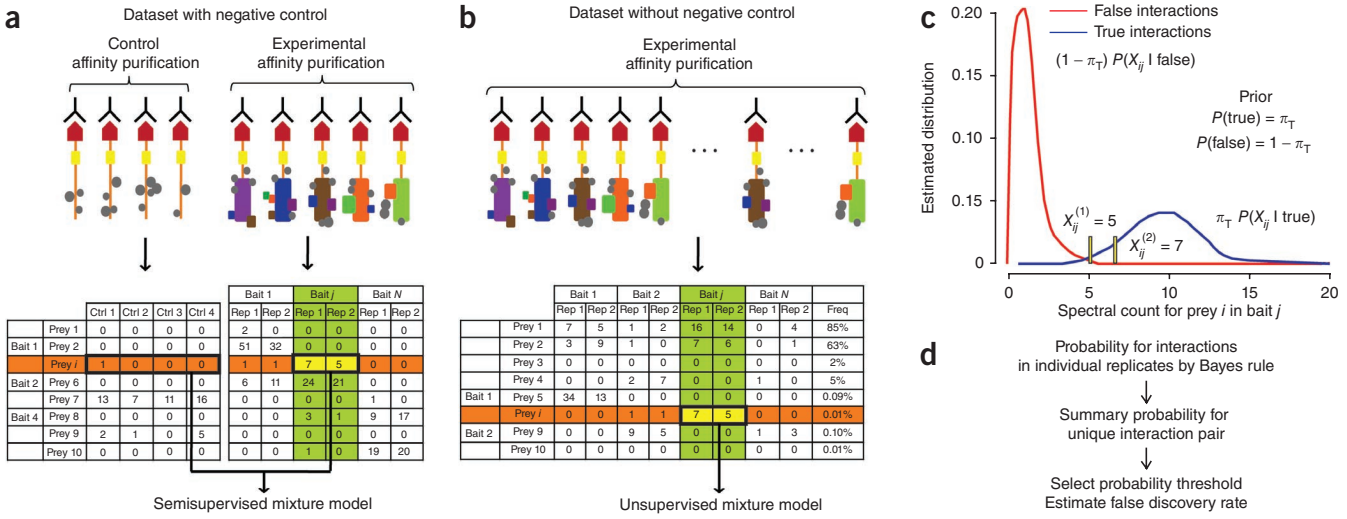

Figure 1 | Probability model in SAINT. (**a**,**b**) Interaction data in the presence (**a**) and absence (**b**) of control purifications. Schematic of the experimental AP-MS procedure is shown at the top and a spectral count interaction table is illustrated at the bottom. Ctrl, control; rep, replicate; freq, frequency. (**c**) Modeling spectral count distributions for true and false interactions. For the interaction between prey *i* and bait *j*, SAINT uses all relevant data for the two proteins, as shown in the column of the bait (green) and the data in the row of the prey (orange) in **a** and **b**. (**d**) Probability is calculated for each replicate by application of Bayes rule, and a summary probability is calculated for the interaction pair (*i*,*j*).

scored in reference to the quantitative information for the same prey across purifications of all other baits in the dataset. Although this is possible for large datasets such as the yeast kinase and phosphatase network[8], and the human deubiquitinating (DUB) enzyme interaction network[7] (which each contain more than 75 baits), this unsupervised approach involves additional assumptions and separate treatment of high- and low-frequency prey proteins (Online Methods).

One challenge in modeling AP-MS data is the limited number of replicates that are typically available for each bait. SAINT addresses this problem by inferring individual bait-prey interaction parameters through joint modeling of the entire bait-prey data. To this end, SAINT defines a protein-specific abundance parameter and establishes a multiplicative model in the mixture component distributions. In other words, if prey *i* and bait *j* interact, then the 'interaction abundance' (the spectral count of the prey *i* in purification with bait *j*) is assumed to be proportional to $\alpha_i \times \alpha_j$. Under this assumption, the protein-specific abundance parameters α_i and α_j can be learned not only from the interaction between the two proteins themselves but also from other bona fide interactions that involve either one of them. The same principle applies to false interactions. Hence, SAINT builds a large number of mixture distributions by pooling data (separate mixture distributions for individual prey-bait pairs), but all models are interconnected through the shared abundance parameters.

The probability distributions $P(X_{ij} \mid \text{true})$ and $P(X_{ij} \mid \text{false})$ are then used to calculate the posterior probability of true interaction $P(\text{true} \mid X_{ij})$ (**Fig. 1c**,**d** and Online Methods). For baits profiled in replicates, the next step involves the computation of a combined probability score from independent scoring of each replicate (Online Methods). Finally, SAINT probabilities can be used to estimate the false discovery rate (FDR). By ordering interactions in decreasing order of probability, a threshold can be selected that considers the average of the complement probabilities as the Bayesian FDR[9]. Although the accuracy of FDR estimates remains to be validated, the availability of an objective reliability measure that has been widely used is an advantage over other methods.

We first tested performance of the generalized SAINT model using a human dataset[6] centered around four key protein complexes that are involved in chromatin remodeling: prefoldin, hINO80, SRCAP and TRRAP or TIP60 (referred to as the TIP49 dataset). Although the original work focused the analysis on the interaction network between a core set of 65 proteins, here we analyzed the entire dataset provided by the authors of that study. The dataset consists of 27 baits (35 purifications) and 1,207 preys, and yielded 5,521 unfiltered interactions. The dataset also included 35 negative controls, which allows semisupervised modeling (**Fig. 1a** and **Supplementary Table 1**).

We applied SAINT to these data and compared the results to PP-NSAF[6] and CompPASS *Z* and D^N scores[7,10], which we reimplemented in-house (Online Methods). We note that PP-NSAF[6] removes all interactions involving prey proteins for which the sum of squared NSAF values across the negative control purifications is higher than that in the experiments that contain bait proteins. CompPASS is the only method that does not incorporate negative controls in scoring.

SAINT selected 1,375 interactions at the probability threshold 0.9, which was approximately equivalent to an estimated FDR of 2%. In PP-NSAF, as arbitrary cutoffs were set to define high, moderate and low probability interaction sets, the same number of top-scoring interactions was selected (corresponding to a PP-NSAF probability 0.2 or higher). In CompPASS, the same number of interactions corresponded to a D^N-score threshold of 1.48 (**Supplementary Table 1**).

We evaluated the performance of each algorithm first by benchmarking the selected interactions against two interaction databases named BioGRID[11] and iRefWeb[12] (**Fig. 2a**), and second by assessing the co-annotation rate of interaction partners to common Gene Ontology (GO) terms in 'biological processes' (**Fig. 2b** and **Supplementary Table 1**). SAINT-filtered interactions (with

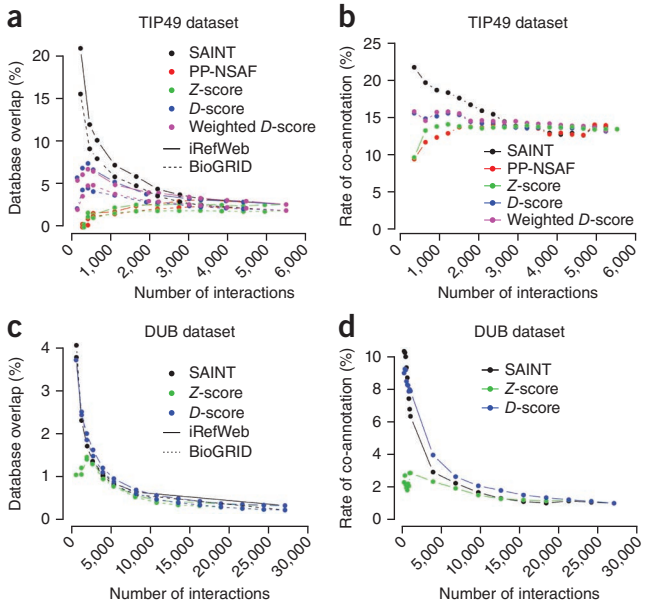

Figure 2 | Analysis of TIP49 and DUB datasets. (**a**) Benchmarking of filtered interactions in the TIP49 dataset by the overlap with interactions previously reported in BioGRID and iRefWeb databases. (**b**) Co-annotation of interaction partners to common GO terms in 'biological processes' in the TIP49 dataset. (**c**) Benchmarking against BioGRID and iRefWeb in the DUB dataset. (**d**) Co-annotation to GO terms in the DUB dataset.

controls) consistently showed the highest overlap with previously reported interactions and co-annotation rates to terms relevant to chromatin remodeling, including histone acetylation, protein amino acid acetylation, chromatin organization and modification, and cellular macromolecular complex assembly. Variation of the SAINT probability thresholds (~0.8–0.95) did not qualitatively change this conclusion (data not shown). Note that omission of negative controls from SAINT modeling decreased the overlap with the interactions reported in BioGRID and iRefWeb (**Supplementary Fig. 1**). Explicit incorporation of the negative control data improved the robustness of modeling, especially in small or medium datasets.

We then tested the performance of SAINT for large-scale datasets without negative controls (**Fig. 1b**) on the human deubiquitinating enzymes (DUB) dataset[7] (this dataset was used in the development of CompPASS). High confidence interactions from SAINT were compared to the high confidence set from CompPASS (**Supplementary Table 2**). Owing to the absence of negative controls, it was not possible to apply PP-NSAF to this dataset. SAINT probabilities and D^N scores were notably correlated (Pearson correlation, $r = 0.79$). At a probability threshold of 0.8, SAINT selected 1,300 interactions, whereas a threshold of CompPASS $D^N \geq 1$ (as used in ref. 7) reported 1,377 interactions. Of these, 1,051 interactions were identified by both methods. Reflecting the similarity of selected interactions, SAINT and CompPASS recovered previously reported interactions at comparable rates (**Fig. 2c**). In the top 1,000 interactions, SAINT showed higher overlap with published data. The co-annotation of interaction partners to the common GO terms also showed similar results between the two methods (**Fig. 2d**), including relevant terms such as positive and negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle, proteasome, and so on (**Supplementary Table 2**). Although SAINT and CompPASS recovered largely overlapping interactions, SAINT removed the interactions identified with 1–2 spectral counts, which were still scored by CompPASS if they were specific to a single bait protein and detected in duplicates.

Another advantage of SAINT over other methods is that it is applicable to the analysis of small-scale datasets for which control purifications are available; this extends to the case of a single bait. We illustrate this by using a recent dataset[13] that contains three experimental purifications of the bait CDC23 and three control purifications. In the original analysis, the authors of the study identified true interactions using ion intensity–based quantification followed by a simple *t*-test. We applied the SAINT approach to the same dataset by using spectral counts (the data were re-searched in-house; Online Methods). The results obtained by SAINT were nearly identical to those in the initial report (**Supplementary Table 3**), the sole exception being the single peptide hit C11orf51, which was reported as a new interactor in the original analysis[13] but which was removed by SAINT.

The SAINT model presented here is based on label-free quantification using spectral counts, a parameter that is easily extracted from most AP-MS datasets. SAINT can also be extended to model other types of quantitative parameters such as peptide ion intensity[14] or other continuous variables[15], which can be accommodated by simply substituting the likelihood with an appropriate continuous distribution. SAINT is available as **Supplementary Software**.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

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

ACKNOWLEDGMENTS

Supported by grants from the Canadian Institute of Health Research to A.-C.G. (MOP-84314) and M.T. (MOP-12246); the US National Institutes of Health to M.T. (5R01RR024031), A.I.N. and A.-C.G. (R01-GM094231), and A.I.N. (R01-CA126239); a Royal Society Wolfson Research Merit Award and a Scottish Universities Life Sciences Alliance Research Chair to M.T.; a Canada Research Chair in Functional Proteomics to A.-C.G.; and the Lea Reichmann Chair in Cancer Proteomics to A.-C.G. We thank M. Sardiu, M. Washburn and M. Sowa for providing additional information regarding the datasets used in this work, G. Bader for discussions and W. Dunham for reading the manuscript.

AUTHOR CONTRIBUTIONS

H.C. and A.I.N. developed, implemented and tested the SAINT method; H.C. wrote the software; B.L., A.B., Z.-Y.L., A.-C.G. and M.T. generated data for the initial SAINT modeling and provided feedback on the model performance; D.M. and D.F. assisted with data analysis and processing; Z.S.Q. contributed to statistical model development; H.C., A.-C.G. and A.I.N. wrote the manuscript; A.I.N. and A.-C.G. conceived the study; A.I.N. directed the project with input from A.-C.G.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Ewing, R.M. *et al. Mol. Syst. Biol.* **3**, 89 (2007).
2. Gavin, A.C. *et al. Nature* **440**, 631–636 (2006).
3. Jeronimo, C. *et al. Mol. Cell* **27**, 262–274 (2007).
4. Krogan, N.J. *et al. Nature* **440**, 637–643 (2006).
5. Nesvizhskii, A.I., Vitek, O. & Aebersold, R. *Nat. Methods* **4**, 787–797 (2007).
6. Sardiu, M.E. *et al. Proc. Natl. Acad. Sci. USA* **105**, 1454–1459 (2008).

7. Sowa, M.E., Bennett, E.J., Gygi, S.P. & Harper, J.W. *Cell* **138**, 389–403 (2009).
8. Breitkreutz, A. *et al. Science* **328**, 1043–1046 (2010).
9. Müller, P., Parmigiani, G & Rice, K. in *Bayesian Statistics* Vol. 8 (eds., Bernardo, J.M. *et al.*) 349–370 (Oxford University Press, 2007).
10. Behrends, C., Sowa, M.E., Gygi, S.P. & Harper, J.W. *Nature* **466**, 68–76 (2010).
11. Breitkreutz, B.J. *et al. Nucleic Acids Res.* **36**, D637–D640 (2008).
12. Turner, B. *et al. Database (Oxford)* **2010**, baq023 (2010).
13. Hubner, N.C. *et al. J. Cell Biol.* **189**, 739–754 (2010).
14. Rinner, O. *et al. Nat. Biotechnol.* **25**, 345–352 (2007).
15. Griffin, N.M. *et al. Nat. Biotechnol.* **28**, 83–89 (2010).

ONLINE METHODS

Label-free quantification by spectral counting. Label-free quantification in this work was based on spectral counting. Spectral counts are the sum of every successful instance of sequencing a peptide from a particular protein by mass spectrometry, including redundant spectra. With proper normalization, spectral counts can be used as a quantitative measure of protein abundance in the sample. This method is conceptually similar to the approach of measuring gene expression using SAGE, EST or RNA-Seq fragment count data. For both DUB and TIP49 datasets, spectral count data were taken exactly as provided by the authors[6,7]. Briefly, the DUB dataset and the TIP49 dataset were searched using SEQUEST[16] using target-decoy database strategies against human databases; selected parameter sets were defined by the authors for filtering. The DUB dataset accepted peptides based on the following criteria. (i) High-stringency set: XCorr 2+ ≥ 2.5; 3+ ≥ 3.2; 4+ ≥ 3.5; +1 charge states were not collected. (ii) Complementary peptide set for proteins identified with high confidence: XCorr thresholds ≥ 1.0; ΔCn ≥ 0.05. The parameters selected by the authors of the TIP49 dataset were: XCorr 1+ ≥1.8 for 2+ ≥ 2.5, and 3+ ≥3.5 (fully tryptic peptides of at least seven amino acids long with max Sp score of 10). The reported spectral FDR for the entire TIP49 dataset was 0.065%; for the DUB data, a set FDR of 2% was selected to populate the interaction tables. No control data were used for the DUB dataset. In the case of the TIP49 dataset, 35 controls were provided alongside the experimental samples. These controls were generated from HeLa and HEK293 cell lines under nine different conditions. We merged the 35 measurements to 9 by taking the largest spectral count for each prey in each condition (**Supplementary Table 1**). For the analysis of the CDC23 data, the data were downloaded from Tranche (trancheproject.org), and re-searched in-house using X!Tandem/*k*-score against the RefSeq database using search parameters similar to those used in ref. 13. The search results were processed using PeptideProphet and ProteinProphet[5], and filtered to achieve a protein-level FDR of less than 0.5%. The spectral counts were extracted using the in-house software Abacus (D.F. and A.I.N., unpublished data).

SAINT model. This section describes the generalized statistical modeling framework for the datasets with and without control purifications (**Fig. 1a,b**). In both cases, the spectral counts for prey *i* in purification with bait *j* are considered to be either from a Poisson distribution representing true interaction (with mean count λ_{ij}) or from a Poisson distribution representing false interaction (with mean count κ_{ij}). In the form of probability distribution, we write

$$P(X_{ij} \mid \bullet) = \pi_T\, P(X_{ij} \mid \lambda_{ij}) + (1 - \pi_T)\, P(X_{ij} \mid \kappa_{ij}) \qquad (1)$$

where π_T is the proportion of true interactions in the data, and dot notation represents all relevant model parameters estimated from the data (here, specifically for the pair of prey *i* and bait *j*). The individual bait-prey interaction parameters λ_{ij} and κ_{ij} are estimated from joint modeling of the entire bait-prey association matrix, with the probability distribution (likelihood) of the form $P(X \mid \bullet) = \prod_{i,j} P(X_{ij} \mid \bullet)$. The proportion π_T is also estimated from the model, which relies on latent variables in the sampling algorithm (see below).

When at least three control purifications are available, and assuming that the control purifications provide a robust representation of nonspecific interactors, the parameter κ_{ij} can be estimated from spectral counts for prey *i* observed in the negative controls. This is equivalent to assuming

$$P(X_{ij} \mid \bullet) = \prod_{i,j:\, j \in E} (\pi_T\, P(X_{ij} \mid \lambda_{ij}) + (1 - \pi_T)\, P(X_{ij} \mid \kappa_{ij})) \times \prod_{i,j:\, j \in C} (P(X_{ij} \mid \kappa_{ij})) \qquad (2)$$

where *E* and *C* denote the group of experimental purifications and the group of negative controls, respectively. This leads to a semisupervised mixture model in the sense that there is a fixed assignment to false interaction distribution for negative controls. As negative controls guarantee sufficient information for inferring model parameters for false interaction distributions, Bayesian nonparametric inference using Dirichlet process mixture priors can be used to derive the posterior distribution of protein-specific abundance parameters in the model. As a result, the mean parameters in the Poisson likelihood functions follow a nonparametric posterior distribution, allowing more flexible modeling at the proteome level. Under this setting, all model parameters are estimated from an efficient Markov chain Monte Carlo algorithm[17].

To elaborate on the two distributions, the mean parameter for each distribution is assumed to have the following form. For false interactions, it is assumed that spectral counts follow a Poisson distribution with mean count

$$\log(\kappa_{ij}) = \log(l_i) + \log(c_j) + \gamma_0 + \mu_i \qquad (3)$$

where l_i is the sequence length of prey *i*, and c_j is the bait coverage, the spectral count of the bait in its own purification experiment, γ_0 is the average abundance of all contaminants and μ_i is prey *i* specific mean difference from γ_0. For true interactions, it is assumed that spectral counts follow a Poisson distribution with mean count

$$\log(\lambda_{ij}) = \log(l_i) + \log(c_j) + \beta_0 + \alpha_{bj} + \alpha_{pi} \qquad (4)$$

where β_0 is the average abundance of prey proteins in those cases where they are true interactors of the bait, α_{bj} is bait *j* specific abundance factor and α_{pi} is prey *i* specific abundance factor. In other words, the mean spectral count for a prey protein in a true interaction is calculated using a multiplicative model combining bait- and prey-specific abundance parameters. This formulation substantially reduces the number of parameters in the model, avoiding the need to estimate every λ_{ij} separately.

For datasets without negative control purifications, the mixture component distributions for true and false interactions have to be identified solely from experimental (noncontrol) purifications. In this case, a user-specified threshold is applied to divide preys into high-frequency and low-frequency groups, denoted as $Y_i = 1$ or 0 if prey *i* belongs to the high- or low-frequency group, respectively. An arbitrary 20% threshold was applied in the case of the DUB dataset; however, the results were not very sensitive to the choice of the threshold. For preys in the high frequency group, the model considers spectral counts for the observed prey proteins (ignoring zero count data, which represent the absence of protein identification), as there are sufficient data to estimate distribution parameters. In the low-frequency group, nondetection of a prey is included to help the separation of high-count from low-count hits. The entire mixture model can then be expressed as

doi:10.1038/nmeth.1541

$$P(X_{ij} \mid \bullet) = \Pi_{i,j} \left(\pi_T P(X_{ij} \mid \lambda_{ij}) + (1 - \pi_T) P(X_{ij} \mid \kappa_{ij})\right)^{Zij} \quad (5)$$

where $Z_{ij} = 1(Y_i = 0) + 1(Y_i = 1, X_{ij} > 0)$ and the false and true interaction distributions are modeled by equations (3) and (4), respectively.

The posterior probability of a true interaction given the data is computed using Bayes rule

$$P(\text{true}|X_{ij}) = T_{ij} / (T_{ij} + F_{ij}) \quad (6)$$

where $T_{ij} = \pi_T P(X_{ij} \mid \lambda_{ij})$ and $F_{ij} = (1 - \pi_T) P(X_{ij} \mid \kappa_{ij})$. If there are replicate purifications for bait j, the final probability is computed as an average of individual probabilities over replicates. Note that one alternative approach is to compute the probability assuming conditional independence over replicates, that is, $\prod_{k \in j} P(X_{ijk} \mid \lambda_{ijk})$ and $\prod_{k \in j} P(X_{ijk} \mid \kappa_{ijk})$ for true and false interactions, with additional index k denoting replicates for bait j. Unlike average probability, this probability puts less emphasis on the degree of reproducibility, and thus may be more appropriate in datasets where replicate analysis of the same bait is performed using different experimental conditions (for example, purifications using different affinity tags) to increase the coverage of the interactome.

When probabilities have been calculated for all interaction partners, the Bayesian false discovery rate (FDR) can be estimated from the posterior probabilities as follows. For each probability threshold p^*, the Bayesian FDR is approximated by

$$\text{FDR}(p^*) = (\Sigma_k 1(p_k \geq p^*)(1 - p_k)) / (\Sigma_k 1(p_k \geq p^*)) \quad (7)$$

where p_k is the posterior probability of true interaction of protein pair k. The output from SAINT allows the user to select a probability threshold to filter the data to achieve the desired FDR.

Implementation of other scores. CompPASS[7,10] calculates two different scores. First, Z score is constructed by mean centering and scale normalization in the conventional Z statistic, where mean and s.d. are estimated from the data for each prey. D score is based on the spectral count adjusted by a scaling factor that reflects the reproducibility of prey detection over replicate purifications of the same bait. If X_{ij} denotes the spectral count between prey i and bait j, then $D_{ij} = ((k / f_i)^{pij} \cdot X_{ij})^{1/2}$, where k is the total number of baits profiled in the experiment, f_i is the number of experiments in which prey i was detected and p_{ij} is the number of replicate experiments of bait j in which prey i was detected. After computing the scores, a threshold D^T is selected from simulation data so that 95% of the simulated data falls below the chosen threshold. Note that CompPASS merges replicate data for bait j to produce a unique spectral count X_{ij} for a given pair. In doing so, it takes nonzero counts only when the prey is identified in a single replicate or otherwise averages counts over multiple replicates. In the analysis of TIP49 dataset, we used both the original D score and the more recently implemented 'weighted D score', which is designed for datasets with large protein complexes[10]. The weighted D scores are shown in **Figure 2** for the TIP49 dataset.

To replicate PP-NSAF[6], we removed 330 contaminants from the dataset using the vector magnitude approach. After filtering, probabilities were computed using an in-house script following the method presented in ref. 6. Although our implementation did not reproduce exactly the same scores for the interactions reported in ref. 6, the scores computed by the in-house implementation showed a clear linear correspondence to the reported scores (Pearson correlation 0.89).

Software. The source C code and a user manual for the generalized SAINT model described in this work (SAINT 2.0) can be downloaded from http://saint-apms.sourceforge.net/, where updates will be distributed. The published version is also available as **Supplementary Software**.

16. Eng, J.K., McCormack, A.L. & Yates, J.R.I. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
17. Ishwaran, H. & James, L.F. *J. Am. Stat. Assoc.* **96**, 161–173 (2001).