An Advanced Method for Identifying Protein-Protein Interaction by Analyzing TAP/MS Data

Xiaoyun Sun, Pengyu Hong
Department of Computer Science
Brandeis University
Waltham, MA USA
{xsun, hongpeng}@brandeis.edu

Meghana Kulkarni, Young Kwon,
Norbert Perrimon
Department of Genetics
Harvard Medical School
Boston, MA USA

Abstract—Tandem affinity purification coupled with mass-spectrometry (TAP/MS) analysis has been increasingly used to identify novel endogenous protein-protein interactions (PPIs) in a high-throughput manner. Computational analysis of TAP/MS data is critical, however, and remains challenging because of the noisy nature of the data. We developed an advanced method for identifying PPIs from TAP/MS data. Our approach APPIC, which stands for Advanced Protein-Protein Interactions Capturing, incorporates an improved statistical method and is more powerful than existing tools.

Keywords-PPIs, TAP/MS, APPIC, Spectral Counts

I. INTRODUCTION

Protein-protein interactions (PPIs) play critical roles in many cellular processes, such as signaling cascades and regulatory complex formation. In addition, information acquired from PPI data is definitive (i.e., protein A and B interact with each other) and could contain quantitative features (i.e., the strength of interactions could vary). Thus PPI data sets have been core resources for biological network construction [1, 2], which is important for understanding the behaviors and functions of biological systems.

Three biochemical experiments are widely used to identify PPIs: Co-immunoprecipitation (Co-IP), yeast two-hybrid [3] and tandem affinity purification (TAP) coupled with mass-spectrometry (MS) [4-6]. Co-IP is a classic method to identify interactions between specific proteins. Although it is straightforward, it lacks the ability to explore interactions at a whole proteome scale. On the other hand, both yeast two-hybrid and TAP/MS are high-throughput technologies that can be used to explore putative interaction partners at large-scales. However, yeast two-hybrid is false-positive prone, and it detects interactions in a heterologous context (with the exception for yeast PPIs). In recent years, TAP/MS has been frequently optimized to detect and study endogenous PPIs, and has been increasingly used to identify novel PPIs under physiologically relevant conditions in large-scale studies [7, 8]. The TAP/MS approach has been successfully used to characterize protein complexes from a variety of cells or multi-cellular organisms [9-11]. In addition, this technique can be combined with quantitative proteomics approaches to better understand the dynamics of protein-complex assembly [12, 13].

With the wide application of the TAP/MS technology, there are increasing numbers of large-scale pathway specific TAP/MS PPI datasets, which are highly valuable resources for understanding molecular details of specific signaling pathways. However, currently available analysis tools are not satisfactory because of high false-positive rate due to non-specific protein binding and missing true PPIs due to weaknesses in analysis methods. To meet the challenges of accurately analyzing these accumulating data sets, new and better computational algorithms and analysis methods are required.

II. RELATED WORKS

One of the major challenges in TAP/MS PPI data analysis is to reduce the false-positive rate while increasing the sensitivity to identify true interactions. Early methods for analyzing TAP/MS PPI data only use the presence/ absence information [14]. Recent methods take into account more quantitative information such as label-free quantitative spectral counts (SCs), which are the number of peptides of a protein detected in MS. Currently, there are three popular methods for analyzing TAP/MS PPI data: Normalized Spectral Abundance Factor (NSAF) [15], Comparative Proteomic Analysis Software Suite (CompPASS) [16], and Significance Analysis of Interactome (SAINT) [17].

NSAF estimates the relative abundance of a prey protein *j* in the purification of a bait protein *i*, i.e., the *NSAF* score, by considering the number of its peptides identified by MS, the length of the prey protein, and the total number of the peptides in the sample:

$$NSAF_{i,j} = \frac{\frac{SC_{i,j}}{L_j}}{\sum_{k=1}^{N} \frac{SC_{i,k}}{L_k}} \quad (1)$$

where $SC_{i,j}$ and L_j are the spectral count and the length of the prey protein j, respectively, and N is the number of preys in purification of the bait protein i. The *NSAF* score is specific to each prey-bait pair in one sample. If there is more than one biological replicate, the average *NSAF* scores is assigned to the prey-bait pair.

CompPASS calculates two indexes for each bait-prey pair: Z score (Z_SC) and D score (D_SC). For a bait protein i, the Z_SC of its prey protein j is:

$$Z\_SC_{i,j} = \frac{SC_{i,j} - mean_j}{std_j} \qquad (2)$$

where $mean_j$ and std_j are the mean and standard deviations of the spectral counts of the prey protein j across all baits. D_SC considers both the reproducibility of the peptide detected across biological replicates and the frequency of each observed prey protein in the purifications with different baits. For a bait protein i, the D_SC of its prey protein j is:

$$D\_SC_{i,j} = \sqrt{SC_{i,j}\left(\frac{K}{\sum_{i=1}^{K} f_{i,j}}\right)^N} \qquad (3)$$

where $f_{i,j} = 1$ if a prey protein j is identified in the purification of a bait protein i, otherwise $f_{i,j} = 0$, N is the number of replicates in which the interaction is detected, and K is the total number of samples. The summation of $f_{i,j}$ indicates the total number of occurrences of an interaction in all samples. The ratio between K and the summation of $f_{i,j}$ is the frequency of a prey protein j being observed across all samples. An interactor with higher detection frequency (frequently observed to interact with many or all baits) will be penalized because it is very likely that it belongs to the sticky protein group that "interact with" many proteins non-specifically. However, when all baits work in the same signaling pathway, some important prey proteins can also interact with most baits and appear to be "sticky". Thus, it is arguable that this feature is necessary. Based on our observations, this feature would be beneficial if there are a large number of baits in the TAP/MS experiments. Otherwise, it may not be a good choice to penalize high detection frequency.

CompPASS also computes a weighted D_SC as:

$$WD\_SC_{i,j} = \sqrt{SC_{i,j}\left(\frac{K}{\sum_{i=1}^{K} f_{i,j}} W_j\right)^N} \qquad (4)$$

where $W_j = std_j / mean_j$, if $std_j / mean_j > 0$, otherwise $W_j = 1$. ($mean_j$ and std_j are the mean and standard deviation of the SCs of the prey protein j in the purification samples of all baits, respectively). CompPASS believes when protein j has higher detection frequency, std_j is more likely to be higher than $mean_j$ if protein j is a true interactor of a bait protein. The weight factor W_j is a multiplicative factor designed to deal with preys frequently detected in the protein samples of all baits, however, with low SCs. If multiple biological replicates are available, CompPASS calculates the Z_SC, D_SC, and WD_SC scores of a prey protein in each sample, and then uses their averages to decide if a prey-bait pair is a true PPI.

SAINT uses a Bayesian approach to estimate $P(true|SC_{i,j})$, which is the probability of true interaction between proteins i and j given $SC_{i,j}$. The probability distribution $P(SC_{i,j}|true)$ and $P(SC_{i,j}|false)$ are assumed to be Poisson. The parameters of $P(SC_{i,j}|true)$ are estimated from joint modeling of the entire bait-prey association matrix in the bait purification samples. The parameters of $P(SC_{i,j}|false)$ are estimated from the spectral counts of prey j observed in the negative controls. The prior probability $P(true)$ is the proportion of true interactions in the data. SAINT then calculates the scoring of a bait-prey pair using Bayes rule:

$$P(true|SC_{i,j}) = \frac{P(SC_{i,j}|true)P(true)}{P(SC_{i,j}|true)P(true) + P(SC_{i,j}|false)P(false)}$$

where $P(false) = 1 - P(true)$. If a bait-prey pair has more than one biological replicates, a SAINT score will be calculated for it in each replicate. Those SAINT scores will then be averaged to represent the bait-prey pair.

The above approaches can effectively analyze the data sets for which they have been developed. However, each one of them has its own intrinsic weaknesses. For example, neither *NSAF* nor CompPASS utilizes information from negative controls, which are extremely valuable for reducing false positives. SAINT over-penalizes interactions that can be true, but "appear" to be non-reproducible. It averages the posterior probability of an interaction across all the biological replicates. Failing to detect an interaction in one of the replicates can drag down the average significantly and cause the algorithm to classify the interaction as false. Therefore, SAINT can miss some statistically-significant PPIs that are not 100% reproducible – a common phenomenon in reality due to experimental variations. With the fast accumulation of large scale TAP/MS PPI datasets, there is a high demand for developing a better algorithm which is capable of eliminating a maximum number of false positives and can accurately capture as many true interactions in TAP/MS PPI datasets as possible.

III. Advanced Protein-Protein Interactions Capturing

We have developed an advanced Protein-Protein Interactions capturing (APPIC) method for analyzing TAP/MS PPI datasets. APPIC is able to capture more true interactions and eliminate more false positives simultaneously. We have successfully applied it to the analysis of pathway-specific TAP/MS PPI datasets generated using *Drosophila* cells.

The scoring function of APPIC is an improvement over those of NSAF and CompPASS. First, we introduce a term

R to indicate the reproducibility of detecting a prey protein across all biological replicates. Differently from the reproducibility term *N* used in CompPASS (see equations 3 & 4), which is simply the number of replicates in which the prey protein is detected, we calculate *R* as *N* divided by the total number of biological replicates *T*. Our definition of the *R* term is relative to the total number of replicates, and hence can interpret experimental reproducibility more accurately. For example, if we have an interaction present in two out of three biological replicates, the value of *N* in CompPASS equals 2. It is not clear how high/low this number is. By contrast, our *R* value is 2/3, and the highest it can get is 1. A bait-prey pair is more likely to be a true PPI if it is detected in many replicates even though it SCs are relatively low. On the other hand, it is less likely to be a true PPI if it has relatively few sporadic high SCs. Taking consideration of the reproducibility factor, we apply *R* to the power of the weighted frequency term:

$$\left(W_j \times \frac{K}{\sum_{i=1}^{K} f_{i,j}} \right)^R \quad (5)$$

Second, we would like to utilize negative controls in a more meaningful way. Among previous works, only SAINT [17] utilizes the negative control data to maximally reduce the false positive rate. There are many abundant cellular proteins that can bind to the purification beads non-specifically and become contaminants in the TAP/MS samples. SAINT uses a complicated method to estimate the spectral count distribution for false interactions directly from negative controls. In contrast, we utilize the negative control data sets to compute a false discovery rate (FDR) for each observed prey protein, which not only can be used to help filter out the majority of false positives, but also can rescue some interactors that are abundant in bait purification samples while having very low SCs in negative controls (most likely noise in negative controls). Using the method proposed in [18], we can compute the FDR of a prey protein *j* as:

$$FDR_{i,j} = \frac{1}{1 + \frac{1}{e \times p_{i,j} \times \log(p_{i,j})}} \quad (6)$$

where *e* is the base of the natural logarithm, and $p_{i,j}$ is the *p*-value generated by applying the rank-sum test to compare two populations: the SCs of protein *j* in the purification samples of bait *i vs* the SCs of protein *j* in the negative controls. Finally, we have the following function to score any given bait-prey pair (i, j) given its MS data:

$$APPIC_{i,j} = NSAF_{i,j} \times (1 - FDR_{i,j})^T \times \left(\frac{W_j \times K}{\sum_{i=1}^{i=K} f_{i,j}} \right)^R \quad (7)$$

where $NSAF_{i,j}$ is the averaged NSAF score if there are multiple replicates available.

The APPIC of a bait-prey depends on the total number of TAP/MS samples, the number of replicates, and the SC of the prey in each individual replicate. It can be used to rank the bait-prey pairs in each data set. The higher the score, the more significant the interaction is. Users can use known interactions to set up their threshold for obtaining a PPI network.

IV. EXPERIMENTS

A. Data sets

We applied APPIC to analyze two pathway-specific TAP/MS PPI data sets in *Drosophila melanogaster*: the Insulin pathway (20 baits with 3 biological replicates) with three time points (0, 10, 30 minutes) of Insulin treatment and the Hippo pathway (12 baits with 3 biological replicates).

Insulin signaling pathway plays an important role in the control of embryonic development and growth, reproduction and appetite regulation in animals [19, 20]. It is clinically associated with Diabetes Mellitus disease in humans [21, 22]. Thus, defects in Insulin signaling can lead to a range of systemic disorders, such as hypertension, high levels of cholesterol, heart and kidney diseases, and female infertility problems. It is an important research area to understand the downstream signaling pathway deregulation. Since the Insulin signaling system is highly conserved across all metazoans, including *Drosophila* and mammals, we selected *Drosophila* as our model system. Twenty major known components (*Akt, chico, dm, foxo, gig, lkb1, InR, melt, Pi3K21B, Pi3K92E, Pk61C, pten, Rheb, S6K, S6KII, sima, B4, rictor, Thor and TSC1*) in Insulin signaling pathway have been selected as baits in TAP/MS experiments. Three biological replicates are performed for all the purifications.

In addition, we conducted experiments and analyses with the Hippo pathway in *Drosophila melanogaster*. Hippo signaling is a conserved pathway that controls organ size by regulating cell proliferation and survival [23, 24]. Interestingly, a number of components in the Hippo pathway also play important roles in the suppression or formation of cancer [24]. For example, the upstream signaling components in the Hippo signaling, such as *fat*, *dachsous*, *Merlin*, *hippo*, *Salvador* and *warts* can suppress cancer growth by inhibiting *Yorkie* [25], which is an important factor required for tumor growth. Nevertheless, the regulation of this pathway in cancer is poorly understood. Therefore, new signaling components and regulators in the Hippo pathway identified from *Drosophila* will provide new insights into cancer biology. Twelve major known components (*yorkie, scalloped, expanded, Salvador, warts, hippo, fat, Merlin, dachsous, fj, mats and dachs*) in the Hippo signaling pathway have been selected as baits in TAP/MS experiments. Three biological replicates are performed for all the purifications.

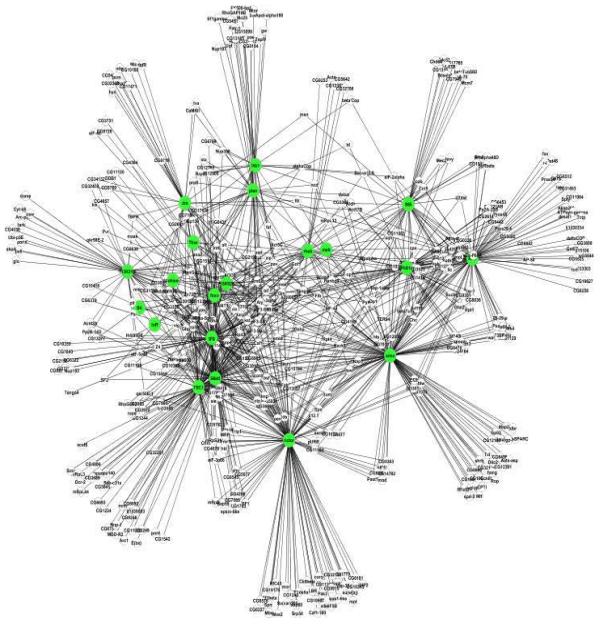

Figure 1. A map of predicted protein-protein interactions (PPIs) in the Insulin pathway computed by APPIC. Bait proteins are highlighted in green.

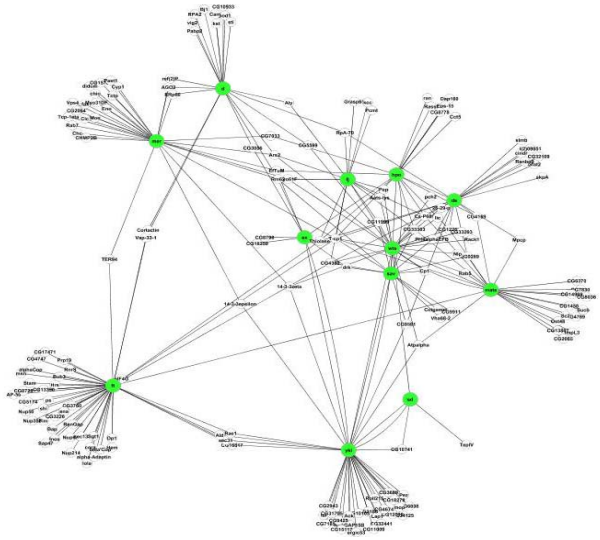

Figure 2. A map of predicted PPIs in the Hippo pathway computed by APPIC. Bait proteins are highlighted in green.

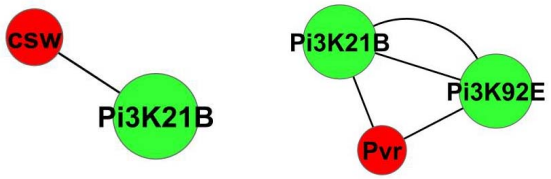

Figure 3. Cross talk between different signal pathways. Two bridge interactions between Insulin and EGF pathways are listed in above. The curve line between Pi3K21B and Pi3K92E indicates the mutually scored interactions between those two baits.

B. Results

Using APPIC, we identified 509 proteins with 1419 interactions in the Insulin pathway and 191 proteins with 286 interactions in the Hippo pathway, respectively. The PPI network built by APPIC includes all but two canonical PPIs between baits in both pathways (the two exceptions are the *dm-Max* interaction in Insulin pathway and *ft-ds* in the Hippo pathway, which may occur in different conditions). We further filtered out heat shock proteins and ribosomal proteins to identify 1080 interactions in the Insulin pathway (Fig. 1) and 255 interactions in Hippo pathway (Fig. 2). Interestingly, the PPI network of the Insulin pathway uncovered novel cross-interactions between the Hippo, Insulin, JAK/STAT, EGF and stress signaling pathways. Two examples of the cross-interactions between the Insulin and EGF pathways are shown in Fig. 3.

To evaluate the quality of APPIC results, we first examined whether APPIC was capable of identifying more known interactions than other approaches, such as NSAF, CompPASS *Z*_SC, CompPASS *D*_SC, and SAINT. We collected known interactions from BioGRID PPI database (http://www.thebiogrid.org/). We only selected physical interactions since genetic interactions could be indirect and may contain more false positives. In total, we retained 27156 previous reported PPIs for *Drosophila*. We then counted the overlaps between the top interactions predicted by each algorithm and the known interactions. The results are shown in Fig. 4 and Fig. 5. The interactions identified by APPIC consistently demonstrated the highest overlap with the known interactions.

V. CONCLUSIONS AND DISCUSSIONS

We have showed that APPIC is able to identify more PPIs with higher accuracy than other methods, such as NSAF, CompPASS, and SAINT. Especially, APPIC can uncover unique and important interactions missed by other approaches, meanwhile efficiently reducing false-positive interactions. For example, APPIC will filter out an interaction with only one SC in each of the three replicates because one SC is very likely due to experiment error or noise. However, SANIT will call this interaction as a significant one because it appears in all three replicates. An interaction that has high SCs in all replicates but has significantly low SCs in negative controls will be assigned a significant score by APPIC. Taking the bait-prey pair *sima–hang* as the example, the SCs of *hang* in *sima*'s three purified replicates are very high: 12, 10, and 6, respectively. However, *hang* has one SC in 1 of 9 control replicates. This interaction is mysteriously filtered out by SAINT, but is

captured by APPIC. It is known that *sima* translocates to the nucleus upon Insulin stimulation to activate downstream target genes [26]. Similarly, *hang* was shown to be a predominantly nuclear protein [27]. Thus it is possible that in Insulin signaling pathway, *hang* can either positively or negatively regulate *sima* activity through direct interactions. It should also be noted that *hang* is broadly expressed in the *Drosophila* nervous system, and is a critical component in the stress pathway for alcohol tolerance development [27]. It was recently reported that *hang* negatively regulates bouton addition at *Drosophila* neuromuscular junctions [28]. Since the Insulin pathway has been implicated in the control of many neuronal activities, such as aging, cognitive maintenance and food intake control [29, 30], it will be interesting and important to study these activities with flies deficient for *hang* gene expression, and characterize the detailed roles of *hang* in the Insulin signaling pathway.

To more broadly evaluate the performance of APPIC, we

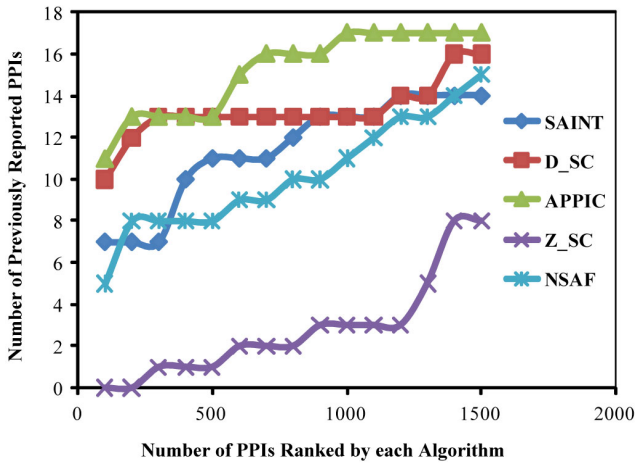

Figure 4. The performance of five algorithms for Insulin pathway analysis. APPIC (in green) demonstrated the ability of reporting more known interactions than other approaches.

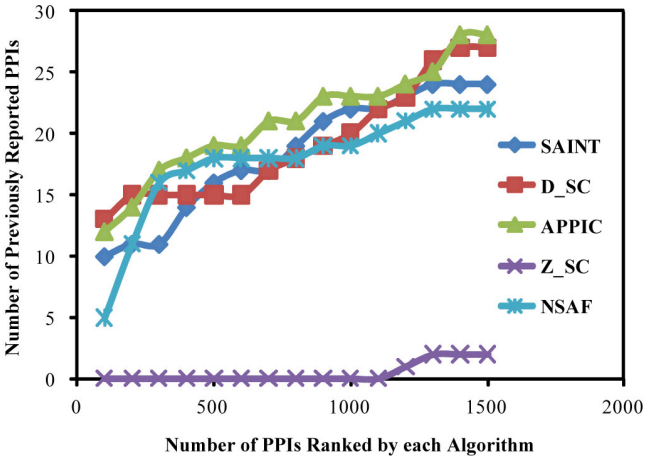

Figure 5. The performance of five algorithms for Hippo pathway analysis. APPIC (in green) demonstrated the ability of reporting more known interactions than other approaches.

examine its ability of identifying known interactions in general. TABLE I compares the canonical interactions in the Insulin pathway detected at three different time points (0, 10 and 30 minutes) by APPIC and/or SAINT. Here, equivalent thresholds are set as APPIC score ≥ 0.005 and SAINT score ≥ 0.9843 to select the same number of significant interactions in each analysis method. Two interactions *lkb1* – *Mo25* at 10 minutes and *dm* – *Max* at 0 minutes were not identified by both methods due to poor SC recovery. These two interactions may occur in different experimental conditions. The interactions *rictor* – *Sin1* at 30 minutes, *TSC* – *gig* at 0 and 30 minutes, and *InR* – *chico* at 0 minute were captured by APPIC but not SAINT. These examples show that APPIC has a better ability to achieve a good balance between penalizing the non-reproducibility in tagged samples and penalizing sporadic appearances in control due to experimental noise.

APPIC can also "rescue" some interactions that seem to be non-reproducible mainly due to noise. For instance, a known interaction *InR* – *chico* has SC = 3/21/0 in three replicates (TABLE I). That is, it is not observed in one of the three replicates. This interaction is filtered out by SAINT because it heavily penalizes the non-reproducibility, but in APPIC, this interaction is considered as significant since it has zero SC in 6 control replicates and relatively high SCs in two replicates, which can compensate occasional SC losses in experiments. *S6KII* – *Shc* is another example identified by APPIC, but not by SAINT. This pair has the observed SCs = 15/3/0 in three replicates before Insulin treatment. *Shc* is a known adaptor protein interacting with some receptor tyrosine kinases including Insulin receptor [31]. The interaction between these two proteins after Insulin treatment was not observed (data not shown). It is likely that *Shc* interacts with *S6kII* to constitutively inhibit the activity of *S6kII*. The activation of the Insulin pathway somehow disrupts the interaction between them, thus relieving the inhibition. It is therefore interesting to investigate the molecular mechanism of the Insulin-induced dissociation between *Shc* and *S6kII*, and determine whether it is phosphorylation-dependent and/or involving induced protein degradation.

In summary, we developed an advanced statistical model APPIC for analyzing TAP/MS experimental data to detect PPIs. It is based on label-free quantification by spectral counts and incorporates the frequency and reproducibility of observed interactors across all biological replicates. The false positive rate of interactors interpreted from statistical *p*-values is integrated in the calculation to filter out non-specific interactors. This approach has been successfully applied to the analysis of two pathway-specific TAP/MS data sets in *Drosophila*: the Insulin and Hippo pathways. The result shows that APPIC is capable of identifying more known interactions than previous approaches and yields high confidence PPI results. The statistical analysis of APPIC requires three or more biological replicates and

negative control data sets. While these steps demand higher experimental cost, they are necessary for producing high confidence results. We demonstrated that APPIC can serve as an exceptional alternative to existing approaches.

TABLE I. CANONICAL INTERACTIONS IN INSULIN PATHWAY SCORED BY APPIC (A) AND SAINT(S). EQUIVALENT THRESHOLD USED FOR SIGNIFICAN INTERACTIONS: APPIC ≥ 0.005 AND SAINT ≥ 0.9843. THREE TIMES POINT ARE SHOWN HERE (0, 10, AND 30 MINUTES). Y = SIGNIFICANT INTERACTIONS, N = NONSIGNIFICANT INTERACTIONS.

bait	prey	A	S	Time	SCs
chico	*14-3-3epsilon*	Y	Y	0	25, 21, 21
chico	*14-3-3epsilon*	Y	Y	10	53, 50, 50
chico	*14-3-3zeta*	Y	Y	0	8, 13, 15,
chico	*14-3-3zeta*	Y	Y	10	34, 20, 32
dm	*Max*	N	N	0	0, 0, 1
gig	*TSC1*	Y	Y	0	5, 8, 3
gig	*TSC1*	Y	Y	10	3, 4, 5
gig	*TSC1*	Y	N	30	8, 3, 0
TSC1	*gig*	Y	N	0	4, 0, 4
TSC1	*gig*	Y	Y	10	7, 5, 4
TSC1	*gig*	Y	N	30	12, 12, 0
InR	*chico*	Y	N	0	3, 21, 0
InR	*chico*	Y	Y	10	19, 17, 16, 7
InR	*chico*	Y	Y	30	12, 6
chico	*InR*	Y	Y	0	17, 21, 21
chico	*InR*	Y	Y	10	8, 8, 9
lkb1	*Mo25*	Y	Y	0	16, 3, 1
lkb1	*Mo25*	N	N	10	0, 1, 3
Pi3K21B	*chico*	Y	Y	0	11, 8, 3
Pi3K21B	*chico*	Y	Y	10	18, 25, 18
Pi3K21B	*chico*	Y	Y	30	25, 31, 3
chico	*Pi3K21B*	Y	Y	10	2, 3, 9
Pi3K92E	*Pi3K21B*	Y	Y	0	41, 13, 33
Pi3K92E	*Pi3K21B*	Y	Y	10	38, 32, 31
Pi3K92E	*Pi3K21B*	Y	Y	30	52, 2, 5
Pi3K21B	*Pi3K92E*	Y	Y	0	34, 39, 2
Pi3K21B	*Pi3K92E*	Y	Y	10	56, 57, 52
Pi3K21B	*Pi3K92E*	Y	Y	30	51, 58, 69
rictor	*Sin1*	Y	N	30	0, 3, 3
S6kII	*RpS6*	Y	Y	0	4, 6, 7
S6kII	*RpS6*	Y	Y	10	8, 5, 6
S6kII	*RpS6*	Y	Y	30	8, 8, 11
Thor	*eIF-4E*	Y	Y	0	60, 64, 40
Thor	*eIF-4E*	Y	Y	10	39, 52, 45
Thor	*eIF-4E*	Y	Y	30	60, 34, 46

REFERENCES

[1] Dittrich, M.T., et al., Identifying functional modules in protein-protein interaction networks: an integrated exact approach. Bioinformatics, 2008. **24**(13): p. i223-31.

[2] Gehlenborg, N., et al., Visualization of omics data for systems biology. Nat Methods, 2010. **7**(3 Suppl): p. S56-68.

[3] Fields, S., High-throughput two-hybrid analysis. The promise and the peril. FEBS J, 2005. **272**(21): p. 5391-9.

[4] Gavin, A.C., et al., Proteome survey reveals modularity of the yeast cell machinery. Nature, 2006. **440**(7084): p. 631-6.

[5] Krogan, N.J., et al., Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature, 2006. **440**(7084): p. 637-43.

[6] Jeronimo, C., et al., Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK APPICng enzyme. Mol Cell, 2007. **27**(2): p. 262-74.

[7] Veraksa, A., A. Bauer, and S. Artavanis-Tsakonas, Analyzing protein complexes in Drosophila with tandem affinity purification-mass spectrometry. Dev Dyn, 2005. **232**(3): p. 827-34.

[8] Friedman, A.A., et al., Proteomic and functional genomic landscape of receptor tyrosine kinase and ras to extracellular signal-regulated kinase signaling. Sci Signal, 2011. **4**(196): p. rs10.

[9] Gingras, A.C., et al., Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol, 2007. **8**(8): p. 645-54.

[10] Knuesel, M., et al., Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. Mol Cell Proteomics, 2003. **2**(11): p. 1225-33.

[11] Bouwmeester, T., et al., A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol, 2004. **6**(2): p. 97-105.

[12] Gavin, A.C., et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002. **415**(6868): p. 141-7.

[13] Wang, X. and L. Huang, Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. Mol Cell Proteomics, 2008. **7**(1): p. 46-57.

[14] Zhu, X., M. Gerstein, and M. Snyder, Getting connected: analysis and principles of biological networks. Genes Dev, 2007. **21**(9): p. 1010-24.

[15] Sardiu, M.E., et al., Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1454-9.

[16] Sowa, M.E., et al., Defining the human deubiquitinating enzyme interaction landscape. Cell, 2009. **138**(2): p. 389-403.

[17] Choi, H., et al., SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods, 2011. **8**(1): p. 70-3.

[18] Hubbard, R., et al., Confusion Over Measures of Evidence (p's) versus errors in classical statistical testing. The Anerican Statistician, 2003.

[19] Wu, Q. and M.R. Brown, Signaling and function of insulin-like peptides in insects. Annu Rev Entomol, 2006. **51**: p. 1-24.

[20] Saltiel, A.R. and C.R. Kahn, Insulin signalling and the regulation of glucose and lipid metabolism. Nature, 2001. **414**(6865): p. 799-806.

[21] Ashcroft, F.M. and P. Rorsman, Diabetes mellitus and the beta cell: the last ten years. Cell, 2012. **148**(6): p. 1160-71.

[22] Leibiger, I.B., B. Leibiger, and P.O. Berggren, Insulin signaling in the pancreatic beta-cell. Annu Rev Nutr, 2008. **28**: p. 233-51.

[23] Saucedo, L.J. and B.A. Edgar, Filling out the Hippo pathway. Nat Rev Mol Cell Biol, 2007. **8**(8): p. 613-21.

[24] Pan, D., The hippo signaling pathway in development and cancer. Dev Cell, 2010. **19**(4): p. 491-505.

[25] Kango-Singh, M. and A. Singh, Regulation of organ size: insights from the Drosophila Hippo signaling pathway. Dev Dyn, 2009. **238**(7): p. 1627-37.

[26] Dekanty, A., et al., The insulin-PI3K/TOR pathway induces a HIF-dependent transcriptional response in Drosophila by promoting nuclear localization of HIF-alpha/Sima. J Cell Sci, 2005. **118**(Pt 23): p. 5431-41.

[27] Scholz, H., M. Franz, and U. Heberlein, The hangover gene defines a stress pathway required for ethanol tolerance development. Nature, 2005. **436**(7052): p. 845-7.

[28] Schwenkert, I., et al., The hangover gene negatively regulates bouton addition at the Drosophila neuromuscular junction. Mech Dev, 2008. **125**(8): p. 700-11.

[29] Bishop, N.A., T. Lu, and B.A. Yankner, Neural mechanisms of ageing and cognitive decline. Nature, 2010. **464**(7288): p. 529-35.

[30] Schwartz, M.W., et al., Central nervous system control of food intake. Nature, 2000. **404**(6778): p. 66

[31] Luschnig, S., et al., The Drosophila SHC adaptor protein is required for signaling by a subset of receptor tyrosine kinases. Mol Cell, 2000. **5**(2): p. 231-41.