

The Texas Medical Center Library

DigitalCommons@TMC

The University of Texas MD Anderson Cancer
Center UTHealth Graduate School of
Biomedical Sciences Dissertations and Theses
(Open Access)

The University of Texas MD Anderson Cancer
Center UTHealth Graduate School of
Biomedical Sciences

12-2010

NETWORK TOPOLOGY IN HUMAN PROTEIN INTERACTION DATA PREDICTS FUNCTIONAL ASSOCIATION

HUA LI

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations



Part of the [Bioinformatics Commons](#)

Recommended Citation

LI, HUA, "NETWORK TOPOLOGY IN HUMAN PROTEIN INTERACTION DATA PREDICTS FUNCTIONAL ASSOCIATION" (2010). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 93.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/93

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

The
TMC LIBRARY
Health Sciences Resource Center

CHAPTER 1
INTRODUCTION

Biological Networks are Highly Complex Networks Governed by Universal Laws

In the past century, traditional biology study focused on individual molecules has been a great success on discovering their cellular functions. From genomic sequences to protein structures and beyond, we have already collected unprecedented amount of information on various species, including *Homo sapiens*. It has been increasingly realized that most biological functions are not carried out by merely a single molecule; instead, they always come from very complex interactions of different cellular constituents (e.g. DNA, RNA, proteins) which form a highly sophisticated network [1-9]. Therefore, it becomes a very prevailing idea to treat living cells as a complex network which shares many common characteristics of other networks, such as Internet and social network.

Biological networks, at a highly abstract level, can be reduced to a simple graph with nodes representing its components and links representing the interactions between any two components. Examples include biological networks [10-11], social networks [12-15] and technological networks [16-18]. Biological networks can be either directed or undirected depending on the nature of interaction. For example, gene regulatory networks are always considered as directed networks because regulatory interaction causes changes in the expression of target genes. By contrast, protein interaction networks are undirected since the interactions do not have any assigned direction [9].

In 1960, Paul Erdos and Alfred Renyi proposed a model of the random network and studied its mathematical properties [19]. In the model, they assumed that a fixed number of nodes are randomly connected to each other (i.e. edges are placed randomly among nodes); in

other words, each node has the same probability to be connected by other nodes. In this type of network, the degrees of nodes follow a Poisson distribution, which means it's very hard to observe nodes with significant larger or smaller degree than the average degree of the network [9].

After a series of studies focused on the real networks, such as biological networks and social networks, topological properties were found to be different from what was proposed in the random network [19]. Instead of the Poisson degree distribution, the degrees of nodes actually follow a power-law distribution in most real networks, which means the probability that a node has k links follows $P(k) \sim k^{-\gamma}$ (γ is the degree exponent with values being between 2 and 3 in most cases) [9, 20]. Networks with a power-law degree distribution are called scale-free networks, where most nodes only have a few link and they are connected to so-called “hub” nodes that have a large number of connectivity in the network [20]. In such a network, hubs play a much more important role than those with only a few links, which is totally different from the random network.

Most networks in living cells approximate a scale-free topology, including protein-protein interaction (PPI) networks. In several recent publications, PPI networks have been shown to be scale free in many eukaryotic species, including human [21-27]. For example, the scale-free topology is obvious in a human PPI network (Fig. 1) which shows that most proteins have only a few direct interacting neighbors, while hub proteins have significantly more neighbors.

High-throughput Screening Enables Mapping of Protein-Protein Interaction Networks on a Large Scale

Due to technological advances on DNA sequencing, biologists are now discovering novel genes at an unprecedented speed, which makes it a more challenging and important task to annotate genes' functions [76]. Large-scale mapping of PPI data by high-throughput screenings is one significant step for accomplishing this task, which enables very fast assignment of functional annotations by network-based annotating schemes and algorithms [76]. So far, two methods have been widely used to generate large-scale PPI maps: yeast two-hybrid (Y2H) and co-affinity purification followed by mass spectrometry [25, 26, 28-33].

Y2H screens use mating assays illustrated in Fig. 2: proteins need to be tested for interactions will be expressed as hybrid fusion proteins in yeast [34-36]. In Fig. 2*a*, protein X is fused to a DNA-binding domain (BD) of a transcription factor (TF) which will later bind to the upstream of a reporter gene. Protein Y is fused to a transcription activation domain (AD). If Protein X and Y interact with each other, the TF (usually Gal4 or LexA) will have the function of activation on the reporter gene. Since reporter genes are always required for yeast to grow, only yeasts with the interaction of X and Y will form a colony on media. In some cases, LacZ is used as a reporter whose activity can be detected by blue color on indicator plates. In order to test which AD strains a BD strain may have interaction with, the BD strain is mated to an array of AD strains which effectively tests the BD strain against each AD strain in the array (Fig. 2*b*). To test a large amount of BD strains against AD strains, libraries of AD strains are retrieved from cDNA or pools of AD arrays, and then individual BD strains are

mated with the libraries of AD strains. Clones with reporters activated will survive in the media and be picked for sequencing (Fig. 2c). [36-41]

Affinity purification followed by mass spectrometry allows identification of interacting partners [42]. This technology has been used to explore the relationships between proteins because it allows rapid characterization of any protein present in a complex mixture as long as the complex is purified sufficiently and has enough quantity [42-46]. With the help of complete genome sequencing of several species, proteins can now be identified by mass spectrometry more conveniently [42]. Not only has Affinity purification been used to generate large-scale PPI networks [31, 32], but also has been an independent approach to valid the PPIs obtained from Y2H screens [23].

Network-based Prediction Schemes have been Widely Used to Predict Protein Functions

To date, PPI networks for several organisms, including human, have been mapped partially [23-26, 49-54]. These PPI networks are valuable resources which provide many new insights into protein functions. From a system biology point of view, compiling PPI networks are also very important as they may help to uncover the underlying organization principles of cellular networks when times and locations for PPIs are taken into consideration [24, 55]. For this reason, different algorithms and methods have been formulated to investigate these networks, trying to figure out protein functions [56-71] and their roles in diseases [72-74]. However, due to notorious noises present in the process of high-through put data generation [47, 48], it now becomes critical to develop algorithms which reduce influence of these

noises and improve the overall quality of our functional inferences [76].

The large volume of PPI data in the recent decade has presented the opportunity to systematically analyze the topology of PPI networks and develop network-based statistics for large-scale functional prediction. One popular method to suggest biological function is to compare real PPI networks with similar random networks (i.e. random networks with the same degree distribution) to find unusual topological connectivity among proteins. Typically, if two proteins share an unusual large number of common neighbors, functional association between the two proteins may be claimed [75]. Here we call this method “common-neighbor statistics”. Such statistics has been used to assess the functional relationship between proteins in a yeast PPI network, and functional predictions based on statistical significance have been made from these relationships [75, 76].

Common-neighbor statistics can be further improved to provide better functional prediction

Lots of evidences showed that such human PPI network is nonrandom with respect to its network topology [27, 33, 75, 76, 78]. Here we assumed that most of the nonrandomness is necessary for human PPI networks to perform proper biological functions. We further hypothesize that, in human PPI networks, if two proteins share a number of interacting neighbors which is significantly larger than that occurred on average in randomized scale-free networks (i.e. random networks with the same degree distribution as human PPI networks), the two proteins are more likely to share a common or related biological function. In prior work on yeast PPI network, Samanta and Liang (2003) developed an algorithm (i.e.

common-neighbor statistics) to rank the nonrandomness of network topology, and hence the functional associations between proteins [75]. In this dissertation, we developed an additional algorithm to overcome a deficiency in their previous work and make the ranking more accurate [76]. We found that the combination of these two algorithms lead to better ranking of protein functional associations, thereby enhancing the overall quality of functional inferences. We applied our method to a comprehensive human PPI data set which is publicly available from www.thebiogrid.com [77]. We also developed a clustering method to analyze protein functions and pathways systematically. With the clustering method, we built a cluster consisting of 1729 proteins and we found most functionally related proteins stayed together in the cluster [76]. After cutting the cluster into subclusters, we performed pathway analysis and identified many subclusters significantly enriched in different signaling pathways [76]. In particular, we made an in-depth analysis of the transforming growth factor β (TGF- β) pathway which is important in cell proliferation and tumorigenesis, and suggested a list of proteins presumably involved in several signaling pathways [76].

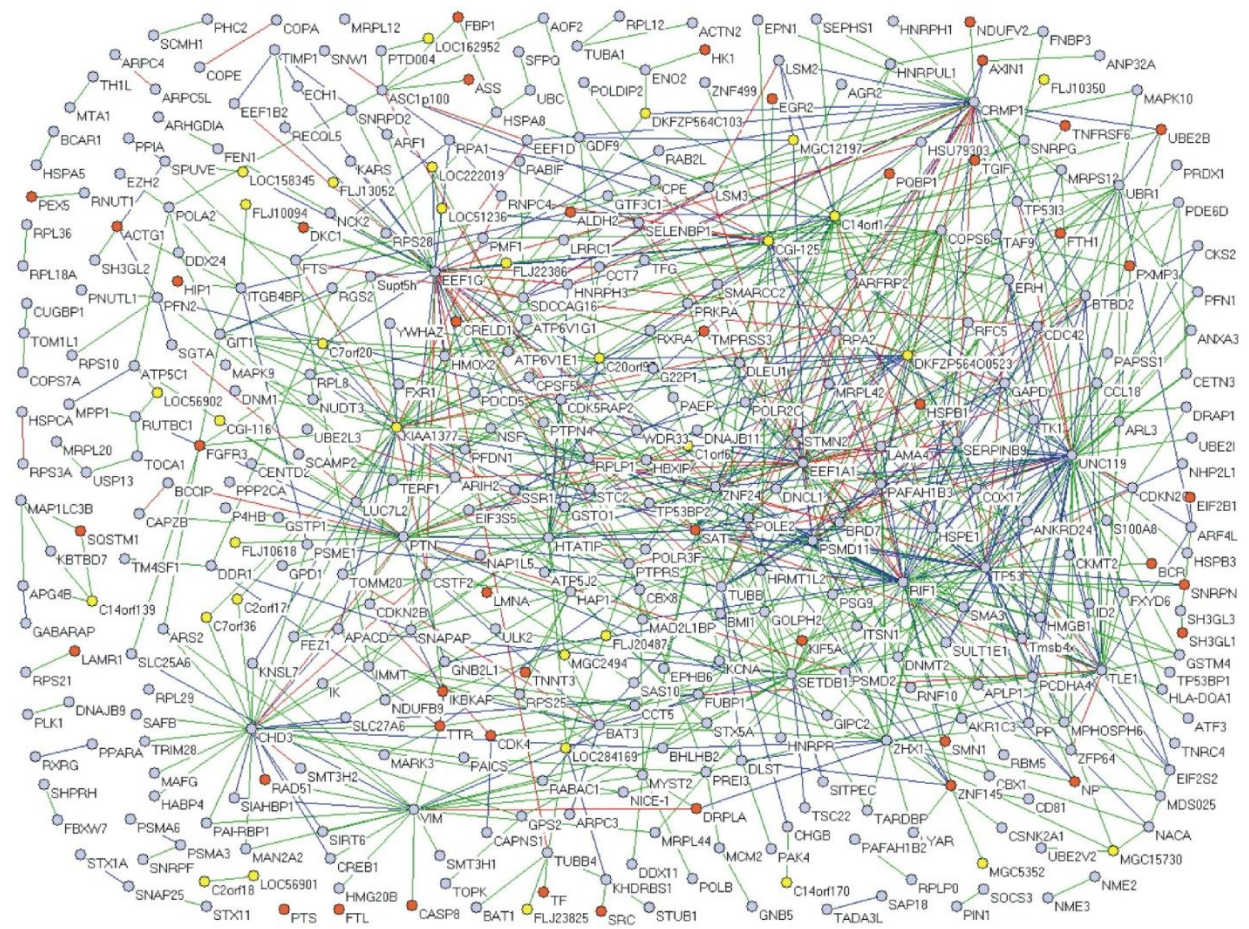


Figure 1 A graph of the human PPI network involving 401 proteins and 911 interactions [26]. The color of nodes represents proteins' annotation status: Orange: disease proteins (according to OMIM morbidmap, NCBI); light blue: proteins with GO annotation; yellow: proteins without GO and disease annotation. The color of links between proteins represents the confidence score of interactions: Green: 3 quality points; blue: 4 quality points; red: 5 quality points; purple: 6 quality points (the high the score is, the more reliable the interaction is). All links here were considered high-confidence interactions (see Ref 24 for more details). Fig. 1 was reproduced with permission from REF. 24.

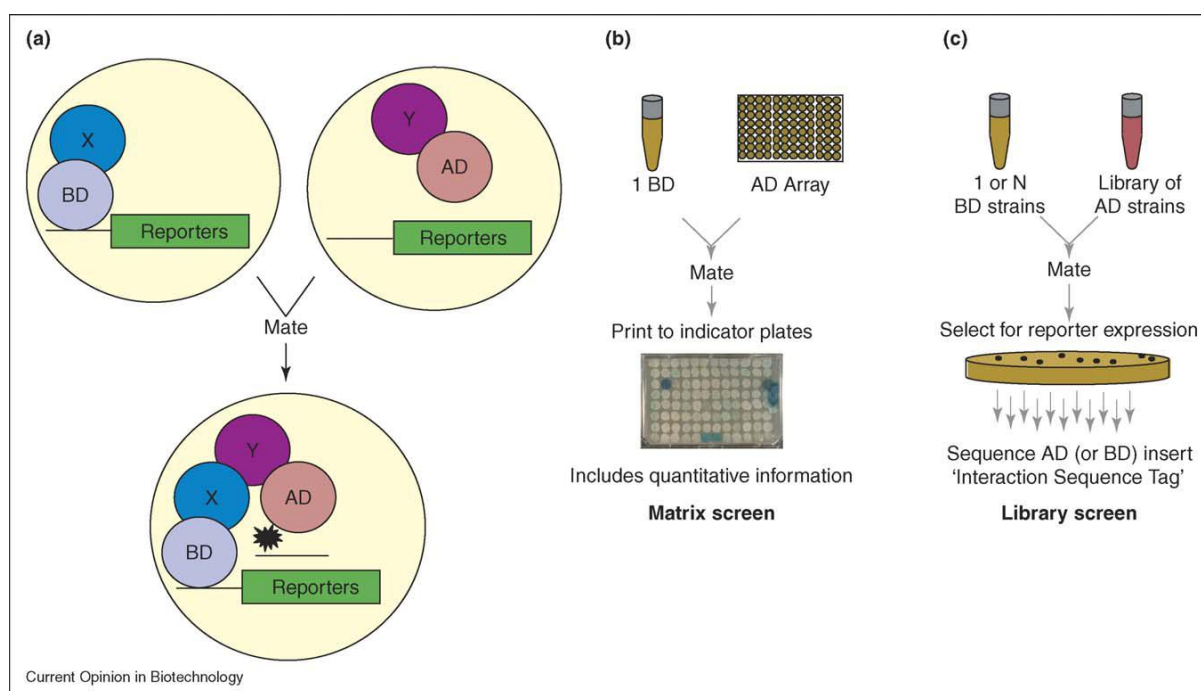


Figure 2 Yeast two-hybrid screening methods. (a) The basic idea of high-throughput Y2H screening. After yeast mating, the expression plasmids from two different haploid yeast strains are brought together. In the first strain, protein X is fused to a DNA-binding domain (BD) and will bind the upstream of the reporter gene. A transcription activation domain (AD) which is fused to protein Y in the second strain will activate the expression of the reporter gene after mating of the two strains if protein X and Y interact with each other. After the reporter gene is expressed, election can be made on media. (b) The basic idea of matrix approach for testing a single BD strain against each AD strain on the array. (c) The basic idea of library approach for testing single or multiple BD strains against libraries of AD strains. Fig.2 was reproduced with permission from REF 36.

CHAPTER 2

RESULTS

Our human PPI network is a scale-free network

The degree k (or connectivity) of a node shows how the number of links it has connected to other nodes in a network. In human PPI networks, recent publications have shown that the degree distribution approximates a power law, $P(k) \sim k^{-\gamma}$, where γ is the degree exponent. For our human PPI network (7,362 nodes with 20,019 links), the degree k also follows a power-law distribution, where the exponent γ is 1.93 (Fig. 3).

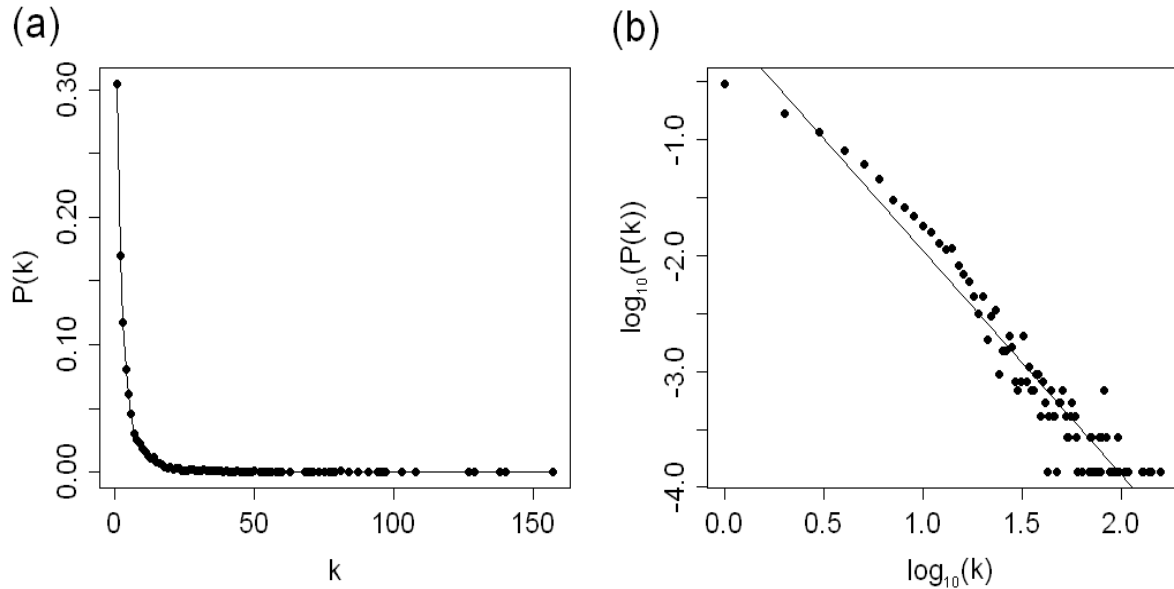


Figure 3 The degree distribution of our human PPI network. Here k represents the degree of node, $P(k)$ gives the probability that a node has k links. Data were plotted in linear-linear scale (a) and log-log scale (b). In (b), the solid black line was fitted with simple linear regression ($R^2 = 0.93$, $P < 2.2e-16$); it has a slope of -1.93 ($P < 2e-16$). Thus, the exponent γ is 1.93 in the power-law degree distribution for our human PPI network.

The Derivation of Algorithm I and II

Suppose that we have a random PPI network of size N . The total number of edges (interactions) is pre-determined, and edges randomly connect nodes (interacting neighbors) with an equal probability. This specifies a random network same as what was proposed in REF. 19. If we randomly choose two proteins, X and Y (X has n_x interacting neighbors and Y has n_y interacting neighbors), and we find that X and Y share m ($m \geq 0$) common interacting neighbors in the random PPI network. Here we denote the set of common neighbors as $A = \{Z_1, Z_2, \dots, Z_m\}$, the set of all proteins as $D = \{1, 2, \dots, N\}$, and the number of interacting partners for each protein in D as $\kappa = \{n_1, n_2, \dots, n_N\}$ [76]. The sample space Ω consists of all the graphs (outcomes) that X and Y randomly connect n_x and n_y proteins in the network, and we calculate the probability of the event that X and Y share m common neighbors.

Suppose that protein X randomly connects neighbors with an equal probability which only depends on the degree of X (the same for protein Y), then all possible graphs have the same probability to happen. The total number of possible graphs in which X and Y have m common partners is a product of three terms: (i) $\binom{N}{m} = \frac{N!}{m!(N-m)!}$ i.e., all possible ways to choose m proteins from N proteins without replacement; (ii) $\binom{N-m}{n_x-m}$ i.e., all possible ways to choose $N-m$ proteins from the remaining n_x-m proteins which only interact with protein X ; (iii) $\binom{N-n_x}{n_y-m}$ i.e., choose n_y-m proteins that interact only with protein Y from the remaining $N-n_x$ available proteins. Then we calculated the total number of possible graphs, $\binom{N}{n_x} \binom{N}{n_y}$, in which protein X has n_x neighbors and protein Y has n_y neighbors and the number of common neighbor is unfixed. By multiplying these three terms and then

dividing it by the total number of unrestricted ways for protein X to have n_x and protein Y to have n_y interacting neighbors— $\binom{N}{n_x}\binom{N}{n_y}$, we got the following formula (Algorithm I)

developed by Samanta and Liang [75]:

$$P_1(m | N, n_x, n_y) = \frac{\binom{N}{m}\binom{N-m}{n_x-m}\binom{N-n_x}{n_y-m}}{\binom{N}{n_x}\binom{N}{n_y}}.$$

We call this formula Algorithm I (P_1) in this dissertation. Our simulations showed that, for the random PPI network as defined above, the probability P_1 is accurate: the observed occurrence probabilities of m agreed with the probabilities (P_1) calculated from Algorithm I (Fig.4).

However, since human PPI network is a scale-free network [23, 24,], hub proteins will be connected at a higher probability than proteins with only a few links, which is different from the random network proposed in [19]. Thus, it's more realistic to use a randomized scale-free PPI network as the control which keeps the same total number of edges and the same degree distribution as our human PPI network. For such a random scale-free network (we use “random scale-free network” hereafter to refer to the randomized scale-free PPI network, unless otherwise specified), our simulations showed that P_1 becomes inaccurate: it deviates more sharply from the observed occurrence probabilities of m as m becomes larger (Fig. 5). In addition, we found that, as common neighbors, hub proteins (i.e. high-connectivity proteins) will appear at an extremely high frequency (Fig. 6, 7), comparing with the frequency of their occurrence in a random network; it also becomes relatively rare to observe low-connectivity proteins in A (the set of common neighbors) comparing with their dominant status in the random network. Since observing a hub protein in A is much easier

than observing a non-hub protein in A , the significance of P_1 needs to be down-weighted if there exists many hub proteins in A . An obvious example to justify this statement is that, if one protein interacts with most proteins in the network, it is not a surprise to find two proteins sharing it as a common neighbor. Therefore, since P_1 cannot tell whether or not common neighbors are hub proteins, we developed another algorithm to measure the connectivity of common neighbors so that we can reduce the influence of hub proteins on our assessment of protein functional associations: we assume that any protein in D , except X and Y , randomly and independently connects other proteins with an equal probability which only depends on its own degree, and all possible (connecting) graphs in the network have the same probability to happen. Thus we were able to use the degree κ of D (excluding the degree of X and Y) to compute the probability that only $A = \{Z_1, Z_2, \dots, Z_m\}$ connects to X and Y , and we arrived at the following formula:

$$\begin{aligned}
& P_2(X \text{ and } Y \text{ share } A \mid \kappa, N) \\
&= \left[\prod_{i=1}^m P(Z_i \text{ connect both } X \text{ and } Y \mid \kappa, N) \right] P(\text{no other protein connect both } X \text{ and } Y \mid \kappa, N) \\
&= \prod_{i \in A} \frac{\binom{N-1}{n_i-1} \binom{N-2}{n_i-2}}{\binom{N}{n_i} \binom{N-1}{n_i-1}} \prod_{i \notin A, i \in \Omega} \left(1 - \frac{\binom{N-1}{n_i-1} \binom{N-2}{n_i-2}}{\binom{N}{n_i} \binom{N-1}{n_i-1}} \right) \\
&= \prod_{i \in A} \frac{n_i(n_i-1)}{N(N-1)} \prod_{i \notin A, i \in \Omega} \left(1 - \frac{n_i(n_i-1)}{N(N-1)} \right)
\end{aligned}$$

We call this formula Algorithm II. In Appendix A, we showed that the second product is bounded within a very narrow interval and can be considered as a constant, and we used the

approximation $P_2 = \prod_{i \in A} \frac{n_i(n_i-1)}{N(N-1)}$ for convenience in our dissertation.

As shown above, each protein pair with at least one common neighbor can be assigned with both P_1 and P_2 . In previous work on yeast PPI network, Samanta and Liang [75] only used P_1 to rank the functional association of protein pairs. In our dissertation, we added P_2 as a complementary algorithm to improve the quality of biological inferences. We later showed that, by reducing the influence of hub proteins in the network, the use of both P_1 and P_2 allowed us to identify a more reliable functional relationship than that identified by P_1 alone. [76]

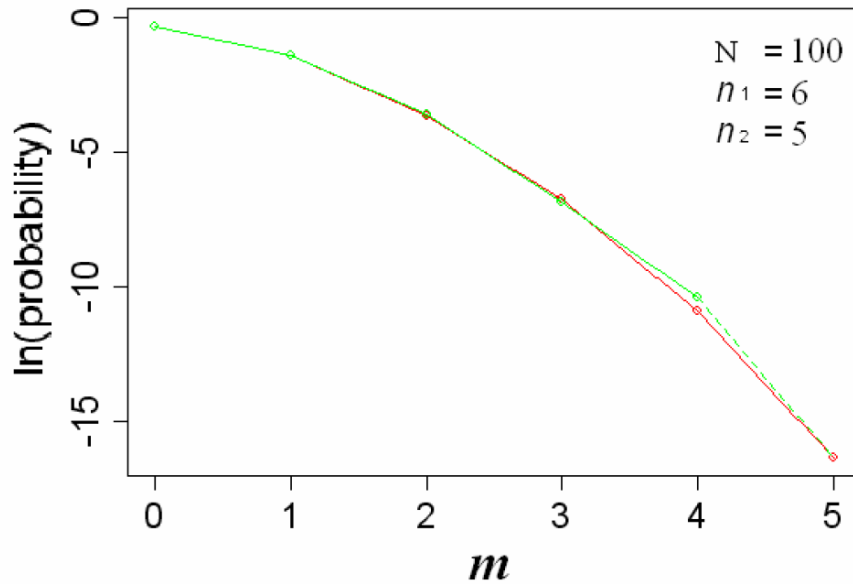


Figure 4 Probabilities observed from random networks vs. probabilities calculated from Algorithm I. Green points (line) represent the observed probabilities for different m in random networks, and green dashed line represents the expected observation from the random networks ($m=5$ cannot be observed based on the times of our simulation). Red points (line) represent the theoretical probabilities calculated from algorithm I. We performed simulations 1000 times for a random network with 100 nodes and 300 interactions (n_1 and n_2 are the degrees of protein X and Y , N is the size of the network). Fig. 4 is an example taken from our simulation.

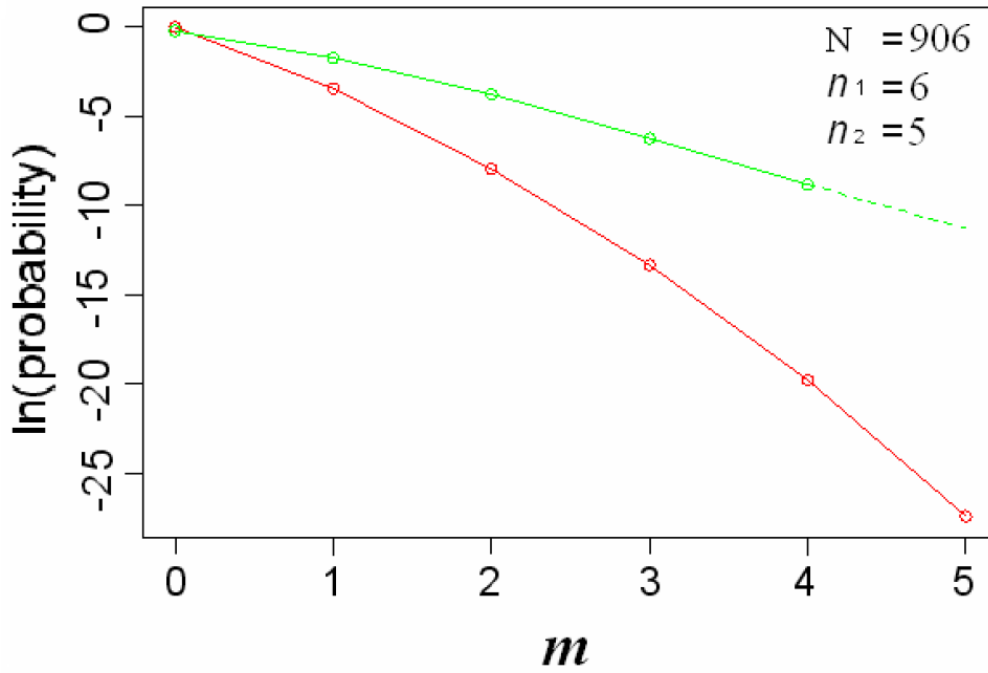


Figure 5 Probabilities observed from random scale-free networks vs. probabilities calculated from Algorithm I. Green points (line) represent the observed probabilities for different m in random scale-free networks, and green dashed line represents the expected observation from random scale-free networks ($m=5$ cannot be observed based on the times of our simulation). Red points (line) represent the theoretical probabilities calculated from algorithm I. Here the random scale-free networks were simulated 30 times based on the mouse PPI network (906 nodes and 1054 interactions) for computational convenience (i.e., the random scale-free networks used here have the same total number of edges and degree distribution as the mouse PPI network) [77]. In Fig. 5, n_1 and n_2 are the degrees of protein X and Y , N is the size of the network in our simulation.

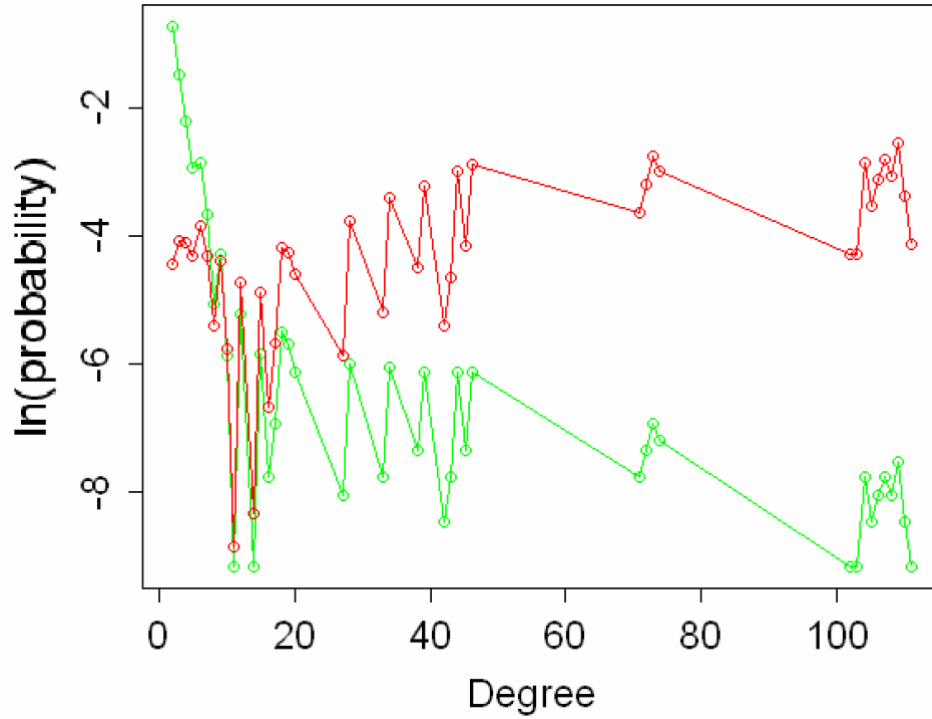


Figure 6 Hub proteins always appear as common neighbors. Green points (line) represent the observed probability of seeing proteins with the same degree in random scale-free network, and red points (line) represent the observed probability of seeing proteins with the same degree in the common neighbor set A in the random scale-free network.

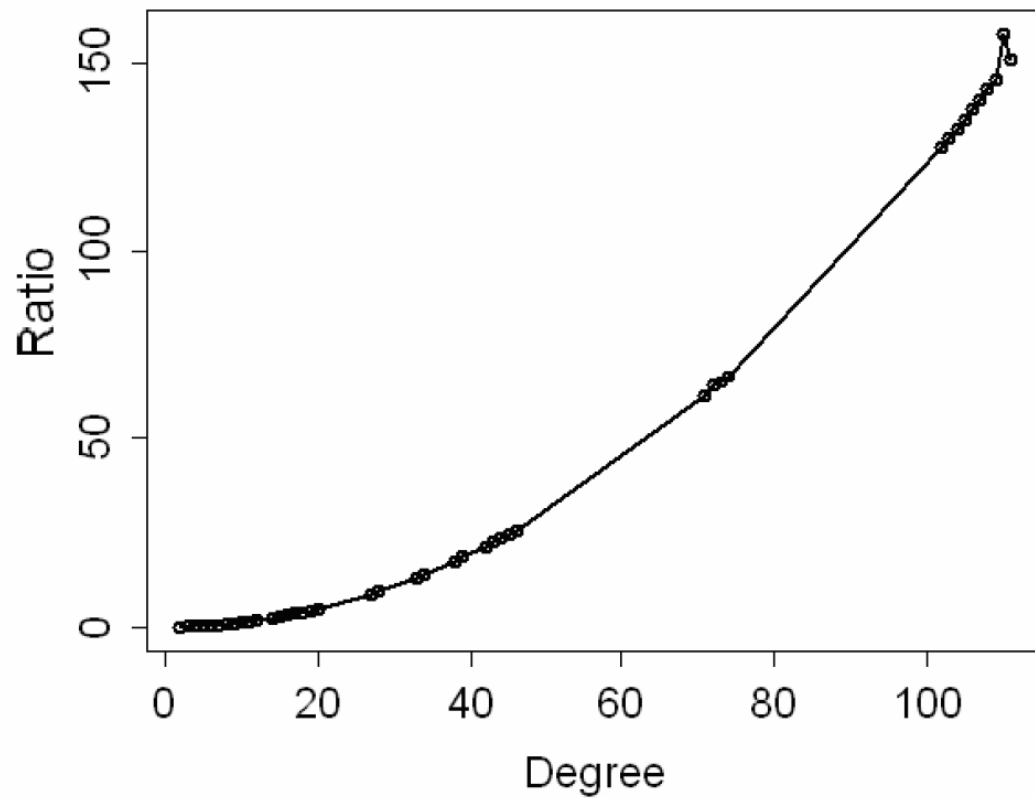


Figure 7 Hub proteins are more inclined to appear as common neighbors. The black points stand for the ratios of each two corresponding observed probabilities from Fig. 6 (red/green). Ratio becomes larger as degree increases, which means hub proteins are more inclined to appear as common neighbors.

Algorithm I and II Distinguishes the Human PPI Network from Random Networks and Random Scale-free Networks

We first applied the two algorithms on our human PPI data and computed the probabilities (P_1 and P_2) for 311,023 protein pairs that had at least one common neighbor. We then applied the two algorithms on the aforementioned random networks and random scale-free networks and calculated P_1 and P_2 for around 500,000 and 150,000 proteins pairs, respectively. To assess the statistical significance of the topological connections in the human PPI network, we first compared the number of common neighbors between the three types of networks, then we compared the distribution of probabilities calculated from them and plotted the distributions (Fig. 8). Comparing to the random network, the random scale-free network have a much more similar behavior as the human PPI network. Since biological networks, including PPI networks, are scale-free networks, it is more reasonable to use random scale-free networks as the benchmark for comparisons. As expected, the human PPI network has many highly improbable topological connections which would almost impossibly happen in random and random scale-free networks. [76]

*In this dissertation, all probabilities (P_1 and P_2) have been natural logarithm (base e) transformed.

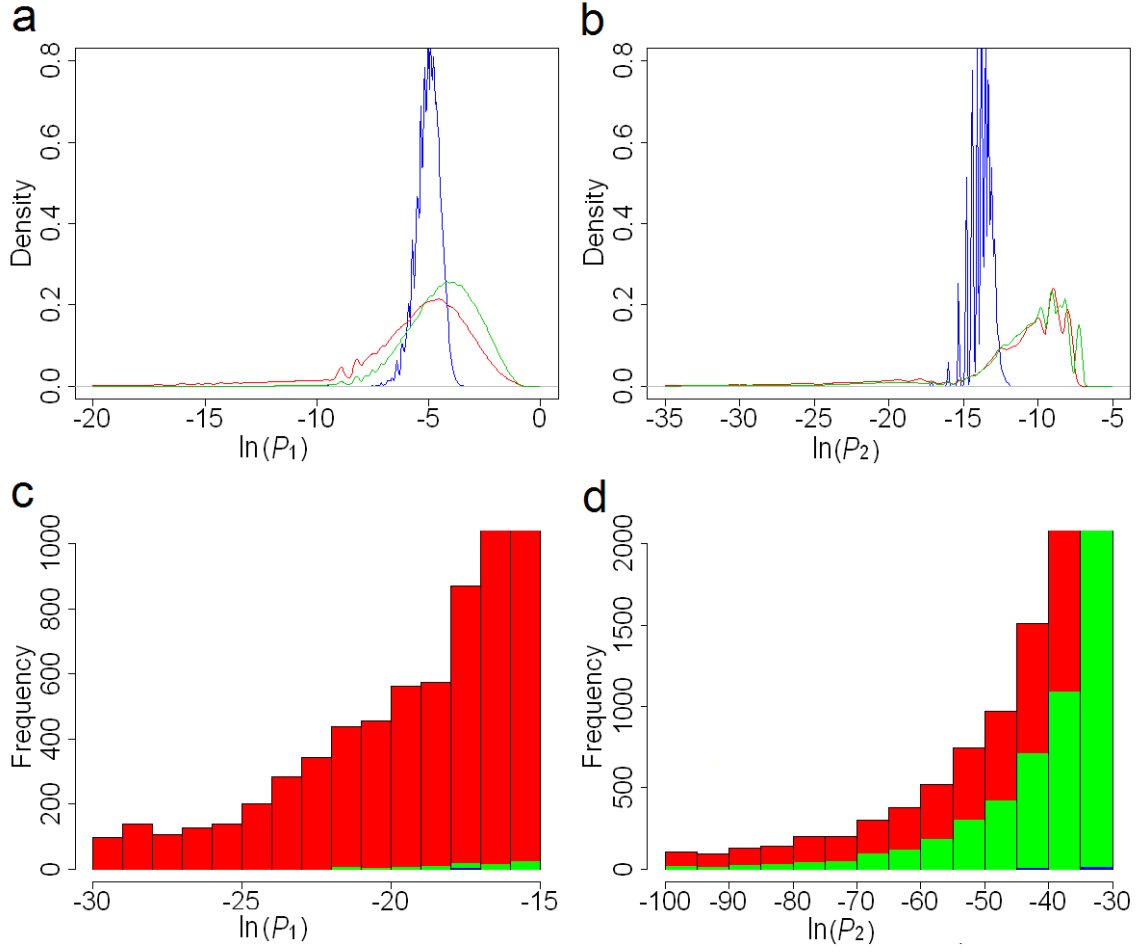


Figure 8 Density distributions and histograms of P_1 and P_2 derived from Algorithm I and II. In all four plots, red, green and blue stands for the results (density distribution and histogram) calculated from the human PPI network, the random scale-free network and the random network, respectively. (a) Density distributions of P_1 . (b) Density distributions of P_2 . (c) Histograms of P_1 . (d) Histograms of P_2 . Fig.8 was reproduced from REF. 76.

Algorithm I Identifies Functionally Associated Proteins in Human PPI Networks

P_1 is able to assess the degree of nonrandomness and suggest protein functional associations in the yeast PPI network [75]. Since human and yeast PPI networks share plenty of common characteristics, we expected P_1 to do the same thing in our human PPI network: ranking protein pairs in terms of their functional association. Therefore, we hypothesized that a higher ranking (the smallest P_1 is ranked highest) corresponds to a closer functional relationship. With the Gene Ontology (GO) annotations [79] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations [80] as benchmarks, we used annotation overlap rate (see Methods and Materials) to test our hypothesis: if protein pairs of higher ranking have better annotation overlap rate, our hypothesis is validated; otherwise rejected. From Fig. 9 we can see that, overall, annotation overlap rate becomes higher as ranking goes higher, which validates our hypothesis. Here we also noted in Fig. 9 that, in the top 5,000 protein pairs, each 1,000 pairs always had a higher overlap rate than those beyond the top 5,000 pairs, and that the region of high overlap will give us a high level of confidence in presenting reliable predictions. Thus, we chose the 5,000th value (-17.11) of P_1 as the cutoff from Algorithm I, which also perfectly matches the Bonferroni correction $\ln(\frac{2}{N(N-1)})$ in which $N=7,362$ is the size of the whole protein network. In order to further assess the performance of our method, we defined a false discovery rate (FDR) [81] (see methods and materials). For the top 5,000 protein pairs selected by Algorithm I, the FDR is 0.40.

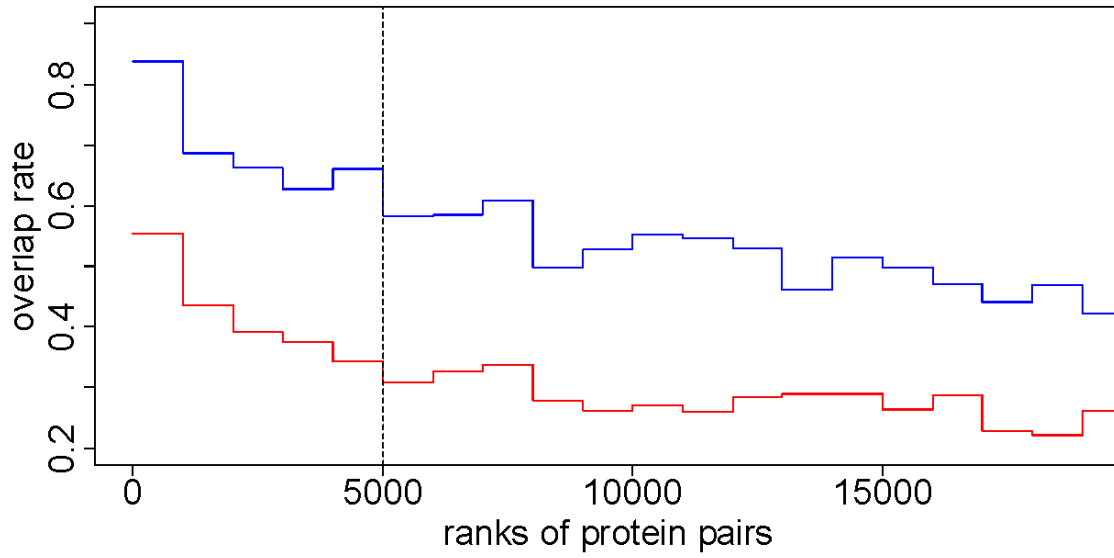


Figure 9 Annotation overlap rate with GO and KEGG as the benchmarks. Protein pairs are ranked by P_1 . The top-ranked 20,000 pairs are divided equally into 20 bins. In each bin (1,000 protein pairs), we calculated the GO overlap rate and the KEGG overlap rate. The red curve stands for the GO overlap rates and the blue one stands for the KEGG overlap rates. The dashed line is the cutoff (-17.11) at the 5,000th protein pair. The correlation coefficient between the two groups of overlap rates is 0.928 ($P < 0.0001$). Fig. 9 was reproduced from REF. 76.

Algorithm II Improves the Evaluation of Functional Associations

As mentioned in previous paragraphs, P_2 is designed to reduce the influence of hub proteins which commonly exist in scale-free networks. Here we hypothesize that, within the top 5,000 protein pairs selected by P_1 , P_2 can identify and remove protein pairs whose small P_1 is mainly caused by hub proteins in their common neighbor set, hence identify better functional associations among proteins. After manually inspecting the top 5,000 protein pairs, we found a considerable amount of protein pairs whose good P_1 s are caused by hub protein, and we listed some as examples in Table 1. With GO and KEGG as the benchmarks, we validated our hypothesis on P_2 by the following assertions: “(i) the protein pairs with a good P_2 (Group I) always have a lower FDR (here a lower FDR means a closer functional relationship) than those without a good P_2 (Group II; Fig. 10a); and (ii) the protein pairs with a good P_2 (Group I) always have a lower FDR than the same number of top protein pairs ranked by P_1 only (Group III; Fig. 10b). We also noted that because P_1 and P_2 have a very low linear correlation (Pearson’s correlation coefficient = -0.033, $P < 10^{-16}$; also see Fig. 11a) and rankings of functional association by P_1 and P_2 are significantly different ($P < 10^{-16}$), an additional cutoff from Algorithm II makes difference from simply tightening the cutoff from Algorithm I. As the cutoff for P_2 changes, the difference in FDR between Groups I and II varies; the difference maximizes when the cutoff goes to -30.03, which is the value we used for the second cutoff from Algorithm II (Fig. 10a). ”^ψ Therefore, 4,233 protein pairs (see Table S1 in Appendix B) were considered to have significant functional associations ($P < 0.001$) in terms of the cutoffs from Algorithm I (-17.11) and Algorithm II (-30.03), and we call them significant protein pairs in this dissertation for convenience. Furthermore, FDR

of the 4,233 significant pairs is only 0.35, compared with 0.39 of the top 4,233 pairs ranked by P_1 only (Fig. 10*b*). [76]

^Ψ In this dissertation, all quoted paragraphs were cited from REF. 76.

Due to the incompleteness of functional annotations for human genes and proteins, the annotation (GO and KEGG) based FDRs (0.35) are probably overestimated. Although it was very hard to obtain “true” FDR from annotations, we still estimated its lower bound by comparing the behavior of the human PPI network with the random scale-free networks: with the same cutoff (-17.11 for P_1 and -30.03 for P_2), we generated false significant protein pairs in random scale-free networks for multiple times; by dividing the averaged number (86) of false significant pairs by our predicted significant pairs (4,233), we determined the lower bound of FDR should be approximately 2% (Fig. 11). [76]

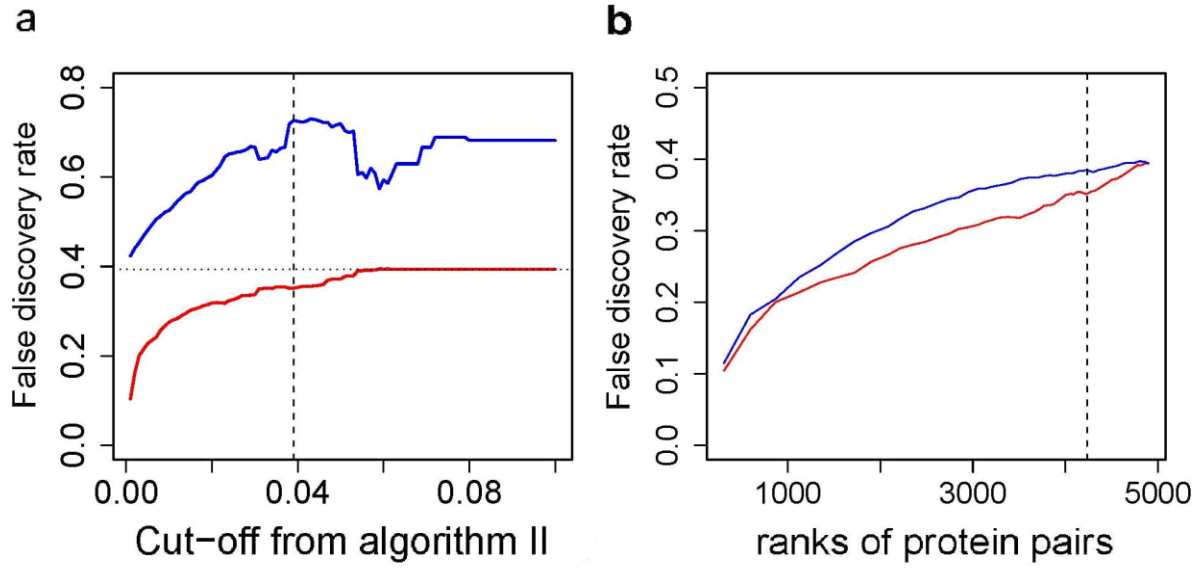


Figure 10 Algorithm II decreased the false discovery rate (FDR) of our predictions.

“(a) For the top 5,000 protein pairs ranked by P_1 , each cutoff value from P_2 (on the x axis is the quantile of P_2 we used as the cutoffs) divided them into two groups: Group I (red line), whose P_2 was better than the cutoff, and Group II (blue line), whose P_2 was worse than the cutoff. In this plot, the maximal difference between the two groups is at 0.039 (vertical dashed line), which corresponds to the cutoff of -30.03 from Algorithm II. The horizontal dotted line stands for the FDR (0.40) of the top 5,000 protein pairs ranked by P_1 . (b) The blue line (Group III) shows the FDR of protein pairs ranked by P_1 only (x axis stands for the amount of selected top protein pairs), and the red line (Group I) shows the FDR of the significant protein pairs selected by P_1 and P_2 together.”^ψ Fig. 10 was reproduced from REF. 76.

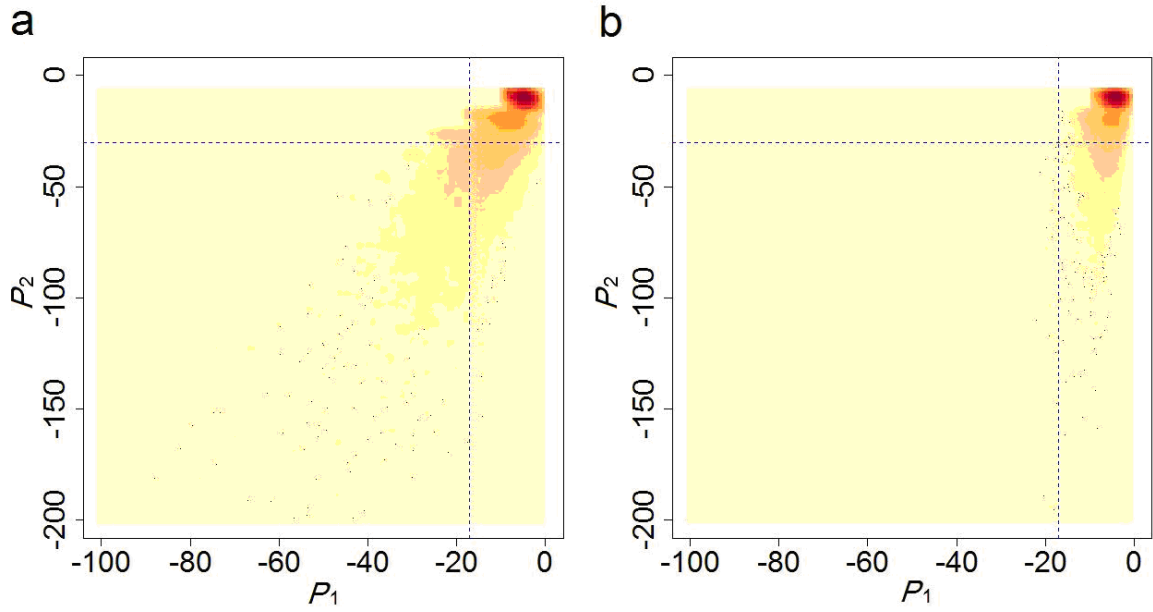


Figure 11 Smoothed color density representation of the scatter plot of P_1 and P_2 . (a) The human PPI network. (b) The random scale-free network. The vertical and horizontal lines represent the cutoffs for P_1 (-17.11) and P_2 (-30.03), respectively. In a random scale-free network, the expected number of significant protein pairs is 86 (lower left area in *b*), compared with the number of 4,233 in our human PPI network (lower left area in *a*). Fig. 11 was reproduced from REF. 76.

Significant Protein Pairs Are Highly Informative in Functional Inference

Within the top 4,233 significant protein pairs, we observed strong functional relationships: by manual inspection, we found that at least 96 of the top 100 protein pairs (unannotated proteins were excluded) are functionally associated (see Table 2 for the top 10 protein pairs); with GO and KEGG as benchmarks, the FDR of the top 4,233 protein pairs is 0.35, compared with 0.83 of the top 4,233 pairs from the random scale-free network [cutoffs^{*}: -8.90 for P_1 and -11.33 for P_2] and 0.92 of the top 4,233 pairs from the random network [cutoffs^{*}: -6.42 for P_1 and -13.10 for P_2]. [76]

^{*} the cutoff for P_1 is the value of the 5,000th protein pair's P_1 ; the cutoff for P_2 is the value of 4,233th protein pair's P_2 when the 5,000 protein pairs are ranked by P_2 .

Most of the direct interactions in our PPI data are from traditional experiments, and the rest are high-throughput data which were also strictly selected [23, 24, 77] (for details of our dataset, see methods and materials). With GO and KEGG as the benchmarks, FDR for the 23,782 direct interactions is 0.57. This is significantly higher than the FDR of the top 4,233 protein pairs ($P < 10^{-16}$ by two-sample proportion test). From here we can conclude that, our method probably identifies more reliable functional associations than the original PPI data does. Furthermore, among our 4,233 significant protein pairs, only 21.6% share direct interaction, which makes we believe our method reveals a substantial amount of functional information that is not present in experimentally acquired PPI data. [76]

“Among the 1,754 proteins in the top 4,233 protein pairs, 1,220 have qualified GO terms (i.e., GO terms at the highest level without direct or indirect GO “offspring” terms in each ontology), and 834 have KEGG pathway annotations. If a protein has at least one

annotated significant partner (i.e., two proteins are significant partners to each other if they are a significant protein pair), a list of annotation(s) from its partner(s) can be sorted by frequency and annotations occurring at the highest frequency are assigned to this protein (frequency must be at least twice for KEGG and four times for GO; otherwise discarded.”^Ψ For more details, see Fig. 12 and Appendix A). “For an annotated protein (based on GO and KEGG annotations), if an assigned annotation occurs among its known functions, we consider this to be a correct prediction. By this method, we found that 79% (for KEGG) and 70% (for GO) of assigned annotations were correct predictions. (Randomly picking 4233 pairs from 1729 proteins will only yield a 7% correct prediction rate for KEGG and 12% for GO on average from 100 trials.) In the same way, we predicted 466 KEGG pathways for 274 proteins and 123 GO terms for 114 proteins. We estimated that the FDRs of our predictions are much less than 21% (for KEGG) and 30% (for GO) because of the percentage of correct predictions for annotated proteins and the incompleteness of GO and KEGG annotations.”^Ψ In addition, for each prediction, we assigned a *P* value to further assess the significance of the assigned annotation(s) under Fisher’s exact test (see methods and materials). For brevity, we only selected 40 predicted annotations and listed them in Table 3 as an example (20 KEGG annotations and 20 GO annotations). For the list of complete predictions, please refer to Table S2 in Appendix B. [76].

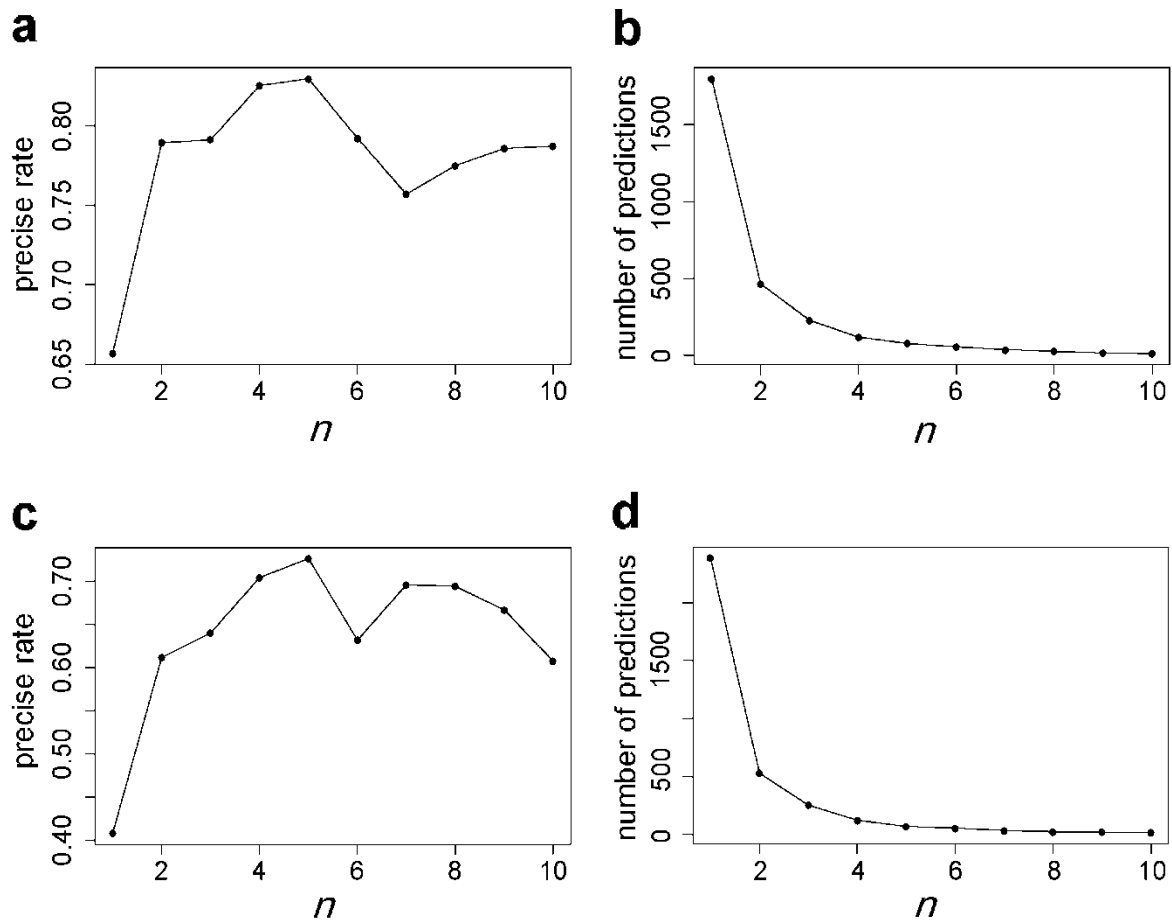


Figure 12 Estimation of prediction precise rates and the number of predictions we can make given different n . “(n is the minimal frequency of annotation occurrence required for functional prediction). (a) Estimated precise rate of predicted KEGG pathway given n . (b) The number of predictions for KEGG pathway we can make given n . (c) Estimated precise rate of predicted GO terms given n . (d) The number of predictions for GO terms we can make given n .”^ψ Fig. 12 was reproduced from REF. 76.

External validation of predicted GO annotations

Our predictions of GO annotations were based on R package: GO, 08-Aug-2006. Among our 123 GO term predictions, 17 annotations have already been validated and approved by an independent GO annotation coordinator - Emily Dimmer (see Table 4 for details). These approved GO annotations will be probably put into GO SeqdbLite with an evidence code of “IC”, where IEA associations are removed and annotations are considered to have a relatively low error rate of 28%-30% [82].

Clustering Based on Significant Pairs can Effectively Identify Signaling Pathway Members

“Because clustering can significantly improve the quality of functional inference [75], we built a cluster consisting of 1,729 proteins (excluding 25 non-human proteins) based on the P_1 of 4,233 significant protein pairs. We constructed the empirical cumulative distribution from these P_1 values; thus, each significant protein pair had a score between 0 and 1 according to its ranking order in the distribution of P_1 . Then we built a 1729×1729 dissimilarity matrix in which each matrix element was assigned either a score (if applicable) or a “10” for pairs with no significant P_1 . The purpose of using such a large value was to minimize background noise. Then the dissimilarity matrix was subjected to agglomerative hierarchical clustering with an unweighted pair-group average. The whole cluster is given in Fig. S1 in Appendix C.”^ψ [76]

We made detailed analysis on functional modules with significant P values. “In the cluster of 1,729 proteins, most of the functionally related proteins were correctly clustered into their corresponding functional modules, in which they are characterized by similar functions or the same pathway (Fig. 13). The largest subcluster derives directly from the root of the whole cluster and consists of 959 proteins; the second-largest subcluster has only 51 members (Appendix C).”^ψ We used Ingenuity Pathway Analysis (IPA 5.0) to analyze the largest subcluster and found that a large number of signaling pathways were highly enriched there: 57 signaling pathways had extreme P values less than 10^{-10} after P value adjustment [81] (Fig. 14). “We cut the 959-member subcluster with different cutoff values and analyzed the corresponding subclusters by using both manual inspection and Ingenuity Pathway

Analysis (IPA). We conducted a detailed analysis for one prominent subcluster (the subcluster related to the TGF- β signaling pathway) as a reference.”^Ψ [76]

“The TGF- β signaling pathway-related subcluster (Fig. 15a) has a total of 45 protein members, 35 of which are known to participate in the TGF- β signaling pathway, according to the Ingenuity database. The probability of observing this by chance is $< 10^{-54}$, according to the calculation from Ingenuity software (right-tailed Fisher’s exact test). With respect to this extreme P value, we reasoned that probably all the cluster members cooperate to mediate signal transduction. To investigate the role of the other 10 proteins in the TGF- β signaling pathway, we generated a functional relationship network using Osprey software (<http://biodata.mshri.on.ca/osprey>) [83] to explicitly elucidate the relationships between the 45 proteins (Fig. 15b): the 10 proteins not related to TGF- β according to the Ingenuity database are located inside a circle, whereas the other 35 TGF- β member proteins lie on the circle; common neighbors which do not belong to the 45-member subcluster stay outside the circle.”^Ψ [76]

“The cluster and the association network (Fig. 15a and 15b) intuitively suggest possible roles that the inner proteins play in the TGF- β signaling pathway, which have not yet been incorporated into the Ingenuity pathway. Take Fig. 15b for instance: SKI functions as both the significant partner and the direct interacting neighbor of SMAD2 and SMAD3, and the three proteins’ common neighbors (five violet nodes) all share the function of transcriptional regulation. From this we infer that SKI may regulate the TGF- β signaling pathway on a transcription level, which is in accordance with findings in the literature (but has not been incorporated into the Ingenuity database) that SKI regulates downstream DNA

transcription by forming a protein complex with SMAD2 and SMAD3 [84-85]. With respect to IGSF1's significant partners, direct-interaction partners, and the previous work identifying IGSF1 as a potential receptor that could affect cellular response through its cytoplasmic region [86], we suspect that IGSF1 could function as a coreceptor for inhibin and/or activin. SOSTDC1 and NOG may regulate TGF- β by interacting with BMP receptors, which is in accordance with the findings that both of them function as BMP antagonists [87, 88]. In addition to positive regulatory functions [89], DAB2 may serve as an antagonist of STRAP, which has a negative regulation on TGF- β -mediated transcriptional activation [90, 91]. FMOD, CTGF, and SLITL2 may be involved in regulating receptor binding of TGF- β s, in accordance with published findings [92-94], and they may interact with each other.”^ψ [76] Therefore, based on our analysis of TGF- β signaling pathway-related subcluster, we suggested possible roles for eight proteins in TGF- β signaling pathway (Table 5). Some of the suggested functions have supporting literatures, while others (shown in bold in Table 5) have none. Those functions without any literature support are worth further experimental investigation by biologists.

“To facilitate analysis of this type, we proposed eight signaling pathways that are worthy of further investigations (Fig. 16), under the assumption that proteins within the same signaling pathway tend to stay together in the same subcluster. This is shown for the largest 959-member subcluster (Fig. 16a; cluster members are indexed from 1 to 959). From IPA-based classification of the proteins into each of the eight pathways, we calculated a density distribution for all eight signaling pathways along the cluster (Fig. 16b–e). Each pathway is expected to have a distinct distribution (its own peaks). The peaks in Fig. 16b–e

map to some areas (i.e., subclusters) that are probably highly related to their corresponding pathways.”^ψ Functionally intercrossed pathways, like death-receptor/NF-κB signaling, may probably have peaks close to each other. Moreover, we calculated the density distribution of the same eight signaling pathways in the largest cluster from random network and found no obvious peaks (Fig. 17). We believe these distribution patterns from real PPI networks are useful in identifying pathway-specific regions in the cluster. “We selected another 4 subclusters that are presumably involved in six signaling pathways (excluding TGF-β) with respect to pathway member distributions, and listed the potential pathway members in Fig. 18. We expect that the clusters and distributions will help biologists to find their subcluster of interest and discover new pathway members.”^ψ [76]

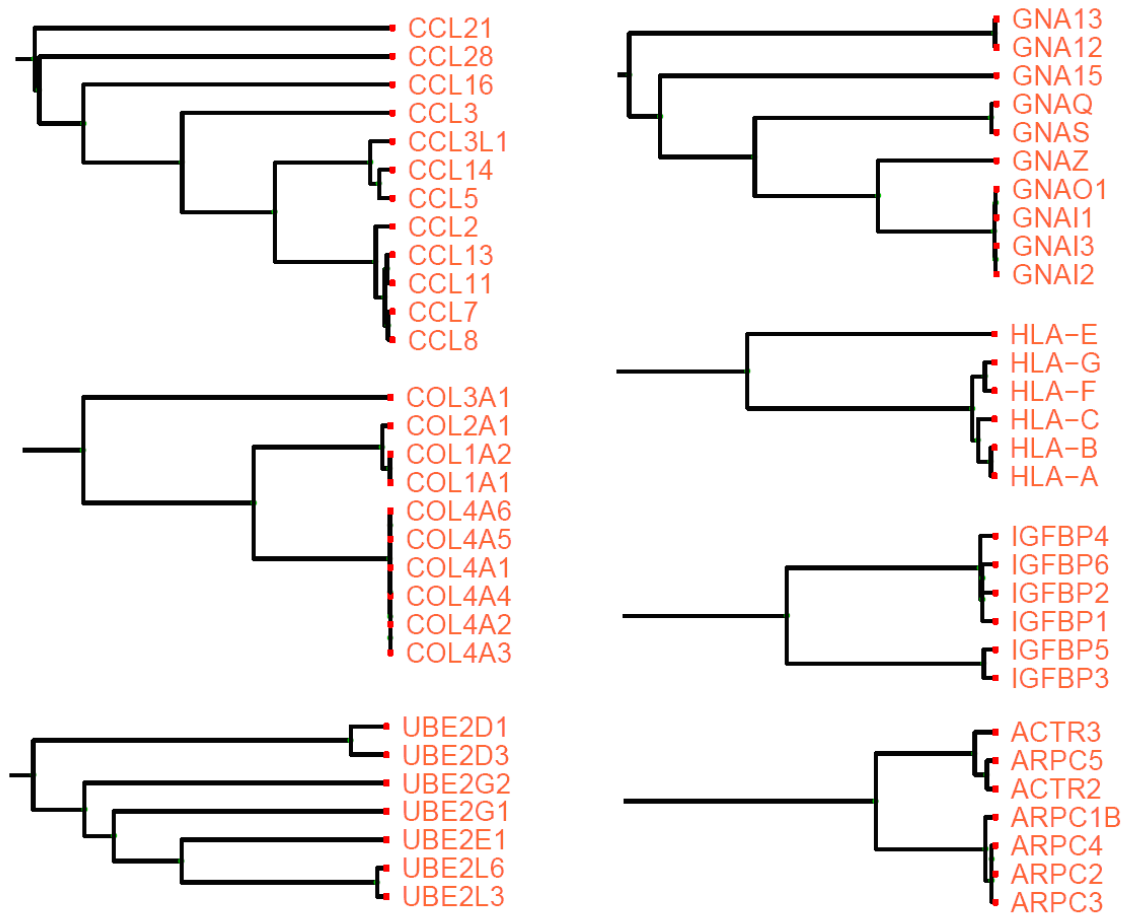


Figure 13 7 subclusters derived from the 4,233 significant protein pairs. It is obvious that they all belong to their own functional module where members perform similar or related biological functions. Fig. 13 was reproduced from REF. 76.

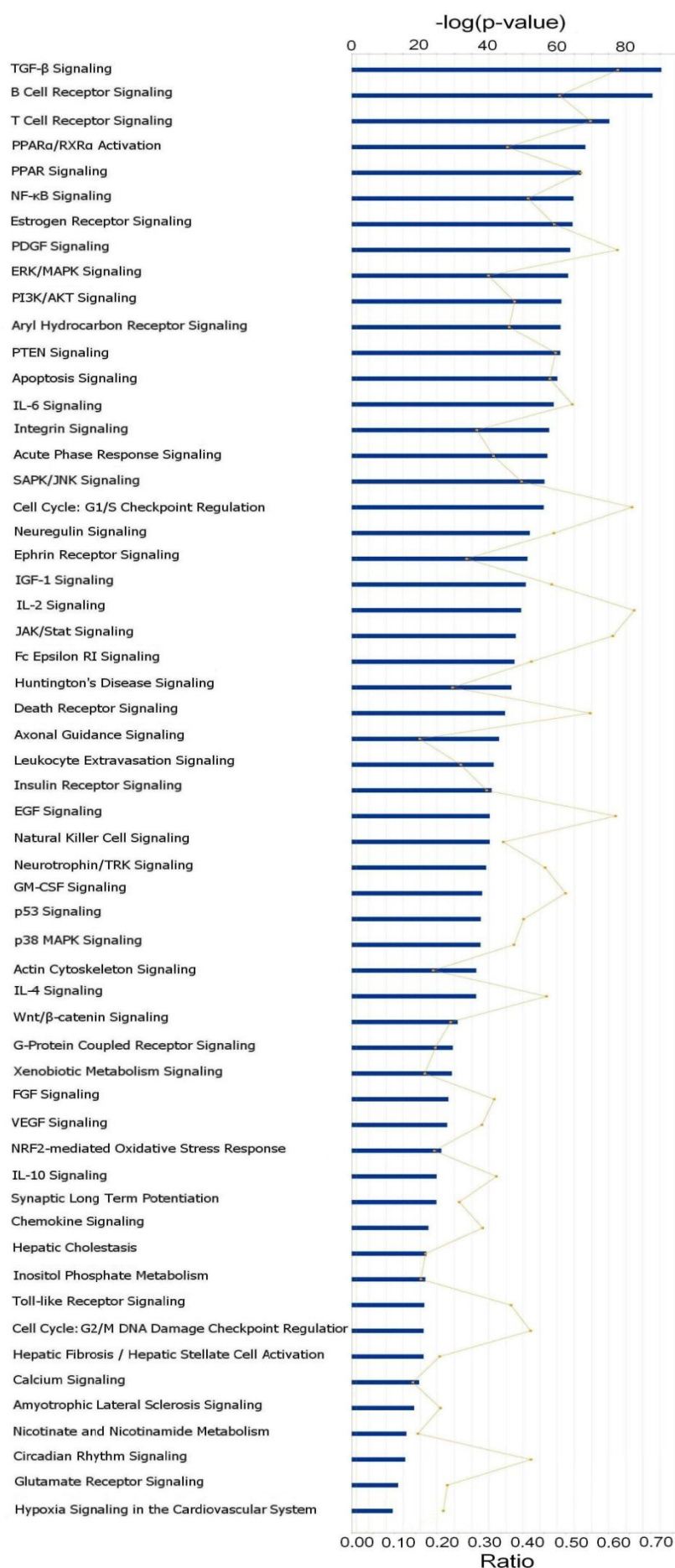
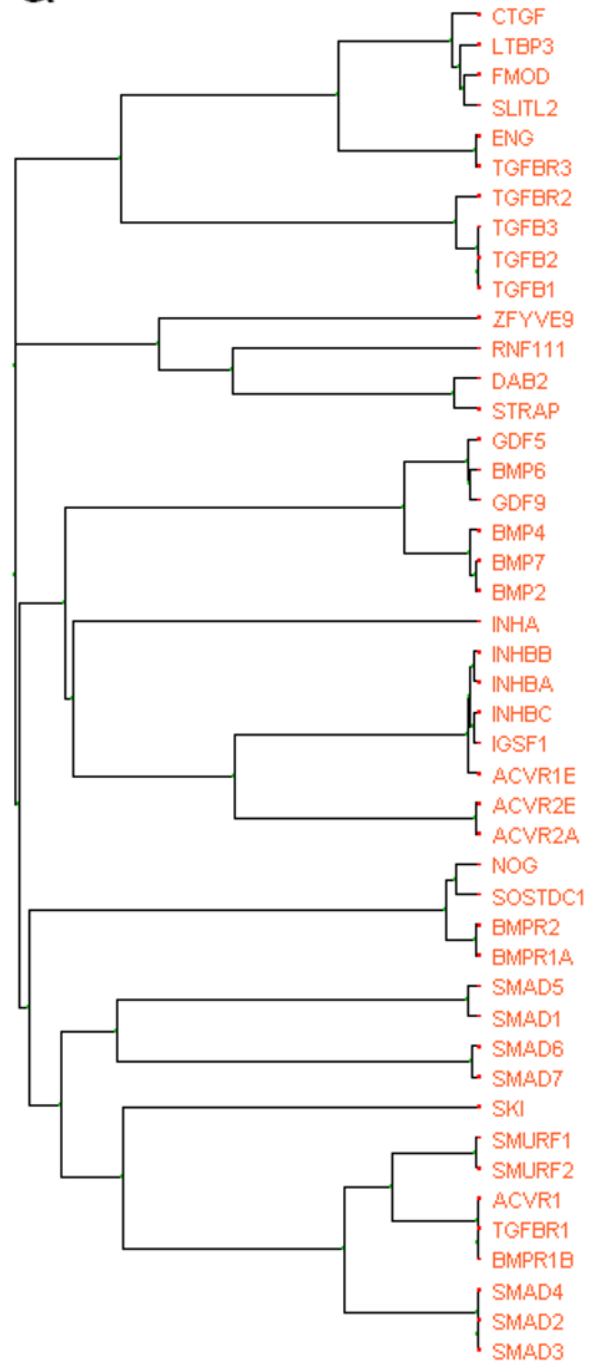


Figure 14 57 highly enriched signaling pathways with P values less than 10^{-10} in the largest subcluster. The height of blue bars indicates unadjusted P value for each signaling pathway. The yellow dots mean the ratio of pathway members in the subcluster over all pathway members in IPA 5.0.

a



b

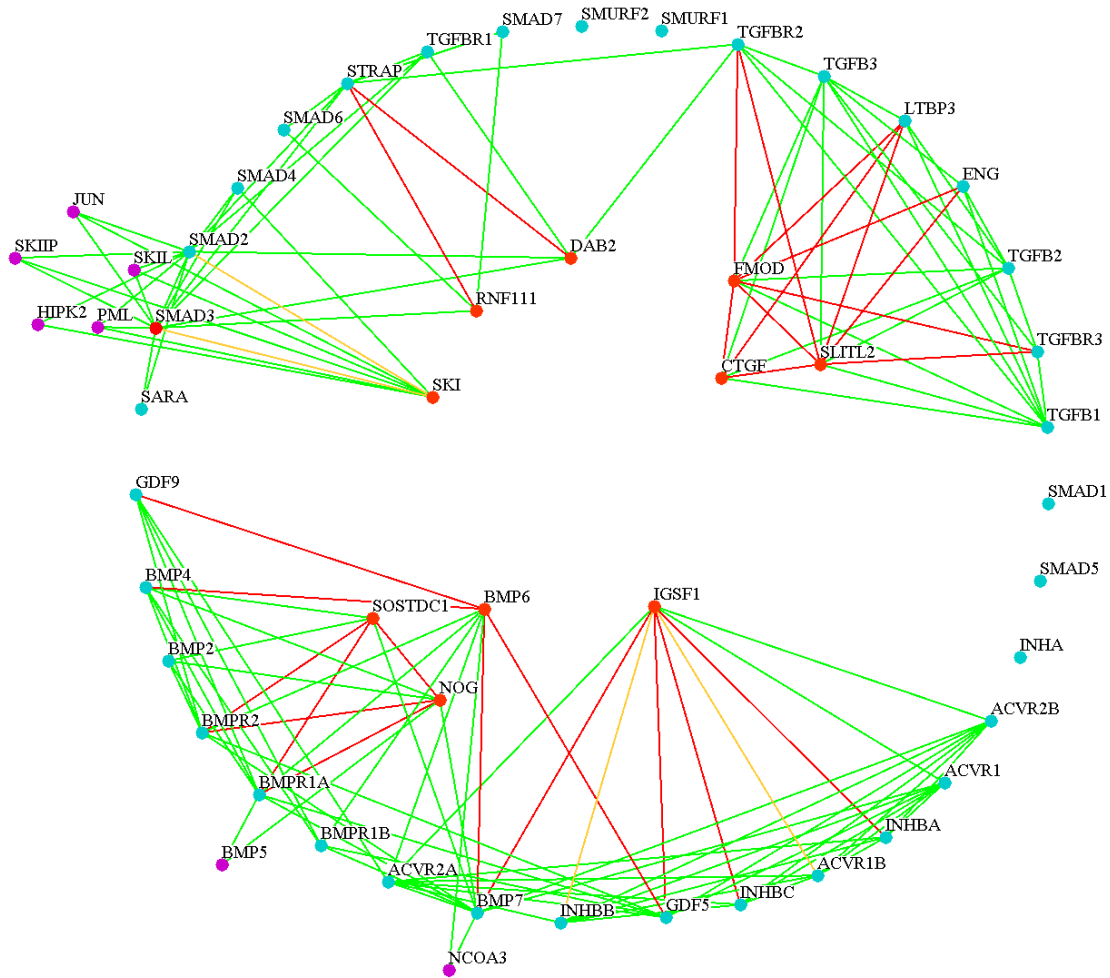


Figure 15 TGF- β signaling pathway–related subcluster. “(a) One subcluster identified by our method consists of proteins presumably involved in the TGF- β signaling pathway. (b) Detailed interpretation of the relationships between each protein from the subcluster. On the basis of the Ingenuity Pathway Analysis 5.0, the 35 blue-green proteins on the circle participate in the TGF- β signaling pathway, and the 10 red proteins inside the circle are unrelated. The violet proteins outside the circle are common neighbors that do not belong to the subcluster in panel *a*. Red lines represent significant protein pairs, green lines represent direct protein–protein interactions, and yellow lines represent both.”^ψ Fig. 15 was reproduced from REF. 76.

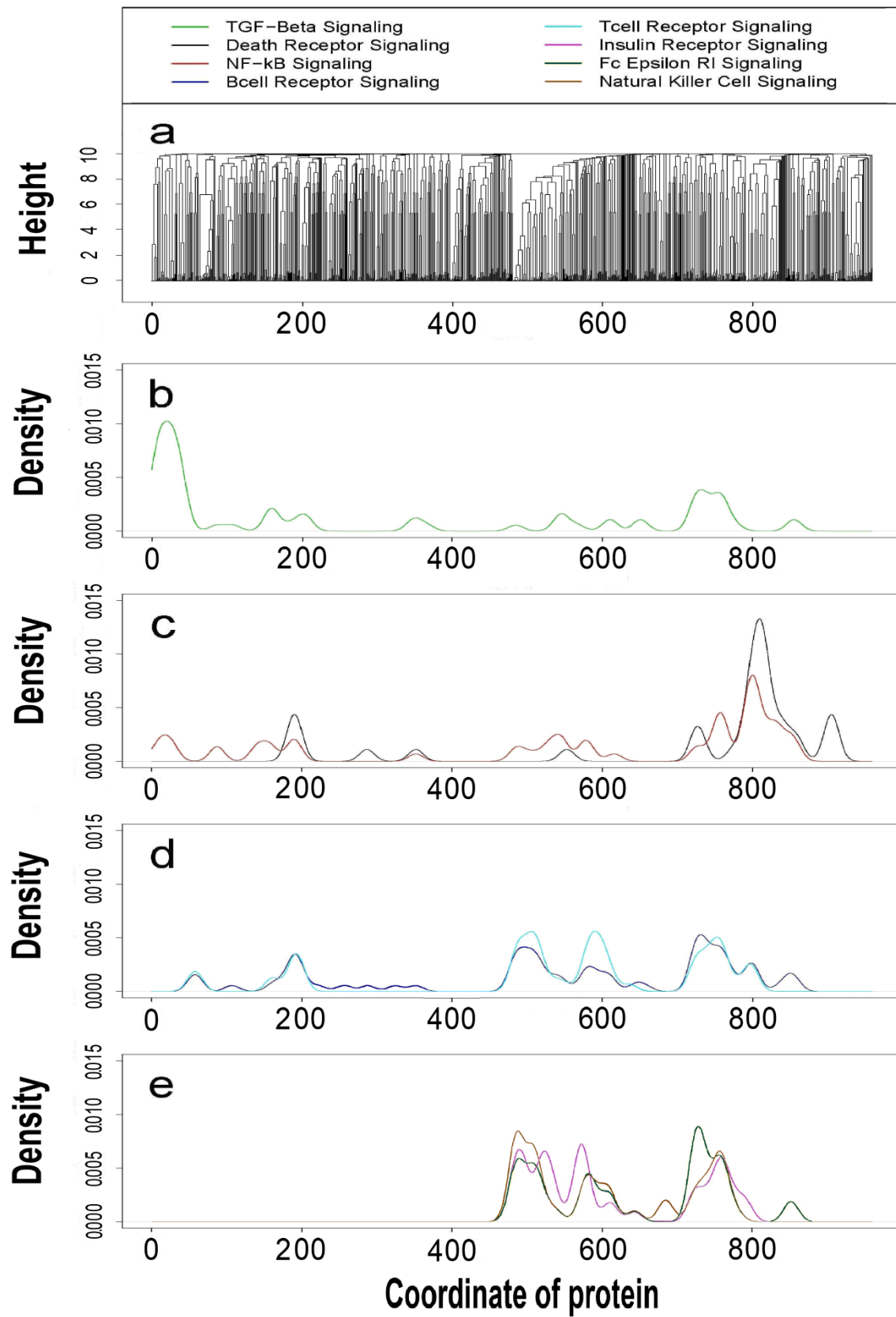


Figure 16 Distribution patterns of eight different signaling pathways. “(a) The largest subcluster of 959 proteins is derived from the root of the whole 1729-member cluster. Each

protein in this subcluster has a coordinate with respect to its order in the 959 members (from left to right); a pathway distribution is generated from the distribution of its members' coordinates under the bandwidth of 10 (R 2.25; IPA 5.5). (b) Distribution of the TGF- β signaling pathway. (c) Distributions of death-receptor and NF- κ B signaling pathways. (d) Distributions of B- and T-cell receptor signaling pathways. (e) Distributions of insulin receptor, Fc epsilon RI and natural killer cell signaling pathways.”^ψ Fig. 16 was reproduced from REF. 76.

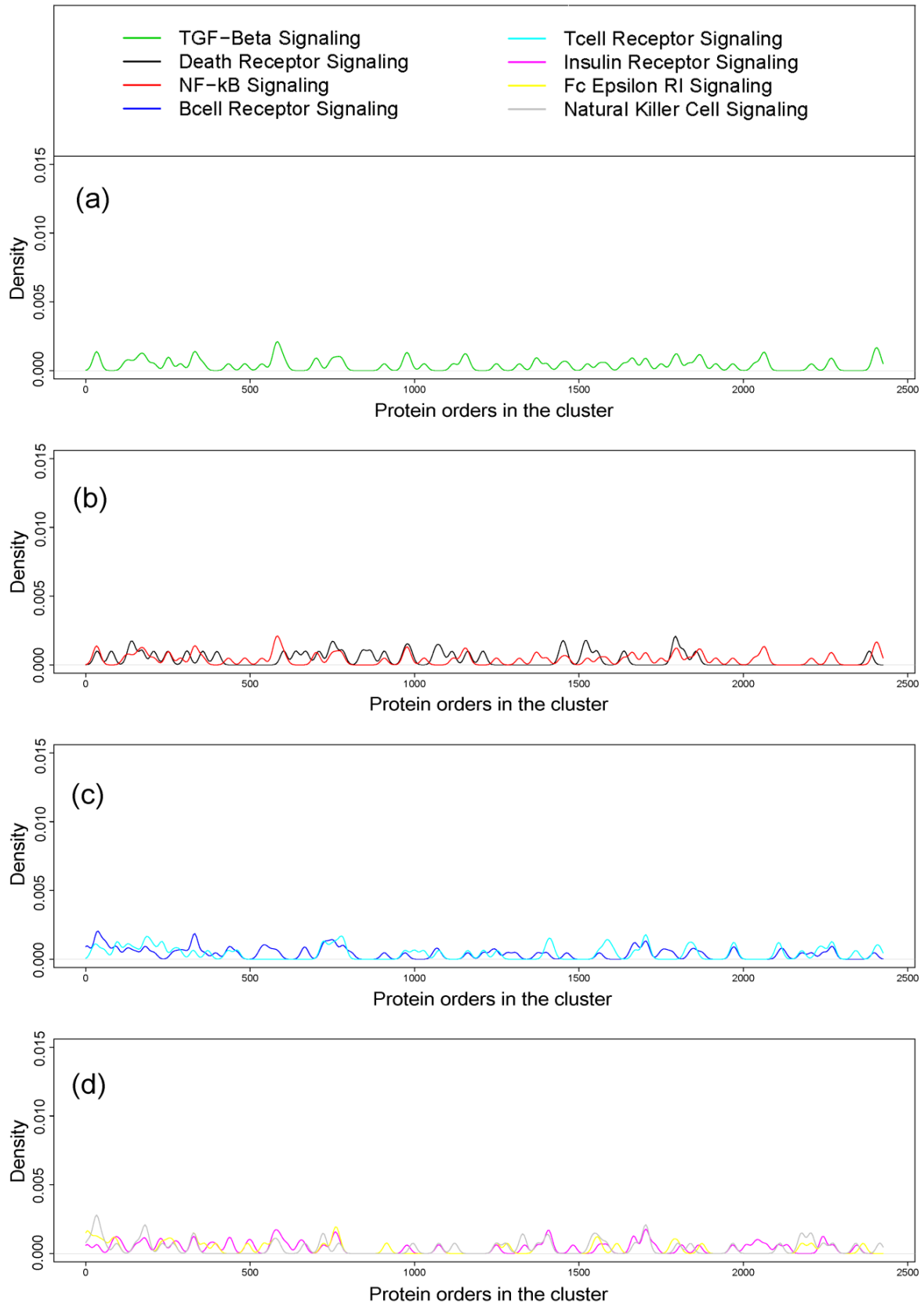


Figure 17 Distribution patterns of eight signaling pathways (the same as those in Fig. 16) from the random scale-free network. This largest of 2425-protein subcluster was derived from the root of the whole 3033-protein cluster which came from random scale-free network.

Similar as Fig. 16, each protein here has a coordinate which was used to generate distributions of signaling pathways under the bandwidth of 10 (R 2.9.0; IPA 8.5). Here, (a), (b), (c) and (d) correspond to Fig. 16 *b*, *c*, *d*, *e*, respectively. (a), (b), (c) and (d) were generated in the same way as in Fig. 16.

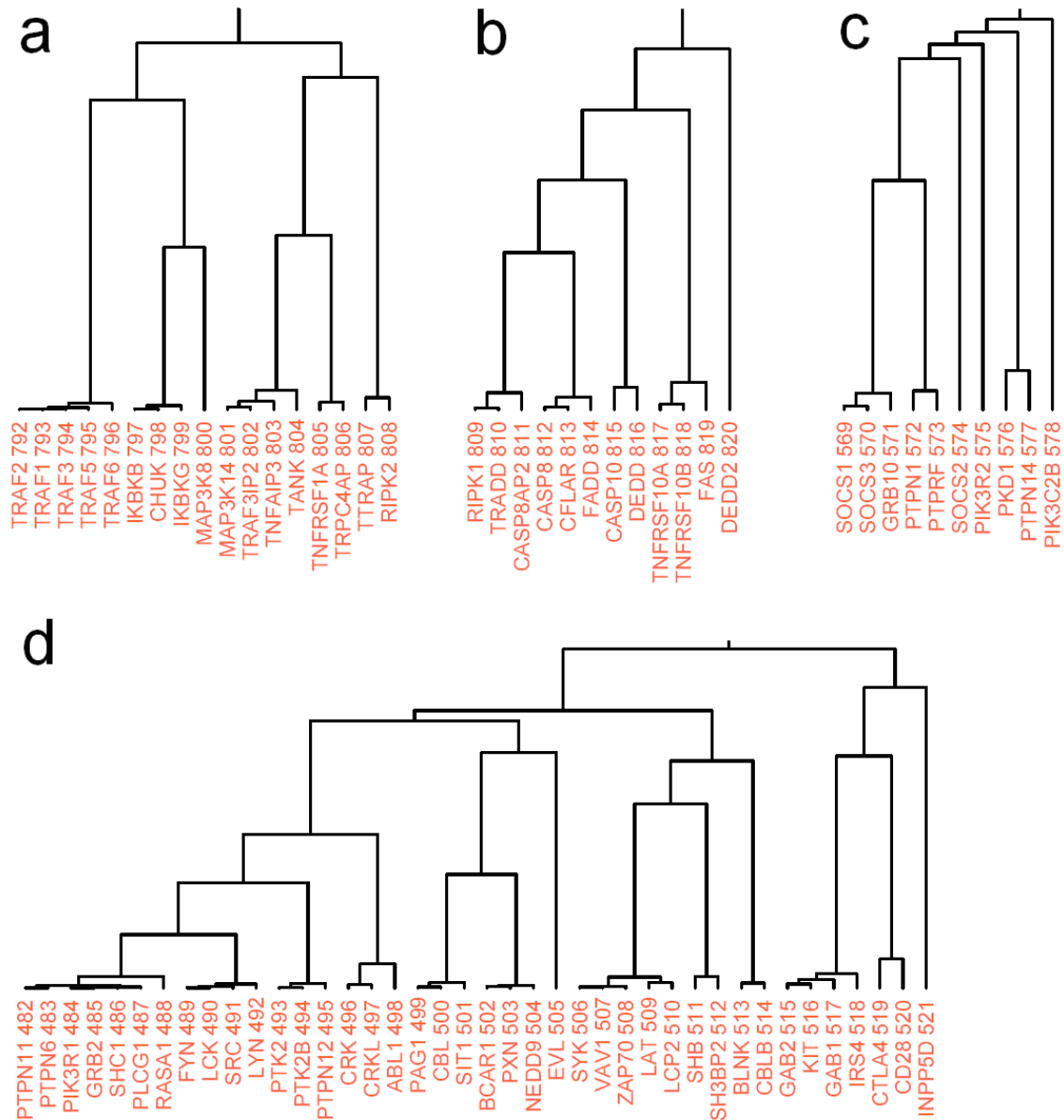


Figure 18 Four subclusters that are presumably involved in six signaling pathways. The indices above protein names are their coordinates in Fig. 16a). (a) The subcluster for NF- κ B signaling pathway ($P < 3.2 \times 10^{-20}$): RIPK2, TRAF3IP2, TRAF1, TRPC4AP and TANK are potential pathway members. (b) The subcluster for death receptor signaling pathway ($P < 2.4 \times 10^{-19}$): DEDD, DEDD2 and CASP8AP2 are potential pathway members. (c) The subcluster for insulin receptor signaling pathway ($P < 2.4 \times 10^{-10}$): PTPN14, PKD1 and SOCS2 are potential pathway members. (d) The subcluster for the immune response [natural killer cell signaling ($P < 1.7 \times 10^{-21}$), T-cell receptor signaling ($P < 4.9 \times 10^{-22}$) and B-cell receptor signaling ($P < 1.1 \times 10^{-17}$): CRK, PTK2, PTK2B, SRC, PTPN12, PXN, CRKL, EVL, SIT1, KIT, IRS4 and NEDD9 are potential members of the above three pathways. Fig. 18 was reproduced from REF. 76.

Comparison of performance between our prediction scheme and others

“We compared the performance of our prediction scheme with that of the direct prediction scheme used by Schwikowski *et al.* (2000) which infers the function of a protein from its direct interacting neighbors in the PPI network [95, 96].”^ψ Under the same criteria [i.e., the same minimum number (n) of annotations (shared by direct interacting neighbors) required to predict a function], our prediction scheme gave better FDRs in most cases (except $n=7$, 10 for KEGG predictions), while Schwikowski *et al.* (2000) gave more predictions at the cost of FDR (Fig. 19). From Fig. 12 and Fig. 19, we decided to use $n=2$ (for KEGG) and $n=4$ (for GO) as the thresholds of minimum number of annotations shared by significant partners, which gave us relatively low FDRs (21% for KEGG and 30% for GO) without sacrificing too many predictions (we obtained 466 predictions for KEGG and 123 predictions for GO were made). Under the same criteria (i.e., $n=2$ for KEGG and $n=4$ for GO), the FDRs of our predictions (30% for GO and 21% for KEGG) have been significantly improved over the FDRs (60% for GO and 49% for KEGG) from the direct prediction scheme [95], while the number of predictions we can make by our scheme has been dramatically decreased compared with that from Schwikowski *et al.* (2000) (Fig. 19).

For GO annotations, when the numbers of predictions were similar, our prediction scheme always gave better FDRs than that from Schwikowski *et al.* (2000) (Fig. 20a). For KEGG pathways, when the number of predictions was greater or equal to 58 (dashed line in Fig. 20b), our prediction scheme still gave better FDRs. When the number of predictions goes below 58, direct annotation scheme may give better predictions in terms of FDR, but the number of predictions becomes very limited. These results are reasonable because our

algorithms identified significant protein pairs that are more functionally associated than the direct-interacting pairs in the human PPI data, and we made functional inferences from these significant pairs, not from direct protein interactions which may suffer large amounts of false positives generated in high throughput assays [76].

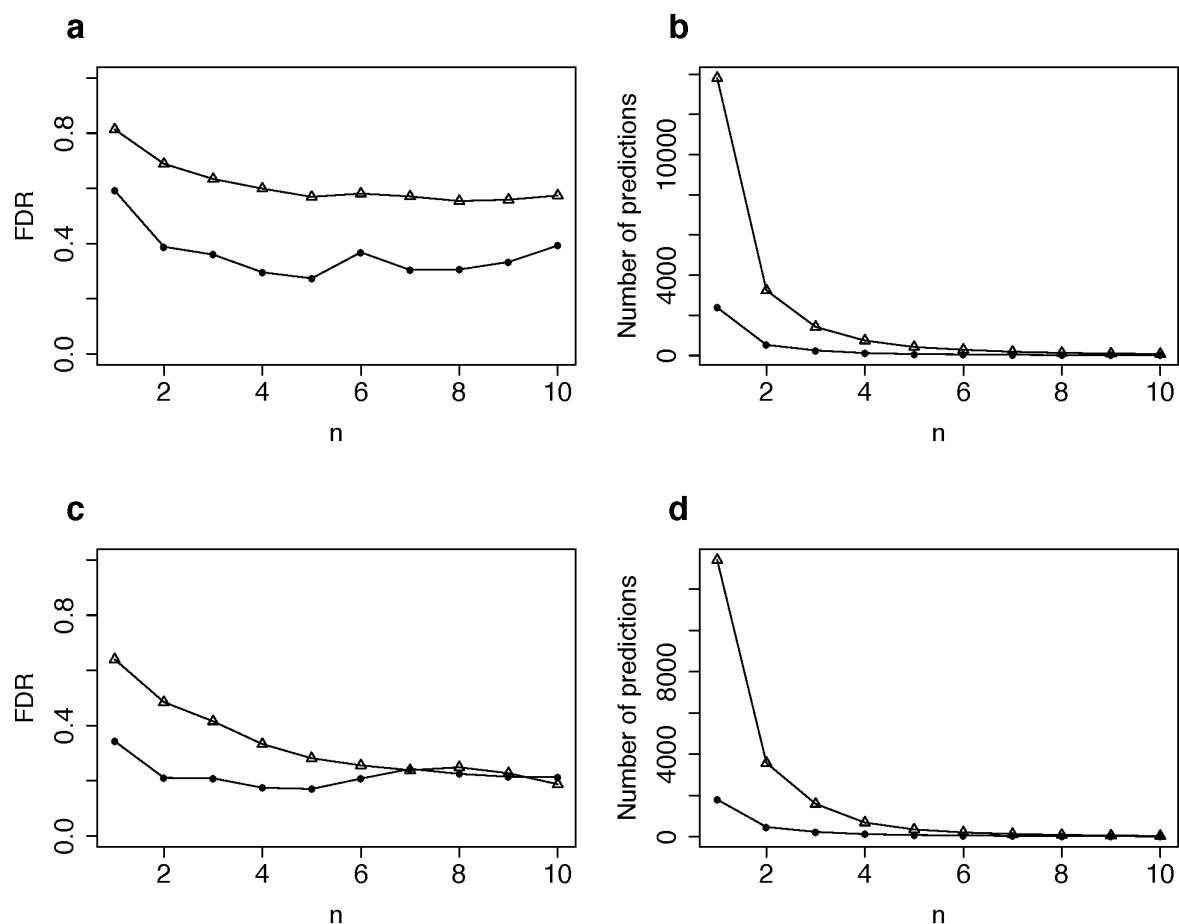


Figure 19 Estimation of the false discovery rates and the number of predictions we can make given different n . The lines with solid dots came from our prediction scheme, which the lines with triangles came from the prediction scheme used by Schwikowski *et al.* (2000). (a) Estimated FDR of predicted KEGG pathway at a given n . (b) The number of predictions we can make at a given n for KEGG pathway. (c) Estimated FDR of predicted GO terms at a given n . (d) The number of predictions we can make at a given n for GO terms.

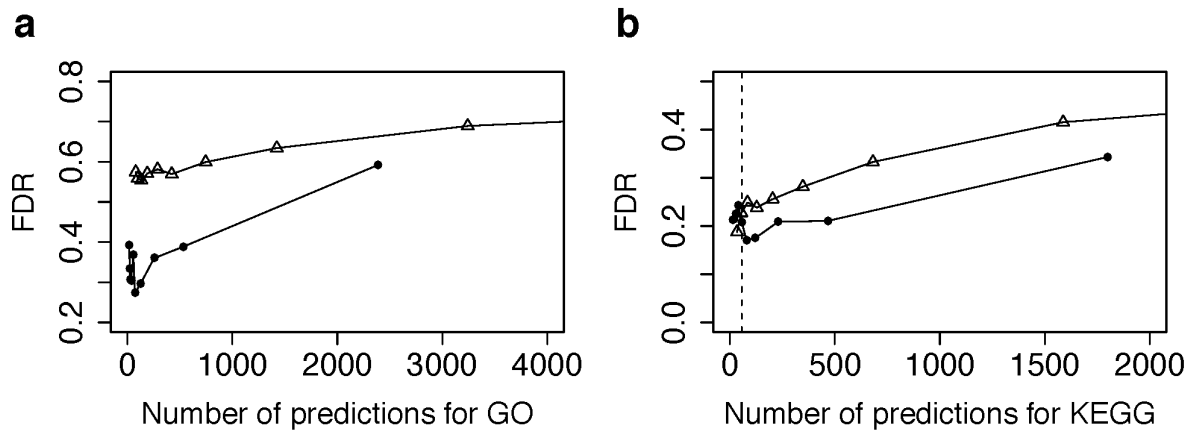


Figure 20 False discovery rates vs. the number of predictions. Solid dots stand for the FDRs of predicted annotations from our prediction scheme, triangles stand for the FDRs of predicted annotations from Schwikowski *et al.* (2000). (a) is for GO annotations and (b) is for KEGG annotations. Dashed line in (b) equals 58 (predictions).

We compared our methods with the latest module-assisted method that used interactive hierarchical cluster to analyze functional modules of PPI data [97, 98]. Arnau et al (2005) used the shortest path length as the distance measure between proteins in PPI networks. When the “ties in proximity” problem [99-101] was encountered, equally valid hierarchical clustering solutions were obtained randomly. Aldecoa et al (2010) improved the interactive clustering algorithm to make the implementation of algorithm faster; with their criteria, they were also able to determine the best partition for the dendrogram. After applying their algorithm (UVCluster as the interactive algorithm and UPGMA as tree algorithm, with 8000 iterations) on our human PPI data, we got 900 functional modules (subclusters) consisting of 7,362 proteins. To compare the performance between their clustering method and ours, we calculated the ratio of proteins that have the most prevalent function (or pathway) within each functional module. By taking the sizes of different functional modules into account, we defined a weighted overlap rate (WOR) based on KEGG and GO annotations to compare the overall quality between their cluster and ours:

$$\text{weighted KEGG overlap rate} = \frac{\sum_{i \in S} k_i \times n_i}{\sum_{i \in S} n_i}$$

$$\text{weighted GO overlap rate} = \frac{\sum_{i \in S} g_i \times n_i}{\sum_{i \in S} n_i}$$

Here g_i and k_i stands for the ratio of proteins that have the most prevalent KEGG and GO annotation in i^{th} functional module; n_i stands for the size of i^{th} functional module; S stands for all functional modules in the cluster. Only those functional modules with at least 10 members [i.e. $n_i > 9$ (usable subclusters)] will be used to calculate the WOR.

Since the best partition of our cluster was not determined, we tentatively cut our cluster at different heights and calculated corresponding WOR. Then we compared the rates with that calculated from Arnau et al (2005) and Aldecoa et al (2010) (Fig. 21). Above the height of 8, the mean sizes of subclusters are comparable with that from Aldecoa et al (2010): for KEGG annotation, our method yields a maximized WOR of 0.394 at the height of 9.44, compared with the WOR of 0.243 from Aldecoa; at the height of 9.44, the mean size of usable subclusters is 17.67, compared with 16.78 from Aldecoa. For GO annotation, our method yields a maximized WOR of 0.418 at the height of 8.24, compared with the WOR of 0.273 from Aldecoa; at the height of 8.24, the mean size of usable subclusters is 15.90, compared with 16.78 from Aldecoa. Since our clustering method didn't use any GO and KEGG annotation, the comparison at least shows that, with GO and KEGG annotation as the benchmarks, our clustering method can give better results than the best partition considered by Aldecoa. As for computing time, our method probably consumes much less time than Aldecoa if written in C/C++ since it doesn't require heavy computation at all.

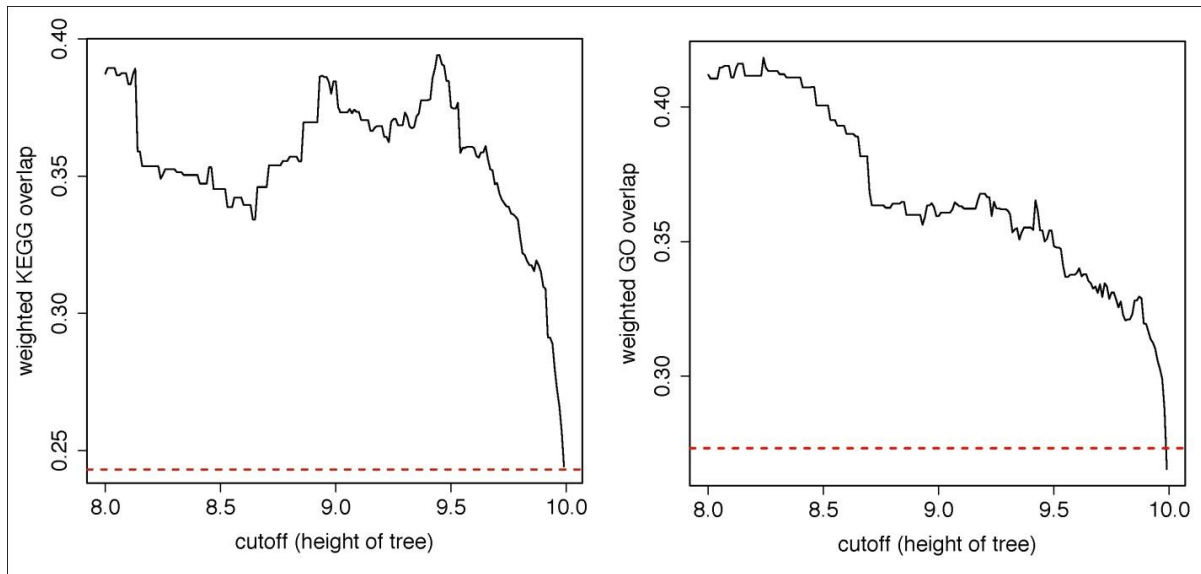


Figure 21 Comparison between the performance of our clustering method and the other one (Arnau et al 2005; Aldecoa et al 2010). X-axis stands for the height where we cut our cluster (dendrogram). Y-axis stands for the weighted overlap rate for our clustering method. Red line stands for the weighted overlap rate calculated from the best partition derived from Arnau et al (2005) and Aldecoa et al (2010).

CHAPTER 3
DISCUSSION

“An advantage of our prediction scheme, inherited from Samanta and Liang (2003), is the insensitivity to the high false positive rate of high-throughput PPI data. After adding 6086 randomly generated interactions (30.4% of the real data, assuming at least 50% false positive rate for high-throughput data), we were still able to recover on average 93.4% of significant protein pairs; furthermore, >90% of falsely generated “significant protein pairs” will become significant if we loosen the cutoffs of P_1 and P_2 a little to double the number of significant protein pairs. This will certainly offer more flexibility when selecting which PPI data to use.”

Ψ [76]

“Human proteins may have multiple functions and belong to different functional modules, so different signaling pathways may also have some pathway members in common. It is thus reasonable to assume that the overlap of distribution (Fig. 16*b–e*), especially of peaks, may reveal the functional relevance of different pathways. For example, the death-receptor and NF-κB signaling pathways overlap in the peak area, and the T- and B-cell receptor signaling pathways have a similar distribution. Therefore, the cluster and its pathway distributions will be useful in multi-pathway analysis and accurate function prediction.” Ψ [76]

With GO and KEGG as the benchmarks, we also compared the performance of P_1 and P_2 in terms of annotation overlap rate (Fig. 22): for the top 5,000 or top 10,000 protein pairs, where we can observe a smaller overlap rate when ranks go lower, P_1 obviously has better performance. From 10,001th protein pairs to 20,000th protein pairs, it becomes hard to tell why one performs better. However, since significant protein pairs are always selected within the top 10,000 pairs where very close functional associations can be expected between

protein pairs [75, 76], we determined that P_1 has a better performance than P_2 on ranking functional association (Fig. 22). However, Algorithm II was not designed to beat Algorithm I by itself: Algorithm I calculates the probability that two proteins, A and B, share m common neighbors. The m common neighbors may contain proteins with any degree: proteins with a large number of connections (i.e. hub protein) or proteins with only a few connections. This difference Algorithm I cannot tell but Algorithm II can. P_1 cannot tell because it does not take into account the degree of common neighbors. Algorithm II was designed so that a small P_2 indicates less connectivity of common neighbors, while large P_2 indicates more connectivity. A very small P_1 with a very small P_2 means that, two proteins have an unusual large number of common neighbors which are not hub proteins, while a very small P_1 with a large P_2 means that, two proteins have an unusual large number of common neighbors which are hub proteins (as shown in Table 1). Since Algorithm II was not designed to calculate the probability of sharing m common neighbors (i.e., not an improved algorithm based on Algorithm I), it's not a surprise to see P_1 gives better ranking results if we directly compare P_1 with P_2 .

GO terms are widely adopted to enhance and/or verify the credibility of functional relationship: the overlap of GO terms indicates functional association [23, 24]. However, because of the highly complex hierarchy of the Gene Ontology tree, we may arrive at some false conclusions if we use GO terms at all levels indiscriminately. For example, the overlap of GO terms at a low level in the GO tree is always less meaningful compared with the overlap of its offspring GO terms. Therefore, in this dissertation, we used GO terms at the highest levels in the GO tree (see methods and materials) which were considered to be the

most detailed functional annotations we could obtain.

The error rate of curated gene annotations can be significant (Jones et al 2007). The overall error rate of curated GO term annotations (excluding evidence code “IEA”) was found to lie between 28%-30%, which was better than the error rate (33% - 43%) of UniProt/SwissProt database annotations (the latter was widely considered to have a very high standard of curation) (Artamonova et al 2005; Jones 2007). The error rate of all GO term annotations (including “IEA”) is, as far as I know, unknown, but could be much higher than 30%. (The error rate of KEGG pathway annotations is also unknown at this point). To test if our result is sensitive to the potentially high error rate of GO and KEGG annotations, we assessed the performance of our method after adding false annotations to 7,362 proteins. We first calculated the annotation distribution from all GO/KEGG annotations of 7,362 proteins; then we generated different amounts of GO/KEGG annotations from the distribution (10%, 20% ... 90%, 100% of all GO/KEGG annotations of the 7,362 proteins) and randomly assigned them to 7,362 proteins. We found that, although FDR will be higher for both blue and red curves (as shown in Fig. 23; please compare with Fig. 10), the result will still be the same: using P_1 and P_2 together will always be better than using P_1 alone.

“We also developed a new algorithm for computing the probabilities that three proteins share m interacting partners (see Appendix A). However, we found that if three proteins have a very low probability of sharing m interacting partners, in most cases two of them will have a very low P_1 . Because this algorithm is highly dependent on Algorithm I (P_1), we do not think it provides more information worthy of further investigation.”^ψ [76]

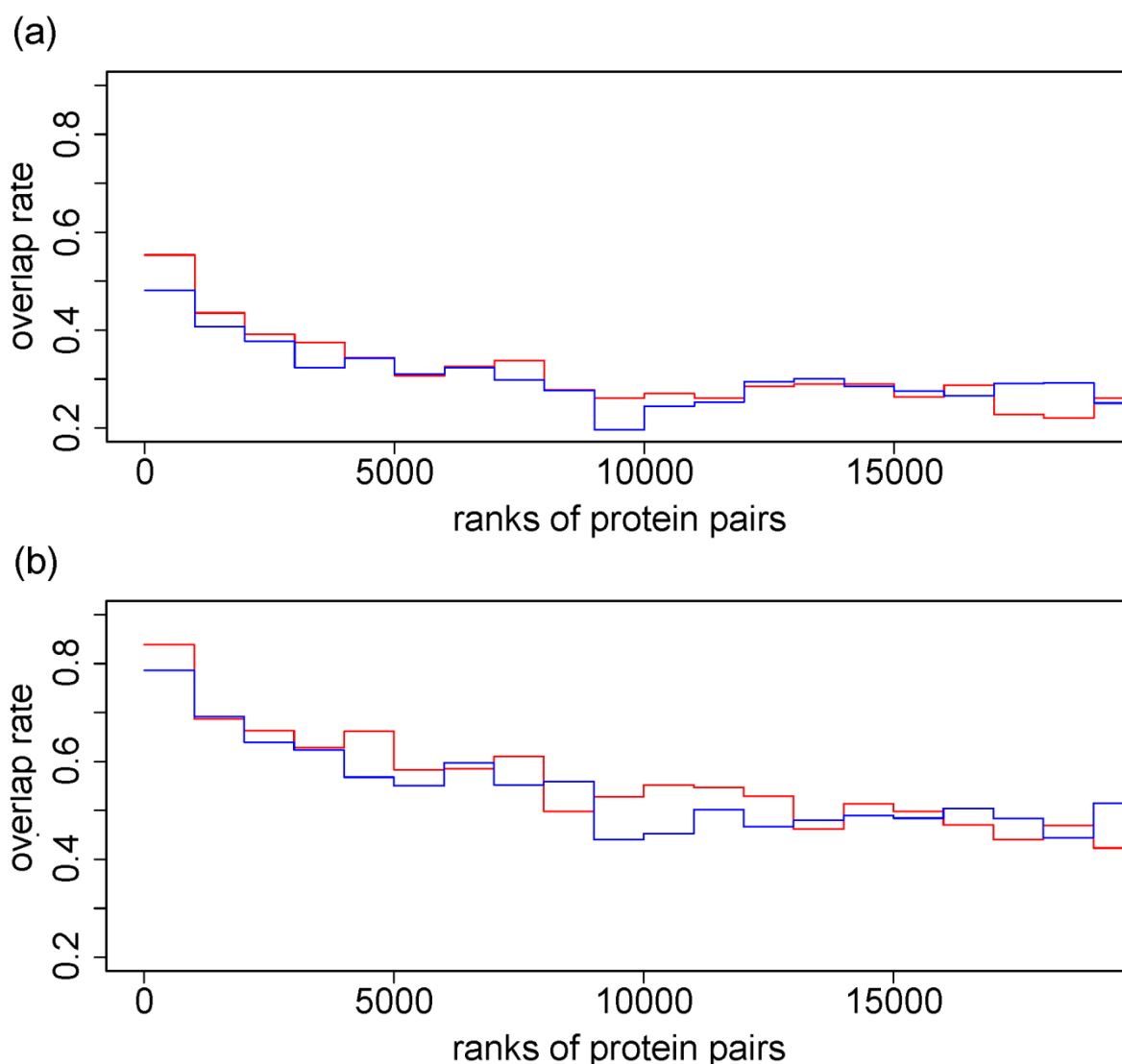


Figure 22 Comparison between the performance of P_1 and P_2 in terms of annotation overlap rate. Protein pairs were ranked by P_1 for the red curve and ranked by P_2 for the blue curve. The top-ranked 20,000 pairs are divided equally into 20 bins. In each bin (1,000 protein pairs), we calculated the GO overlap rate in (a) and the KEGG overlap rate in (b). (a) The red curve stands for the GO overlap rates of proteins pairs ranked by P_1 ; the blue one stands for the GO overlap rates of protein pairs ranked by P_2 . (b) The red curve stands for the KEGG overlap rates of proteins pairs ranked by P_1 ; the blue one stands for the KEGG overlap rates of protein pairs ranked by P_2 .

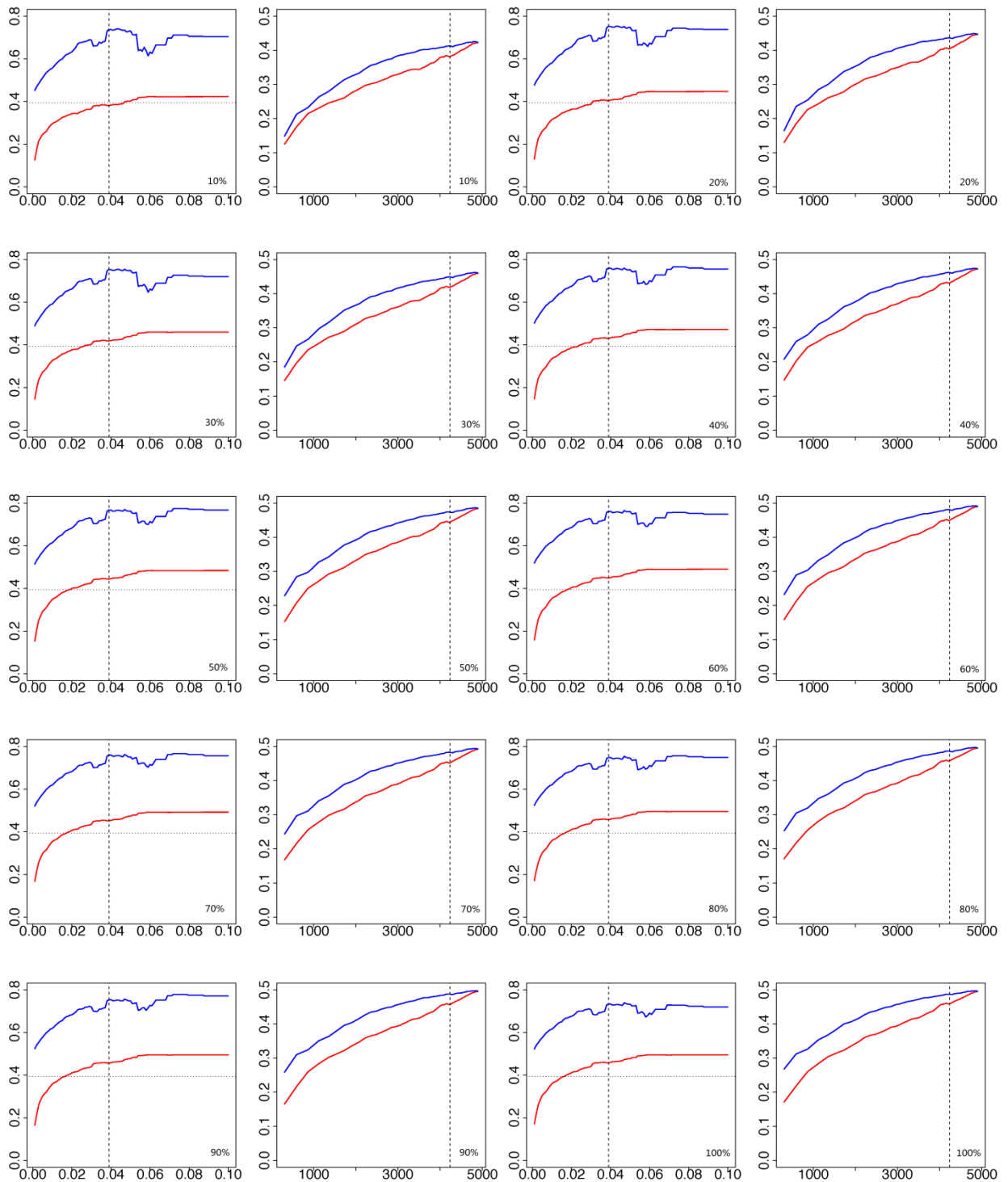


Figure 23 Algorithm II decreased the false discovery rate (FDR) of our predictions even after adding false annotations. The meaning of X-axis, Y-axis, dashed lines and solid colored (red and blue) curves are the same thing as shown in figure 10. These results were generated after we added 10%, 20% ... 100% (indicated at the lower right corner in each plot) randomly generated annotations to the GO and KEGG annotations of all proteins used in this thesis.

CHAPTER 4
METHODS AND MATERIALS

Protein–Protein Interaction Data

“From the BioGRID (www.thebiogrid.org), we downloaded the human PPI data (version 2.20), which derived from both conventional focused studies (~69.6%) and high-throughput studies (yeast two-hybrid and affinity capture; ~30.4%) [77]. There are 20,019 total non-redundant interactions (excluding self-interactions) and 7,362 protein entries in this dataset, including 42 nonhuman proteins that interact with human proteins.”^Ψ [76]

Benchmarks for Evaluating the Functional Association

“We used GO and KEGG as independent benchmarks to assess the functional association of each protein pair. GO and KEGG databases provide specific pathways, functions and cellular components for proteins in our PPI data: we classified the 7,362 proteins into 237 KEGG pathways and 1956 qualified GO terms (including biological process, molecular function and cellular component). These databases are good references for evaluating functional association because of its reasonable coverage of the genome and its large number of categories, which makes it improbable to have random matching of pathways.”^Ψ [76]

Annotation Overlap Rate

“With GO annotation (R package: GO, 08-Aug-2006), we defined the GO overlap rate as follows: $overlaprate = \frac{T_Q}{T_A}$, where T_Q is the number of protein pairs of which both proteins share at least one qualified GO term; T_A is the number of protein pairs of which both proteins are annotated with qualified GO terms. Here “qualified GO terms” means GO

terms at the highest level without direct or indirect GO “offspring” terms in each ontology (the level is defined as the number of nestings from the root node (level 1) in the Gene Ontology DAG file [79]).”^Ψ [76]

“We defined the KEGG overlap rate in the same way as above (R package: KEGG, Release 41.1). We used the GO and KEGG overlap rates to assess the functional association of protein pairs: a higher overlap rate corresponds to a closer functional relationship.”^Ψ [76]

Definition of FDR for the Declared Significant Functional Associations

“Suppose the GO and KEGG pathways are complete: if both proteins in each pair have KEGG pathway identifiers and qualified GO terms, we call them declared positive protein pairs. If they share at least one identifier (either GO or KEGG identifier), we consider this declared association true positive; otherwise we consider it false positive. Therefore, the

FDR can be written as follows:
$$FDR = 1 - \frac{\text{number of true positive protein pairs}}{\text{number of declared positive protein pairs}}.$$
”^Ψ [76]

“This false discovery rate is used to assess the performance of our algorithm as we expect an improved annotation scheme will lower the proportion of wrong predictions among declared significant functional associations.”^Ψ [76]

Fisher’s Exact Test

For each protein with an assigned annotation, we first calculated the number of its total neighbors and the number of neighbors with the assigned annotation; we then obtained the total number of proteins and the number of proteins with the assigned annotation from

corresponding databases (data were downloaded from NCBI and KEGG ftp on Aug 16th, 2009 and Aug 21st, 2009, respectively). With these data, we performed Fisher's exact test (two sided) for each assigned annotation and adjusted P values by "BH" method [81]. We listed the results in Table S2.

Pathway Analysis Tool

In this dissertation, pathway members were identified by Ingenuity Pathway Analysis (IPA 5.0, 5.5 and 8.5; Ingenuity[®] Systems, Inc., Redwood City, CA, www.ingenuity.com). *P* values for signaling pathways were calculated from Ingenuity Pathway Analysis. Fig. 14, 15, 16 and 18 were generated through the use of Ingenuity Pathway Analysis.

SUMMARY AND FUTURE PERSPECTIVE

Protein-protein interaction (PPI) network is a very useful tool to perform a large-scale protein function investigation in the post-genomic era. With high-throughput screening technologies such as yeast two-hybrid assays, biologists have immensely accelerated the accumulation of protein interaction data across human and model species. These newly available data have made it possible to build global PPI networks and hence study protein function in the context of a large-scale network.

So far, partial protein interaction maps of model species have been depicted and network-based annotation schemes are being developed to investigate protein functions. Here we propose a new method to detect certain non-randomness in the protein-protein interaction network. Unlike analyses that measure non-randomness globally, such as on power-law distributions, our method measures and ranks non-randomness for each pair of proteins, which allows us to make large-scale functional predictions (>500) with high reliability. Our method also clusters proteins properly into functional modules that cover a variety of signaling pathways. Further analysis of modules with significant P values reveals the possible roles of proteins from the pathway perspective and highlights potential new members in certain signaling pathways important in cancer research.

The method we proposed in this dissertation can be used as a general tool for high throughput functional annotations; it can also derive functional modules enriched in important pathways and identify potential new members, which will spur the additional interests from biologists. Besides PPI networks, our method is also applicable in other types of networks with a similar distribution of degrees (i.e. scale-free distribution), such as gene

regulatory network and human social network [9, 12].

Future studies focusing on common-neighbor based statistics should first address one important problem: how to combine Algorithm I and II into one single algorithm which takes into account the degrees of all proteins in PPI networks. Since using two algorithms together needs our arbitrary threshold(s), it becomes hard to run automatic computer programs to identify significant functional associations between proteins. By contrast, a single algorithm with a pre-set P value as the cut-off will not only make computer programming much easier and straightforward, but also help biologists better understand and utilize common-neighbor statistics in their research fields.

Another interesting perspective is the future application of our method with nodes (e.g. genes, proteins, annotations) from two totally different networks. For example, to assess the relationship between direct interacting neighbors in a PPI network, biologists probably want to check if any two interacting proteins share the same functional annotations which come from a totally different type of networks [23, 24] (e.g. GO annotations which are hierarchically ordered, and the KEGG pathway network). However, even the overlap of functional annotations does occur between two proteins, this doesn't mean any statistical significant without further analysis: two proteins may only share one common function which is "ubiquitous" for proteins, such as ATP binding, Zinc ion binding and Calcium ion binding. We think our method, which can take into account both the number and the prevalence of the shared annotations, will probably be able to give a reliable statistical significance on the functional association between two proteins and reduce the influence of those "ubiquitous" functions on future predictions. For the same reason, we may also be able to put P values on

our claimed “true positive protein pairs” (see methods and materials) and possibly reduce false positive rates caused simply by random chance. Therefore, we believe our work in this dissertation is a very important and solid groundwork for future research.

CONCLUSION

High throughput gene discovery by sequencing demands corresponding high throughput annotations. The protein-protein interaction data from high throughput assays has made it possible to study protein functions in the context of a large-scale network. However, the interactions acquired from high throughput assays suffer a high error rate. How to extract reliable functional inference from the error-prone network becomes more significant. We have developed a new method to detect certain non-randomness in the protein-protein interaction network. Unlike analyses that measure non-randomness globally, such as on power-law distributions, our method measures and ranks non-randomness for each pair of proteins. This allows us to make large-scale functional predictions with top-ranking protein pairs. We applied our method to human data and assigned >500 annotations for human proteins. In addition, we have built a cluster to better illustrate the functional association between proteins and highlight potential new pathway members in certain signaling pathways important in cancer research.

BIBLIOGRAPHY

1. Hartwell, L.H., Hopfield, J.J., Leibler, S. & Murray, A.W. (1999) From molecular to modular cell biology. *Nature* 402, C47-C52.
2. Hasty, J., Macmillen, D. & Collins, J.J. Engineered gene circuits. (2002) *Nature* 420, 224-230.
3. Kitano, H. Computational systems biology (2002). *Nature* 420, 206-210.
4. Koonin, E.V., Wolf, Y.I. & Karev, G. (2002) The structure of the protein universe and genome evolution. *Nature* 420, 218-223.
5. Oltvai, Z.N. & Barabasi, A.-L. (2002) Life's complexity pyramid. *Science* 298, 763-764.
6. Wall, M.E., Hlavacek, W.S. & Savageau, M.A. (2004) Design of gene circuits: lessons from bacteria. *Nature Rev. Genet.* 5, 34-42.
7. Bray, D. (2003) Molecular networks: the top-down view. *Science* 301, 1864-1865.
8. Alon, U. (2003) Biological networks: the tinkerer as an engineer. *Science* 301, 1866-1867.
9. Barabasi A.-L. & Oltvai Z.N. (2004) Network biology: understanding the cell's functional organization. *Nature Rev Genet* 5, 101-113.
10. Watts, D.J. & Strogatz, S.H. (1998) The collective dynamics of "small-world" networks. *Nature (London)* 393, 440-442.
11. Williams, R.J. & Martinez N.D. (2000) Simple rules yield complex food webs. *Nature (London)* 404, 180-183.
12. Wasserman, S. & Faust, K. (1994) *Social Network Analysis* (Cambridge Univ. Press, Cambridge, U.K.).
13. Scott, J. (2000) *Social Network Analysis: A Handbook* (Sage, London), 2nd Ed.

14. Amaral, L.A.N., Scala, A., Barthelemy, M. & Stanley, H.E. (2000) Classes of small-world networks. *Proc. Natl. Acad. Sci. USA* 97, 11149-11152.
15. Newman, M.E.J. (2001) The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* 98, 404-409.
16. Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999) On power-law relationships of the internet topology. *Comput. Commun. Rev.* 29, 251-262.
17. Albert, R., Jeong, H. & Barabási, A.-L. (1999) Statistical mechanics of complex networks. *Nature (London)* 401, 130-131.
18. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. (2000) Graph structure in the web. *Comput. Networks* 33, 309-320.
19. Erdos, P. & Renyi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17-61.
20. Barabasi, A.-L. & Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509-512.
21. Wagner, A. (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* 18, 1283-1292.
22. Yook, S.-H., Oltvai, Z.N. & Barabasi, A.-L. (2004) Functional and topological characterization of protein interaction networks. *Proteomics*, 4: 928-42.
23. Rual J, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437:1173-1178.

24. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122:957-968.
25. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302:1727-1736.
- 26.** Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303:540-543.
27. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411:41-42.
28. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627.
29. Ito, T., Chiba, T., Ozawa R., Yoshida, M., Hattori, M., Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* 98, 4569-4574.
30. J. Reboul, P. Vaglio, J-F. Rual, P. Lamesch, M. Martinez et al. (2003) *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* 34. 35-41.
31. A.C. Gavin, M Bosche, R Krause, P Grandi, M Marzioch, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-147.

32. Y. Ho, A Gruhler, A Heibut, GD Bader, L Moore, et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180-183.
33. J-D.J. Han, D. Dupuy, N. Bertin, M.E. Cusick & M. Vidal. (2005) Effect of sampling on topology predictions of protein-protein interaction networks. *Nat Biotechnol*, 23, 839-844.
34. RL Finley Jr, R Brent (1994) Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc Natl Acad Sci USA*, 91:12980-12984.
35. C Bendixen, S Gangloff, R Rothstein (1994) A yeast mating-selection scheme for detection of protein-protein interactions. *Nucleic Acids Res* 22: 1778-1779.
36. JP Parrish, KD Gulyas and RL Finley Jr. (2006) Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol* 17:387-393.
37. PL Bartle, JA Roecklein, D SenGupta, S Fields (1996) A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet* 1996, 12:72-77.
38. M Fronmont-Racine, JC Rain, P Legrain (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* 16:277-282.
39. AJ Walhout, M Vidal (2001) High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* 24:297-306.
40. J Zhong, H Zhang, CA Stanyon, G Tromp, RL Finley Jr (2003) A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res*, 13:2691-2699.

41. F Jin, T Hazbun, GA Michaud, M Salcius, PF Predki, S Fields, J Huang (2006) A pooling-deconvolution strategy for biological network elucidation. *Nat Methods*, 3:183-189.
42. O. Puig, F. Caspary, G. Rigaut, B. Rutz, E. Bouveret et al (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24, 218-229.
43. A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm, et al (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* 93, 14440.
44. W.P. Blackstoc and M.P. Weir (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, 121.
45. G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann and B. Seraphin (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* 17, 1030-1032.
46. E. Bouveret, G Rigaut, A Shevchenko, M. Wilm and B. Seraphin. (2000) A sm-like protein complex that participates in mRNA degradation. *EMBO J.* 19, 1661-1667.
47. D'haeseleer P, Church GM (2004) Estimating and improving protein interaction error rates. *Proc IEEE Comput Syst Bioinform Conf* 216-223.
48. Chien CT, Bartel PL, Sternglanz R, Fields S (1991) The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* 88:9578-9582.

49. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, et al. (2001) The protein-protein interaction map of *Helicobacter pylori*. Nature 409:211-215.
50. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, et al. (2000) Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci U S A 97:1143-1147.
51. Lehner B, Fraser A (2004) A first-draft human protein-interaction map. Genome Biol 5:R63.
52. Enright AJ, Iliopoulos J, Kyrpides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. Nature 402:86-90.
53. Ramani AK, Bunesu RC, Mooney RJ, Marcotte EM (2005) Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. Genome Biol. 6, R40.
54. H. Yu, P Braun, M.A. Yildirim, I. Lemmens, K Venkatesan et al. (2008) High-quality binary protein interaction map of the yeast interactome network. Science 322, 104-110.
55. H. Ge, A.J. Walhout and M. Vidal. (2003) Integrating 'omic' information: a bridge between genomics and systems biology. Trends genet. 19, 551-560.
56. Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci U S A 101:4781-4786.
57. Vazquez A, Flammini A, Maritan A, Vespignani A (2003) Global protein function prediction from protein-protein interaction networks. Nat Biotechnol 20:697-700.
58. Karaoz U, Murali TM, Letovsky S, Zheng Y, Ding C, et al. (2004) Whole-genome

- annotation by using evidence integration in functional-linkage networks. *Proc Natl Acad Sci U S A* 101:2888–2893.
59. Nabieva E, Jim K, Agarwal A, Chazelle B, Singh M (2005) Whole proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics* 21:i302–i310.
60. Deng M, Zhang K, Mehta S, Chen T, Sun F (2003) Prediction of protein function using protein–protein interaction data. *J Comput Biol* 10:947–960.
61. Bader GD, Hogue CW (2002) Analyzing yeast protein–protein interaction data obtained from different sources. *Nat Biotechnol* 20:991–997.
62. Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics* 7:207.
63. Sharan R, Ideker T, Kelley B, Shamir R, Karp RM (2005) Identification of protein complexes by comparative analysis of yeast and bacterial protein interaction data. *J Comput Biol* 12:835–846.
64. Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci U S A* 100:12123–12128.
- 65.** Przulj N, Wigle DA, Jurisica I (2004) Functional topology in a network of protein interactions. *Bioinformatics* 20:340–348.
66. King AD, Przulj N, Jurisica I (2004) Protein complex prediction via cost-based clustering. *Bioinformatics* 20:3013–3020.

67. Goldberg DS, Roth FP (2003) Assessing experimentally derived interactions in a small world. *Proc Natl Acad Sci U S A* 100:4372-4376.
68. Llewellyn R, Eisenberg DS (2008) Annotating proteins with generalized functional linkages. *Proc Natl Acad Sci U S A* 105:17700-17705.
69. Geisler-Lee J, O'Toole N, Ammar R, Provart NJ, Millar AH, et al. (2007) A predicted interactome for Arabidopsis. *Plant Physiol* 145: 317-329.
70. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, et al. (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 287: 116-122.
71. Gandhi TKB, Zhong J, Mathivanan S, Karthick L, Chandrika KN, et al. (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* 38:285-293.
72. H. Goehler, M. Lalowski, U Stelzl, S. Waelter, M. Stroedicke, et al. (2004) A protein interaction network links GIT1, an Enhancer of Huntingtin Aggregation, to Huntington's disease. *Mol. Cell.* 15, 853-865.
73. J. Lim, T. Hao, C. Shaw, A.J. Patel, G. Szabo, et al. (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*, 125, 801-814.
74. J. Xu and Y. Li (2006) Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics*, 22, 2800-2805.
75. Samanta MP, Liang S (2003) Predicting protein functions from redundancies in large-scale protein interaction networks. *Proc Natl Acad Sci U S A* 100:12579-12583.

76. H. Li, S. Liang. (2009) Local network topology in human protein interaction data predicts functional association. PLoS ONE 4(7): e6410.
77. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, et al. (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34:D535-D539.
78. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99: 7821-7826.
79. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25-29.
80. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27-30.
81. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J.R. Stat Soc Ser B 57: 289–300
82. Jones C, Brown A, Baumann U. (2007) Estimating the annotation error rate of curated GO database sequence annotations. BMC bioinformatics 8:170.
83. Breitkreutz BJ, Stark C, Tyers M (2003) Osprey: a network visualization system. Genome Biol 4:R22.
84. Chen W, Lam SS, Srinath H, Schiffer CA, Royer Jr WE, et al. (2007) Competition between Ski and CREB-binding protein for binding to Smad proteins in transforming growth factor- β signaling. J Biol Chem 282:11365-11376.
85. Sun Y, Liu X, Eaton EN, Lane WS, Lodish HF, et al. (1999) Interaction of the Ski oncoprotein with Smad3 regulates TGF-beta signaling. Mol Cell 4:499-509.

86. Mazzarella R, Pengue G, Jones J, Jones C, Schlessinger D (1998) Cloning and expression of an immunoglobulin superfamily gene (IGSF1) in Xq25. *Genomics* 48:157-162.
87. Laurikkala J, Kassai Y, Pakkasjarvi L, Thesleff I, Itoh N (2003) Identification of a secreted BMP antagonist, ectodin, integrating BMP, FGF, and SHH signals from the tooth enamel knot. *Dev Biol* 264:91-105.
88. McMahon JA, Takada S, Zimmerman LB, Fan CM, Harland RM, et al. (1998) Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev* 12:1438-1452.
89. Hocevar BA, Smine A, Xu XX, and Howe PH (2001) The adaptor molecule disabled-2 links the transforming growth factor β receptors to the Smad pathway. *EMBO J* 20, 2789–2801.
90. Datta, P. K., Chytil, A., Gorska, A. E., and Moses, H. L. (1998) Identification of STRAP, a novel WD Domain protein in transforming growth factor- β signaling. *J Biol Chem* 273: 34671–34674
91. Datta, P. K., and Moses, H. L. (2000) STRAP and Smad7 synergize in the inhibition of transforming growth factor β signaling. *Mol Cell Biol* 20: 3157–3167.
92. Ikeda Y, Lmai Y, Kumagai H, Nosaka T, Morikawa Y, et al. (2004) Vasin, a transforming growth factor β -binding protein expressed in vascular smooth muscle cells, modulates the arterial response to injury in vivo. *Proc Natl Acad Sci U S A* 101:10732–10737.
93. Abreu JG, Ketpura NI, Reversade B, Robertis EM (2002) Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF- β . *Nat Cell Biol* 4:599-604.

94. Hildebrand A, Romaris M, Rasmussen LM, Heinegard D, Twardzik DR, et al. (1994) Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem J* 302:527-534.
95. Schwikowski B, Uetz P, Fields S (2000) A network of protein–protein interactions in yeast. *Nat Biotechnol* 18:1257–1261.
96. Sharan R, Ulitsky I, Shamir R (2007) Network-based prediction of protein function. *Mol Syst Bio* 3:88
97. Arnau V, Mars S, Marin I (2005) Interactive cluster analysis of protein interaction data. *Bioinformatics* 21: 364-378.
98. Aldecoa R, Marin I. (2010) Jerarca: efficient analysis of complex network using hierarchical clustering. *PLoS ONE* 5(7): e11585.
99. Backeljau T, De Bruyn L, De Wolf H, Jordaens K, Van Dongen S, Winnepenninckx B. (1996) Multiple UPGMA and Neighbor-joining trees and the performances of some computer packages. *Mol. Biol. Evol.* 13:309-313.
100. Takezaki N. (1998) Tie trees generated by distance methods of phylogenetic reconstruction. *Mol Biol. Evol.* 15:727-737.
101. MacCuish J, Nicolaou C, MacCuish N.E. (2001) Ties in proximity and clustering compounds. *J. Chem. Inf. Comput. Sci.* 41:134-146.
102. Artamonova II, Frishman G, Gelfand MS, Frishman D. (2005) Mining sequence annotation databanks for association patterns. *Bioinformatics*, 21(3): ii49-ii57.

Table 1 Examples of protein pairs with a very good P_1 that is caused by hub proteins. P_2 can identify these protein pairs and remove them from the top 5000 protein pairs.

Protein A	Protein B	Common neighbors	Degree of common neighbors
Tssk3	Itk	SMAD4, TGFBR1	108, 138
FBXW5	ZNF439	KRTAP4-12, MDFI	73, 97
CD33	PILRA	PTPN6, PTPN11	58, 69
SLC25A6	MCSP	PLSCR1, KRTAP4-12	68, 73
SHREW1	CA9	CTNNB1, CDH1	81, 28

Table 2 Top 10 protein pairs from our 4,233 significant protein pairs. All of them share close functional relationships. Table 2 was reproduced from REF. 76.

Protein_A	Protein_B	Ln(P_1)	Functional Relationship
SMAD3	SMAD2	−157.6068	SMAD family member
TUBB	TUBB2	−136.0437	Cellular structural activity
PTPN11	PTPN6	−125.8552	Proliferation of cells
BMPR1B	TGFBR1	−124.9466	Differentiation of cells
CALM2	CALM3	−124.9368	Calcium-modulated proteins
MAPK1	MAPK3	−113.0905	MAP kinase family member
CALM1	CALM3	−112.6375	Calcium-modulated proteins
IXL	MED9	−107.7585	Mediator complex
PIK3R1	GRB2	−107.7070	Tyrosine phosphorylation
CALM1	CALM2	−106.1716	Calcium-modulated proteins

Table 3 Selected Predictions of KEGG and GO annotations for human proteins. The 2nd column is the predicted KEGG and GO IDs for proteins in the 1st column. The 3th column is the corresponding KEGG pathway name and GO term. Ratio is the number of significant partners with the assigned annotation(s) divided by the total number of significant partners. P values were first calculated by Fisher's exact test and then adjusted by "BH" method [81]. Table 3 was partially reproduced from REF. 76.

Protein	KEGG	KEGG Pathway Name	Ratio	P value
CDC5L	hsa04110	Cell cycle	4/5	1.09E-05
DEDD	hsa04210	Apoptosis	4/5	4.22E-06
KSR2	hsa04010	MAPK signaling pathway	4/5	0.000171
GMFB	hsa04010	MAPK signaling pathway	6/6	3.44E-07
ITGB1	hsa04640	Hematopoietic cell lineage	4/6	1.07E-05
PTK2B	hsa04630	Jak-STAT signaling pathway	21/68	1.39E-12
GDF9	hsa04350	TGF-beta signaling pathway	5/5	3.91E-08
ZIC1	hsa04340	Hedgehog signaling pathway	3/3	1.09E-05
GRAP2	hsa04664	Fc epsilon RI signaling pathway	5/7	2.97E-07
ACTR2	hsa04810	Regulation of actin cytoskeleton	5/6	6.83E-06
PLCG2	hsa04660	T cell receptor signaling pathway	6/8	5.76E-08
CD2	hsa04660	T cell receptor signaling pathway	5/6	3.46E-07
TRPV4	hsa04670	Leukocyte transendothelial migration	3/10	0.00253
USP7	hsa04060	Cytokine-cytokine receptor interaction	11/15	9.44E-10
CCBP2	hsa04060	Cytokine-cytokine receptor interaction	4/6	0.000432
SLA	hsa04650	Natural killer cell mediated cytotoxicity	4/5	1.98E-05
CSK	hsa04650	Natural killer cell mediated cytotoxicity	10/15	1.32E-10
RGS16	hsa04080	Neuroactive ligand-receptor interaction	5/14	0.00191
STX1A	hsa04130	SNARE interactions in vesicular transport	5/6	6.01E-09
NAPA	hsa04130	SNARE interactions in vesicular transport	4/6	6.16E-07

Protein	GO ID	GO Term	Ratio	P value
KHDRBS1	GO:0005524	ATP binding	5/18	0.0151
GNAI1	GO:0003924	GTPase activity	4/4	9.45E-08
COL1A2	GO:0005587	collagen type IV	6/9	1.16E-16
MCM10	GO:0008270	zinc ion binding	10/26	0.00206
FN1	GO:0005509	calcium ion binding	6/18	0.000478
SAA1	GO:0005509	calcium ion binding	4/11	0.00307
ATP2B4	GO:0030955	potassium ion binding	4/16	1.52E-05
ACTR2	GO:0005885	Arp2/3 protein complex	6/6	5.13E-18
BLNK	GO:0005070	SH3/SH2 adaptor activity	4/15	4.29E-07
CD28	GO:0005070	SH3/SH2 adaptor activity	4/12	1.72E-07
DLG4	GO:0004385	guanylate kinase activity	4/11	1.13E-09
TIF1	GO:0003714	transcription corepressor activity	4/12	5.50E-06
GADD45G	GO:0030521	androgen receptor signaling pathway	4/8	1.22E-08

TNFRSF17	GO:0005031	tumor necrosis factor receptor activity	4/11	7.33E-10
TNFRSF8	GO:0005031	tumor necrosis factor receptor activity	4/14	1.38E-09
SOCS3	GO:0005159	insulin-like growth factor receptor binding	4/10	1.17E-09
PTPN1	GO:0005159	insulin-like growth factor receptor binding	4/14	4.29E-09
FAS	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	5/6	6.74E-10
CASP10	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	4/6	8.47E-08
MAP3K14	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	7/17	8.18E-11

Table 4 GO annotations that have been approved by GO annotation coordinator. The supporting literatures are listed in the 5th column.

Protein	GO ID	GO Term	Ontology	Supporting Literature	Ratio	P value
TRAF1	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	P	Yes. Literature support: PMID: 9774460.	4/10	7.74e-07
BLNK	GO:0005070	SH3/SH2 adaptor activity	F	Yes. This annotation is supported by experimental evidence from: PMID:9697839	4/15	4.29E-07
GAB2	GO:0005070	SH3/SH2 adaptor activity	F	Yes. Supported by evidence from PMID: 10068651	6/34	5.44e-09
CD28	GO:0005070	SH3/SH2 adaptor activity	F	Yes. Supported by evidence from PMID: 18295596	4/12	1.72E-07
GNAI1	GO:0003924	GTPase activity	F	Yes. Supported by evidence from PMID: 2834384	4/4	9.45E-08
CASP10	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	P	Yes. Supported by literature: PMID: 17822854	4/6	8.47E-08
IL4R	GO:0043560	insulin receptor substrate binding	F	Yes. Supported by PMID: 8124718	4/12	9.39e-10
MAP3K14	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	P	Yes. Supported by PMID: 9020361	7/17	8.18E-11
SIN3A	GO:0003714	transcription corepressor activity;	F	Yes. Supported by PMID:11259580	4/16	1.57e-05
NCOA2	GO:0030521	androgen receptor signaling pathway	P	Yes. Supported by PMID:17079484	5/12	7.33e-10
NR0B2	GO:0030521	androgen receptor signaling pathway;	P	Yes. Supported by PMID: 11705994	4/10	3.44e-08
NR0B2	GO:0050681	androgen receptor binding	F	Yes. Supported by PMID: 11705994	4/10	1.02e-08
TNFRSF19	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	P	Yes. Supported by PMID:18202551	4/19	1.00E-05
ACTR2	GO:0005885	Arp2/3 protein complex	C	Yes. Supported by PMID:9000076	6/6	5.13E-18
TNFRSF17	GO:0005031	tumor necrosis factor receptor activity	F	Yes. Supported by PMID:10903733	4/11	7.33E-10
TNFRSF8	GO:0005031	tumor necrosis factor receptor activity	F	Yes. Supported by PMID:9168896	4/14	1.38E-09
TNFRSF11A	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	P	Yes. Supported by PMID:9774460	4/17	6.39e-06

Table 5 Suggested functions for proteins in TGF- β signaling pathway. Functions shown in bold characters don't have any supporting literatures (as far as we know).

Protein	Possible roles in TGF- β signaling pathway
SKI	Transcriptional regulation [84-86]
IGSF1	Coreceptor for inhibin and/or activin
DAB2	Antagonist of STRAP
SOSTDC1, NOG	Interact with BMP receptors [87]
FMOD, CTCF, SLITL2	Regulating receptor binding of TGF- β proteins [92-94]; interact with each other

Appendix A

Formula derivation of Algorithm II

“ $P_2(X \text{ and } Y \text{ share } A \mid \kappa, N)$

$$\begin{aligned}
 &= [\prod_{i=1}^m P(Z_i \text{ connect both } X \text{ and } Y \mid \kappa, N)] P(\text{no other protein connect both } X \text{ and } Y \mid \kappa, N) \\
 &= \prod_{i \in A} \frac{\binom{N-1}{n_i-1} \binom{N-2}{n_i-2}}{\binom{N}{n_i} \binom{N-1}{n_i-1}} \prod_{i \notin A, i \in \Omega} (1 - \frac{\binom{N-1}{n_i-1} \binom{N-2}{n_i-2}}{\binom{N}{n_i} \binom{N-1}{n_i-1}}) \\
 &= \prod_{i \in A} \frac{n_i(n_i-1)}{N(N-1)} \prod_{i \notin A, i \in \Omega} (1 - \frac{n_i(n_i-1)}{N(N-1)})
 \end{aligned}$$

Let $M_1 = \prod_{i_1 \notin A_1, i_1 \in \Omega} (1 - \frac{n_{i_1}(n_{i_1}-1)}{N(N-1)})$ for any randomly picked protein X_1 and Y_1 ; let

$M_2 = \prod_{i_2 \notin A_2, i_2 \in \Omega} (1 - \frac{n_{i_2}(n_{i_2}-1)}{N(N-1)})$ for any randomly picked protein X_2 and Y_2 . Also let

$Q = \{i_1\} \cap \{i_2\}$. We can easily see that there are more than $N - 2m$ common elements in Q .

Therefore, we have the following inequality:

$$\begin{aligned}
 &\frac{\prod_{i \in A_1} n_i(n_i-1)}{\prod_{i \in A_2} n_i(n_i-1)} \leq \frac{M_1}{M_2} \leq \frac{\prod_{i \in A_1} n_i(n_i-1)}{\prod_{i \in A_2} n_i(n_i-1)} \\
 \Rightarrow &\prod_{i=1}^{2m} (1 - \frac{\max(n_i)(\max(n_i)-1)}{N(N-1)}) \leq \frac{M_1}{M_2} \leq \frac{1}{\prod_{i=1}^{2m} (1 - \frac{\max(n_i)(\max(n_i)-1)}{N(N-1)})} \quad \text{'' } \Psi \text{ [76]}
 \end{aligned}$$

For our human PPI network, $N = 7,362$, $1 \leq m \leq 45$, $1 \leq n_i \leq 157$. Simulations 1,000

times of the random PPI network and the random scale-free PPI network also showed the same results for N , m and n_i . Thus, we are able to get the following inequality:

$$\Rightarrow 0.96 \leq \frac{M_1}{M_2} \leq 1.04$$

$$\Rightarrow M_1 \approx M_2 \text{ and } \log(M_1) \approx \log(M_2)$$

Therefore, we can simply consider $\prod_{i \notin A, i \in \Omega} (1 - \frac{n_i(n_i-1)}{N(N-1)})$ as a constant, and we have

$$P_2 \propto \prod_{i \in A} \frac{n_i(n_i-1)}{N(N-1)}. \text{ For simplicity, we used } P_2 = \prod_{i \in A} \frac{n_i(n_i-1)}{N(N-1)} \text{ in our dissertation. [76]}$$

Mathematical expression for the probability that three proteins share m interacting partners

“To compute this probability, we count the number of distinct ways in which three proteins with n_1 , n_2 and n_3 interacting partners have m in common. We divide the whole set of partners of the three proteins into seven nonoverlapping groups: (i) m common protein partners that interact with proteins 1, 2 and 3; (ii) $n_{12} - m$ proteins that interact only with proteins 1 and 2; (iii) $n_{23} - m$ proteins that interact only with proteins 2 and 3; (iv) $n_{13} - m$ proteins that interact only with proteins 1 and 3; (v) $n_1 - n_{12} - n_{13} + m$ partners that interact only with protein 1; (vi) $n_2 - n_{12} - n_{23} + m$ partners that interact only with protein 2; and (vii) $n_3 - n_{13} - n_{23} + m$ partners that interact only with protein 3. We count the total number of distinct ways of assigning these seven groups to N proteins. This is given by:

$$\binom{N}{m} \binom{N-m}{n_{12}-m} \binom{N-n_{12}}{n_{23}-m} \binom{N-n_{12}-n_{23}+m}{n_{13}-m} \binom{N-n_{12}-n_{23}-n_{13}+2m}{n_1-n_{12}-n_{13}+m} \binom{N-n_1-n_{23}+m}{n_2-n_{12}-n_{23}+m} \binom{N-n_1-n_2+n_{12}}{n_3-n_{13}-n_{23}+m}$$

The total number of ways to randomly pick n_1 , n_2 and n_3 proteins from N proteins is given

$$\text{by: } \binom{N}{n_1} \binom{N}{n_2} \binom{N}{n_3}. \text{”}^\Psi \text{ [76]}$$

“Therefore, the probability that three proteins share m interacting partners is given as follows:

$$P_1(m | N, n_1, n_2, n_3) = \frac{\binom{N}{m} \binom{N-m}{n_{12}-m} \binom{N-n_{12}}{n_{23}-m} \binom{N-n_{12}-n_{23}+m}{n_{13}-m} \binom{N-n_{12}-n_{23}-n_{13}+2m}{n_1-n_{12}-n_{13}+m} \binom{N-n_{12}-n_{23}+m}{n_2-n_{12}-n_{23}+m} \binom{N-n_{12}-n_{23}+m}{n_3-n_{13}-n_{23}+m}}{\binom{N}{n_1} \binom{N}{n_2} \binom{N}{n_3}}$$

” Ψ [76]

Assessing the reliability of functional predictions for GO and KEGG annotations

“If a protein has at least one annotated significant partner, a list of annotation(s) from its partner(s) can be sorted by frequency. Suppose that annotations occurring n times or more will be assigned to this protein. For an annotated protein (based on GO and KEGG annotations), if an assigned annotation occurs among its known functions, we consider this a correct prediction. Assuming that GO and KEGG annotations are complete for those annotated proteins, we define a prediction precise rate as $\frac{\text{total number of correct predictions}}{\text{total number of predictions}}$, and as n varies, we have different precise rates (Fig. 12) which we used to estimate the FDRs of our functional predictions ($FDR = 1 - \text{precise rate}$), hence to assess the reliability of our functional predictions. From Fig. 12, we decided to use $n=2$ (for KEGG) and $n=4$ (for GO) as the thresholds of minimum frequency of functions shared by significant partners, which gave us relatively low FDRs (21% for KEGG and 30% for GO) without sacrificing too many predictions (we made 466 predictions for KEGG and 123 predictions for GO).” Ψ [76]

Appendix B

Table S1 The 4,233 significant protein pairs derived by our method (only top 1000 pairs are listed here for brevity). There are totally 1,729 human proteins and 25 nonhuman proteins. Protein pairs are ranked in terms of P_1 . Table S1 was partially reproduced from REF. 76.

	Protein A	Protein B	Ln(P_1)	Ln(P_2)	# of Common Neighbors
1	SMAD3	SMAD2	-157.6068404	-597.2243748	45
2	TUBB	TUBB2	-136.0437454	-289.8610141	21
3	PTPN11	PTPN6	-125.8551562	-424.3147331	33
4	BMPR1B	TGFBR1	-124.9465733	-599.6449534	38
5	CALM2	CALM3	-124.9367504	-291.8124921	21
6	MAPK1	MAPK3	-113.0905345	-384.8694846	27
7	CALM1	CALM3	-112.6375413	-302.5388215	22
8	IXL	MED9	-107.7585446	-279.4824444	18
9	PIK3R1	GRB2	-107.7069756	-489.2596327	42
10	CALM1	CALM2	-106.1716277	-306.6247978	22
11	RAC1	CDC42	-102.793931	-373.9517939	27
12	FYN	LCK	-102.6697375	-358.873044	30
13	HDAC1	HDAC2	-102.2579826	-358.5313968	28
14	CREBBP	EP300	-102.0552572	-420.1182758	32
15	GRB2	SHC1	-99.65932503	-469.6442329	39
16	SMAD2	SMAD4	-99.09164657	-460.3519272	34
17	DLG4	DLG1	-96.76638457	-356.084827	26
18	TRAF2	TRAF1	-95.58565584	-338.2088375	25
19	BCL2L1	BCL2	-94.54649001	-324.8992611	23
20	ACVR1	TGFBR1	-89.65577292	-467.2366865	30
21	CPNE1	CPNE4	-87.91061912	-181.4118777	13
22	MDFI	KRTAP4-12	-87.88978999	-438.6963486	30
23	SMAD3	SMAD4	-85.77343234	-435.3492838	33
24	ACVR1	BMPR1B	-84.52904218	-343.6074392	23
25	COL4A3	COL4A2	-82.20398173	-167.9671211	13
26	PTPN11	GRB2	-81.63255068	-383.1932153	32
27	COL4A2	COL4A4	-80.7276668	-157.9221044	12
28	PTK2	PTK2B	-80.57310834	-240.1404574	22
29	SRC	FYN	-79.46202476	-355.9045821	30
30	COL4A3	COL4A4	-79.33974045	-157.9221044	12
31	FYN	LYN	-79.00597291	-320.8620938	26
32	RELA	NFKB1	-78.83597444	-279.429444	23
33	TRAF3	TRAF1	-78.157227	-213.0497868	16
34	CSNK2A1	CSNK2A2	-77.73600946	-181.4968663	15
35	JAK1	JAK2	-77.22522398	-233.6962148	20

36	PTPN11	PIK3R1	-74.97662001	-296.7887929	26
37	COL4A2	COL4A1	-74.60859428	-151.9508426	12
38	YWHAZ	YWHAE	-74.52029133	-245.9556029	19
39	JUN	FOS	-74.00815036	-250.4496582	21
40	RBL2	RBL1	-74.0078444	-175.0053511	14
41	COL4A5	COL4A6	-73.93034147	-123.3558506	10
42	NR3C1	AR	-73.72738905	-289.2367706	25
43	NR3C1	ESR1	-73.72738905	-292.7791161	25
44	SUMO1	UBE2I	-73.5062065	-258.2503183	20
45	COL4A3	COL4A1	-73.22107635	-151.9508426	12
46	DLG4	DLG3	-73.17886306	-232.2327749	17
47	HDAC1	SIN3A	-72.47932222	-266.6637009	21
48	IRS2	IRS1	-71.88001114	-151.8096496	14
49	COL4A1	COL4A4	-70.72591947	-141.9058259	11
50	IKBKB	CHUK	-69.96305423	-162.9136822	14
51	GADD45B	GADD45G	-69.56739182	-130.2946982	11
52	GRB2	PLCG1	-69.39340099	-337.6520463	28
53	JUP	CTNNB1	-69.25868254	-254.9584259	19
54	PLSCR1	KRTAP4-12	-68.91001314	-329.6901377	23
55	PTPN11	SHC1	-68.90609068	-272.90666	24
56	PIK3R1	SHC1	-68.82244231	-294.9077988	26
57	ESR1	ESR2	-68.16817601	-221.2853544	17
58	DLG2	DLG3	-67.29158995	-151.719371	12
59	TBP	GTF2B	-66.86509738	-191.3382891	16
60	COL4A2	COL4A5	-65.9161763	-123.3558506	10
61	COL4A2	COL4A6	-65.9161763	-123.3558506	10
62	COL4A1	COL4A5	-65.9161763	-123.3558506	10
63	COL4A1	COL4A6	-65.9161763	-123.3558506	10
64	SRC	GRB2	-65.68304701	-369.6892177	31
65	AR	ESR1	-65.36120402	-294.273143	24
66	COL4A3	COL4A5	-64.93398687	-123.3558506	10
67	COL4A3	COL4A6	-64.93398687	-123.3558506	10
68	RARA	RXRA	-64.11053408	-215.3482183	17
69	TRAF3	TRAF2	-64.01676383	-260.0936955	19
70	DLG4	DLG2	-63.98188409	-202.8927639	15
71	COL1A1	COL1A2	-63.24777014	-151.5867474	12
72	TGFBR1	SMURF1	-62.87519103	-423.9212755	27
73	ORC2L	MCM3	-62.62470809	-143.1193491	11
74	GRB2	FYN	-62.50403476	-353.978848	30
75	FLJ22494	LNK	-61.76207239	-191.962379	15
76	TNFSF13	TWE-PRIL	-60.6242878	-105.0219606	8
77	PSEN1	PSEN2	-59.82978047	-172.4340654	12
78	SHANK2	SHANK1	-59.63303509	-125.6225476	9

79	COL4A4	COL4A5	-59.6305881	-113.3108339	9
80	COL4A4	COL4A6	-59.6305881	-113.3108339	9
81	GRB2	CRKL	-59.62441316	-240.3660842	21
82	NIF3L1	LNK	-59.17599865	-179.0708543	15
83	KRT15	DIPA	-59.0531779	-283.1457656	22
84	HDAC4	HDAC5	-58.97406646	-131.1471165	11
85	NCOR1	NCOR2	-58.96941997	-160.9803545	14
86	GRB2	EGFR	-58.92264739	-339.0230987	29
87	DDIT3	CEBPG	-57.53762273	-144.3332532	11
88	NCOA1	NCOA2	-57.49927497	-148.7111631	13
89	TGFB1	TGFB2	-57.37723082	-148.0591844	11
90	TRAF2	TRAF5	-57.35284724	-187.591848	14
91	PIK3R1	PLCG1	-57.07816398	-234.8728064	21
92	PTPN6	VAV1	-56.80604805	-182.1499538	17
93	PTK2	PXN	-56.6283112	-199.3110295	17
94	YWHAE	YWHAB	-56.36045703	-168.5263529	14
95	NCOA1	PPARBP	-56.35164632	-147.0223305	13
96	TCEB2	TCEB1	-56.22888739	-116.397614	8
97	RAC1	RHOA	-55.93112463	-243.8942209	17
98	STX1A	STX4A	-55.66368681	-170.1849433	12
99	IGF2	IGF1	-55.36089149	-129.1709752	9
100	PAG1	CBL	-55.22086551	-119.852138	12
101	YWHAZ	YWHAB	-54.86337875	-193.8031874	15
102	SMN2	DDX20	-54.80372312	-123.8406098	9
103	GNAI2	GNAI3	-54.55428285	-171.4575127	12
104	MCM7	MCM2	-54.4577214	-155.7905121	12
105	ESR1	RXRA	-54.42433382	-233.5142756	19
106	SYK	VAV1	-54.32372654	-168.6185713	15
107	SRC	PIK3R1	-54.26967037	-271.9053986	23
108	GNAI1	GNAO1	-54.13038371	-171.8454402	12
109	RBBP4	HDAC1	-53.99133316	-149.8446323	13
110	RIBC2	TU3A	-53.78938962	-105.218086	10
111	JAK2	PTPN11	-53.6548795	-198.1082041	18
112	JUN	JUND	-53.56563793	-156.5032226	13
113	ERBB2	EGFR	-53.37266001	-194.4930757	17
114	IL8RB	IL8RA	-53.31826358	-125.3057676	8
115	SMN2	GEMIN5	-53.31826358	-107.4878591	8
116	DIPA	KIAA0980	-53.21491811	-248.0511319	18
117	GRB2	LCK	-53.08345707	-258.9204744	22
118	K-ALPHA-1	TUBA2	-53.0093322	-94.17201147	8
119	PTPN6	SHC1	-52.93501896	-209.5488701	19
120	CPNE2	CPNE1	-52.61692073	-105.5499929	8
121	CPNE2	CPNE4	-52.61692073	-105.5499929	8

122	INSR	IGF1R	-52.60883246	-152.7198638	13
123	CCR5	CCBP2	-52.53526685	-146.9777368	10
124	CDC27	CDC16	-52.21869964	-96.02430266	8
125	EGFR	INSR	-52.00018901	-199.9203353	17
126	NCOA1	RELA	-51.4879004	-181.7950369	17
127	TGFB3	TGFB2	-51.2665869	-107.9942436	8
128	ERBB2	ERBB3	-51.14620615	-124.7513678	11
129	KIAA0408	RIBC2	-51.11713523	-107.8027107	10
130	PLCG1	SHC1	-51.09921926	-217.1652406	19
131	NCOR1	SIN3A	-51.06007248	-162.6756772	13
132	TRAF3	TRAF5	-51.05966346	-135.8949254	10
133	PSMA2	PLK1	-50.92646954	-157.7437897	10
134	SYK	ZAP70	-50.85333953	-147.3483587	13
135	NCOA1	NCOA3	-50.47714922	-134.0379987	12
136	JUND	JUNB	-50.379487	-106.6247075	9
137	LRP2	LRP1	-50.17507279	-184.9652614	14
138	TGFB1	TGFB3	-50.15242977	-120.7524255	9
139	PLSCR1	MDFI	-50.09653138	-284.4524951	20
140	NOTCH1	NOTCH2	-50.05417387	-132.2017988	9
141	ACTA1	ACTA2	-49.92697889	-121.5887173	8
142	GEMIN5	DDX20	-49.65742117	-106.2810332	8
143	PPFIA1	PPFIA2	-49.65089455	-111.7036483	8
144	NIF3L1	FLJ22494	-49.41538383	-153.1918347	12
145	RB1	RBL2	-49.35799053	-159.7627593	13
146	TGFBR1	SMAD2	-49.22991515	-313.6395562	23
147	PAG1	PTK2B	-49.13577967	-106.1032354	11
148	DLG1	DLG2	-49.11758986	-149.7428191	11
149	SYK	GRB2	-48.96440679	-216.4612199	19
150	GRB2	VAV1	-48.95379374	-221.5131424	20
151	GNAI1	GNAI3	-48.92562467	-156.5021062	11
152	STAT3	STAT5A	-48.87833789	-143.7916207	13
153	GRIN2B	GRIN2A	-48.70997026	-103.2496301	9
154	BCAR1	PXN	-48.65494906	-144.6837174	13
155	SNAP23	SNAP25	-48.54485032	-136.1281066	10
156	DCN	MMP9	-48.4240181	-121.169453	10
157	LSM2	LSM6	-48.35487606	-97.64839637	8
158	RB1	RBL1	-48.21283721	-171.297408	14
159	MRAS	RAP2A	-48.13695391	-92.11190456	7
160	SYK	CBL	-48.09739775	-146.6243912	14
161	RXRA	PPARG	-48.03681755	-154.551442	12
162	DCN	TGFBI	-47.93369554	-95.13917142	8
163	GNAI2	GNAO1	-47.80939626	-154.1143633	11
164	MCM3	MCM7	-47.73212946	-132.8259441	10

165	DCN	C1QR1	-47.63175485	-113.3746534	9
166	CDK4	CDK6	-47.39346181	-117.4577271	9
167	GADD45A	GADD45G	-47.35497338	-93.59561967	8
168	PPARA	PPARG	-47.31448046	-127.2745442	10
169	LSM5	LSM2	-47.25803203	-97.64839637	8
170	NUP214	NUP153	-47.25694418	-100.9688921	8
171	RIPK1	TRADD	-47.2113429	-117.5311184	10
172	BIRC3	BIRC2	-47.12603232	-83.77493325	7
173	BATF	FOSL2	-47.12127462	-83.25986031	7
174	GNA12	GNA13	-47.01864688	-138.531382	10
175	SYK	PIK3R1	-47.0063129	-166.6326524	16
176	PLAT	PLG	-46.96960868	-133.6613215	10
177	CDH1	CTNNB1	-46.90858967	-179.2859447	14
178	GRB2	NCK1	-46.90800812	-231.1999571	19
179	GRB2	CBL	-46.9007666	-229.5719977	20
180	ORC2L	MCM7	-46.84619219	-132.9907331	10
181	Rps27a	Uhmk1	-46.84323227	-54.41756914	6
182	GRB2	ZAP70	-46.84008345	-188.8756016	17
183	FLJ32855	LNK	-46.75233618	-123.3224073	11
184	ABLIM1	TU3A	-46.74843035	-85.3823523	8
185	PPFIA2	PPFIA3	-46.6321971	-98.54875399	7
186	TGFBR1	SMAD4	-46.56941628	-366.3543753	25
187	RXRA	THRB	-46.41308504	-165.5102055	12
188	ACVR2A	ACVR2B	-46.41127861	-110.1503806	8
189	LSM3	LSM5	-46.40910277	-97.80254705	8
190	PDGFRB	EGFR	-46.40310111	-162.0161989	15
191	BCAR1	NEDD9	-46.24876461	-116.599624	11
192	FXR2	LNK	-46.24552497	-166.307615	14
193	GNAI2	GNAI1	-46.21961531	-166.406671	12
194	TNPO1	RANBP5	-46.05025162	-108.2216242	8
195	JUN	CEBPB	-46.03232224	-163.504687	15
196	LSM4	LSM5	-45.64641909	-97.9456479	8
197	LSM3	LSM2	-45.45440799	-97.96960114	8
198	TGFBR3	ENG	-45.42849633	-82.7769275	7
199	PEPP-2	RPMS	-45.36250418	-157.0131503	12
200	NCOA1	NRIP1	-45.25772539	-131.2515547	12
201	NDP52	DIPA	-45.24092977	-236.7683123	18
202	NCOA2	PPARBP	-45.10586824	-101.2283063	9
203	SIT1	PAG1	-45.03108659	-68.7234239	7
204	ORC1L	MCM3	-44.95708047	-105.1465529	8
205	HABP2	SERPINE2	-44.89732213	-76.79095267	6
206	PTPN11	CRKL	-44.78490422	-156.8651716	14
207	LSM4	LSM2	-44.69186048	-98.11270198	8

208	HRAS	RAP1A	-44.65606489	-128.1956568	10
209	JUN	JUNB	-44.63141161	-129.3848887	11
210	TGFBR1	TGFBR2	-44.56273839	-174.3312409	14
211	PIK3R1	PIK3R2	-44.49201549	-114.1243101	11
212	PIK3R1	FYN	-44.2701179	-232.9912143	20
213	CCR2	CCR5	-44.14390473	-110.3012698	8
214	AP2M1	AP1M1	-43.94719372	-117.6375598	9
215	CDC42	RHOA	-43.85316884	-200.1685715	14
216	LSM3	LSM4	-43.84660373	-98.26685266	8
217	LSM3	LSM7	-43.84660373	-99.12430289	8
218	MMP9	MATN2	-43.76638644	-97.93373905	8
219	EFEMP2	KRTAP4-12	-43.76009548	-178.2902063	13
220	VAV1	ZAP70	-43.73581462	-136.522431	12
221	STAT5B	STAT5A	-43.61607016	-100.7888124	9
222	Rasd2	Rps27a	-43.509397	-54.41756914	6
223	Rasd2	Fbxo3	-43.509397	-54.65430411	6
224	Rasd2	Sqstm1	-43.509397	-55.14657866	6
225	Rasd2	Map2k3	-43.509397	-53.1824134	6
226	Rasd2	Uhmk1	-43.509397	-54.41756914	6
227	Rasd2	Rhod	-43.509397	-55.7396668	6
228	DCN	FBLN2	-43.42110213	-112.3362812	9
229	PIK3R1	LYN	-43.25809006	-199.2928279	17
230	PTPN11	SYK	-43.00462897	-149.9651422	14
231	APEX2	TU3A	-42.95178206	-71.96030807	7
232	VLDLR	LRP8	-42.9507326	-84.28501423	6
233	RPP25	POP5	-42.9507326	-87.80362146	6
234	RPP25	RPP30	-42.9507326	-87.80362146	6
235	POP5	RPP30	-42.9507326	-87.80362146	6
236	DCN	MATN2	-42.83346747	-97.93373905	8
237	ACVR1	SMAD4	-42.72020889	-240.1710342	17
238	GAB2	SHC1	-42.6974178	-112.0751914	11
239	RXRA	RXRB	-42.68817061	-130.8295077	10
240	NCOA1	RXRA	-42.63524817	-170.8421429	14
241	HDAC3	HDAC2	-42.61819069	-158.2345068	13
242	TBP	TAF1	-42.59805866	-137.9181026	12
243	OSM	HABP2	-42.40996916	-76.79095267	6
244	MAPK8IP1	MAPK8IP2	-42.33772675	-83.34262178	7
245	BMP2	BMP7	-42.33582328	-89.46033915	7
246	GADD45B	GADD45A	-42.33582328	-83.22242956	7
247	GNAI1	GNAQ	-42.31927231	-161.3585908	11
248	SNX1	SNX2	-42.20077893	-81.12426297	7
249	ERCC3	GTF2H5	-42.07141913	-100.6980051	7
250	MAP3K14	TRAF3IP2	-42.07141913	-75.87139392	7

251	MYOD1	TCF3	-41.94784723	-134.3091363	11
252	CASP8	CFLAR	-41.93916853	-113.6687053	9
253	PIK3R1	LCK	-41.91302113	-184.8567086	16
254	JAK1	JAK3	-41.82955924	-117.902316	10
255	DDIT3	BATF	-41.7367615	-99.22834816	8
256	C1QR1	MMP9	-41.72314951	-100.4494174	8
257	KRT15	USHBP1	-41.69024654	-158.4353022	12
258	DLG1	DLG3	-41.65494862	-135.4901292	10
259	E2F2	E2F3	-41.56375877	-78.20255709	6
260	BMP7	GDF5	-41.49286277	-71.03511552	6
261	SIP1	SFRS2IP	-41.49286277	-77.52434331	6
262	FN1	DCN	-41.46084085	-135.7774582	11
263	PPFIA1	PPFIA3	-41.44213051	-99.6191954	7
264	DIPA	HOOK2	-41.41509206	-153.9566021	12
265	SMURF2	SMURF1	-41.39619496	-220.2378639	14
266	SFRS1	SFRS2	-41.38096387	-112.9666444	8
267	VAV1	CBL	-41.35790865	-138.5413097	13
268	DNM1	SYNJ1	-41.34420166	-86.97828944	7
269	SNRPE	SNRPD2	-41.32490269	-96.06384734	7
270	LSM5	LSM6	-41.32490269	-84.37378467	7
271	ARF1	ARF6	-41.30877327	-111.7981418	8
272	CASP8	FADD	-41.29842529	-131.406081	10
273	SRC	PLCG1	-41.21726699	-189.9434317	17
274	LYN	LCK	-41.18164688	-162.2855551	14
275	SRC	LCK	-41.1732206	-183.5990962	16
276	PTK2B	PXN	-41.01939063	-145.9634317	13
277	RAP1A	RAP2A	-40.8660863	-92.39958664	7
278	DIPA	USHBP1	-40.78388281	-195.9356015	15
279	HABP2	MATN2	-40.70766738	-76.79095267	6
280	MYOD1	MYOG	-40.62940142	-91.27315733	8
281	PIK3R1	CRKL	-40.57375767	-156.0399911	14
282	SNTA1	SNTB2	-40.55130524	-103.4834277	7
283	C1QR1	MATN2	-40.55130524	-88.12201829	7
284	GHR	EPOR	-40.54843776	-92.54024858	9
285	RPP25	RPP38	-40.46609791	-88.34261796	6
286	POP5	RPP38	-40.46609791	-88.34261796	6
287	RPP30	RPP38	-40.46609791	-90.28852811	6
288	OSM	SERPINE2	-40.46446692	-76.79095267	6
289	TGFB1	MMP9	-40.46367871	-83.64468148	7
290	PTPN6	GRB2	-40.37656558	-215.1490247	19
291	HDAC1	HDAC3	-40.29652281	-184.5879805	15
292	BMPR1A	BMPR2	-40.23924629	-93.96468397	7
293	PXN	NEDD9	-40.2361581	-113.3121523	11

294	NCK2	NCK1	-40.20086252	-134.6177707	11
295	RGS16	RIC8	-40.17773637	-67.38652981	6
296	GRB2	LCP2	-40.17518002	-143.5999216	13
297	ESR1	PPARA	-40.15293346	-139.6695433	12
298	SYK	SHC1	-40.09195087	-153.7464221	14
299	PIK3R1	SOCS1	-40.06668637	-129.5450323	12
300	SMAD4	SMURF2	-40.04433175	-210.6740511	15
301	PIK3R1	PTPN6	-39.93165164	-180.291024	16
302	VAV1	FYN	-39.88833982	-164.0926823	15
303	LSM3	LSM6	-39.85775044	-84.69498943	7
304	SNX6	SNX2	-39.85775044	-83.82843161	7
305	CCND3	CCND2	-39.84162117	-90.72353943	7
306	SRC	PTK2	-39.82567764	-172.8450731	16
307	GRB2	LYN	-39.74804755	-222.7113672	19
308	NRBF2	CGI-63	-39.73158422	-55.92972196	5
309	MNAT1	CCNH	-39.67665305	-85.02417441	7
310	TRAF5	TRAF6	-39.59479632	-117.835573	9
311	SHB	SHC1	-39.57382389	-121.6037075	11
312	RASIP1	RALGDS	-39.39548099	-76.14064973	6
313	C1QR1	HABP2	-39.38977137	-76.79095267	6
314	PDGFRB	PDGFRA	-39.38885079	-106.2921215	9
315	CRK	CRKL	-39.36474033	-131.4257445	11
316	LSM4	LSM6	-39.22873411	-84.83809028	7
317	LSM7	LSM6	-39.22873411	-84.83809028	7
318	TU3A	ZNF638	-39.22873411	-73.58197839	7
319	NCOA1	TRIP4	-39.08813372	-88.31539245	9
320	VAV1	LYN	-38.96454176	-145.5439083	13
321	VAV1	ITK	-38.88788498	-95.53901474	9
322	PIK3R1	CBL	-38.87133268	-159.2201208	15
323	CHUK	IKBKG	-38.86245192	-109.922838	9
324	MPP6	MPHOSPH6	-38.82933948	-89.39471023	6
325	IL2RB	IL2RG	-38.78349477	-100.1553455	8
326	SERPINE2	MATN2	-38.76162185	-76.79095267	6
327	LSM1	LSM7	-38.74995345	-102.4421187	8
328	MYOD1	MYF5	-38.74757995	-86.57241694	7
329	NCOR1	HDAC1	-38.6291783	-153.4759068	13
330	JUN	ATF2	-38.55288725	-122.141834	11
331	SOCS1	SOCS3	-38.48629283	-92.27214114	8
332	PTPN6	CRKL	-38.46424811	-135.5828818	12
333	RBBP4	HDAC2	-38.45389791	-103.3187214	9
334	PTPN11	SOCS3	-38.35768563	-111.7661621	10
335	LSM5	LSM7	-38.35721042	-85.02041183	7
336	JAK2	EPOR	-38.35610561	-111.6535173	11

337	BCAR1	CBL	-38.34818087	-113.6672268	11
338	ACVR1B	ACVR2B	-38.21194941	-95.01677625	7
339	JAK2	GHR	-38.19468926	-101.7545324	10
340	RB1	HDAC1	-38.17580263	-202.9810522	17
341	FN1	C1QR1	-38.15340304	-113.3746534	9
342	KRT19	KRT15	-38.10828079	-120.1304474	9
343	MCM3	MCM2	-38.06647217	-104.5689186	8
344	IGF1R	IRS1	-38.05870836	-108.2068021	10
345	PTPN12	PTK2	-38.03029447	-97.40838181	9
346	ARCN1	COPE	-37.93982475	-67.78837801	5
347	AFTIPHILIN	APIGBP1	-37.93982475	-68.10316201	5
348	Rab38	Map2k3	-37.93982475	-44.04891418	5
349	Rhobtb1	Rps27a	-37.93982475	-45.96449121	5
350	Rhobtb1	Fbxo3	-37.93982475	-45.96449121	5
351	Rhobtb1	Uhmk1	-37.93982475	-45.96449121	5
352	Rhoj	Rhod	-37.93982475	-45.28406991	5
353	Rps27a	Rhoj	-37.93914536	-45.28406991	5
354	Uhmk1	Rhoj	-37.93914536	-45.28406991	5
355	ZAP70	CBL	-37.91376014	-119.9245798	11
356	JAK2	GRB2	-37.89879538	-199.208469	18
357	SMAD2	SMURF1	-37.84663713	-245.4812714	16
358	VAV1	SHC1	-37.80246512	-146.7413383	14
359	YWHAH	YWHAB	-37.67115933	-124.2123049	10
360	TRAF1	TRAF5	-37.66196924	-104.3042011	8
361	GHR	IL2RB	-37.6535237	-81.44727819	8
362	FBLN2	MATN2	-37.62904187	-86.60267342	7
363	CCR3	CCBP2	-37.58415763	-102.8772919	7
364	LSM7	LSM2	-37.58225339	-85.18746592	7
365	CBX5	CBX3	-37.45101917	-96.06384734	7
366	MYF5	TCF4	-37.44970757	-71.24928561	6
367	SNTA1	SNTB1	-37.44481328	-84.86445954	6
368	C1QR1	SERPINE2	-37.44481328	-76.79095267	6
369	ORC2L	MCM2	-37.43143898	-102.846638	8
370	TGFBI	MATN2	-37.37519209	-71.31728241	6
371	MEP50	SMN2	-37.37519209	-81.31323519	6
372	COPG	COPG2	-37.37274511	-81.61758432	6
373	SNX1	SNX4	-37.37274511	-68.86960731	6
374	SNX2	SNX4	-37.37274511	-67.84650257	6
375	CDC2	CCNB1	-37.36662133	-104.0034713	9
376	GRB2	SOCS1	-37.32588335	-151.7299828	13
377	GAB1	SHC1	-37.32014814	-103.9326095	10
378	KIT	GAB2	-37.23916225	-78.5627139	8
379	ORC4L	MCM7	-37.19693349	-89.75000862	7

380	DNMT3B	DNMT3A	-37.06367795	-75.51782783	6
381	YWHAH	YWHAE	-37.01500475	-124.0001303	10
382	NCOA1	ESR1	-37.0092193	-161.0801637	14
383	FBLN2	HABP2	-37.00446152	-76.79095267	6
384	WAS	CBL	-36.98082325	-103.8072983	10
385	CCR5	CCR1	-36.95819592	-101.3368469	7
386	SYK	LYN	-36.79947833	-144.6458661	12
387	PTPN11	LYN	-36.79710806	-165.9931712	14
388	PTPN11	VAV1	-36.71246718	-137.8453417	13
389	CORT	SST	-36.68570291	-77.67344698	5
390	GAB3	CSF1R	-36.68231702	-62.32889307	6
391	HLA-A	HLA-B	-36.67905416	-81.31974988	6
392	BATF	ATF3	-36.67905416	-69.65580396	6
393	GNAI3	GNAO1	-36.66530744	-113.2407623	8
394	THRAP4	PPARBP	-36.63476315	-65.60952835	6
395	KIT	EPOR	-36.63165599	-95.0663538	9
396	XRCC5	XRCC6	-36.59489555	-94.9311125	8
397	IL6ST	LIFR	-36.59474601	-95.92593859	7
398	BCL3	TRIP4	-36.47599033	-71.03108651	7
399	JAK1	IL2RB	-36.39720697	-94.26662965	9
400	JAK2	PTPN6	-36.33127453	-145.5425149	13
401	PDGFRB	INSR	-36.33042752	-118.0134758	10
402	HIST3H3	HIST2H2BE	-36.31673665	-97.46379232	7
403	JAK1	VAV1	-36.30137608	-119.24112	11
404	OSM	MATN2	-36.27644439	-76.79095267	6
405	MMP9	HABP2	-36.26666352	-76.79095267	6
406	GAB1	GAB2	-36.26267605	-72.22069688	7
407	LSM4	LSM7	-36.26267605	-85.48471744	7
408	TRAF2	TRAF6	-36.20107542	-197.1283011	15
409	PTK2B	CRKL	-36.19897285	-110.8158048	11
410	PTPN11	SRC	-36.196943	-173.2372823	16
411	IGSF1	INHBC	-36.14752183	-61.74378984	5
412	ARPC3	ARPC2	-36.14752183	-71.70218514	5
413	ARPC3	ARPC4	-36.14752183	-71.70218514	5
414	ARPC2	ARPC4	-36.14752183	-71.70218514	5
415	Rps27a	Fbxo3	-36.14752183	-45.96449121	5
416	Rps27a	Sqstm1	-36.14752183	-46.45676576	5
417	Rps27a	Map2k3	-36.14752183	-44.4926005	5
418	Rps27a	Rhod	-36.14752183	-45.28406991	5
419	Fbxo3	Sqstm1	-36.14752183	-46.69350074	5
420	Fbxo3	Map2k3	-36.14752183	-44.72933548	5
421	Fbxo3	Uhmk1	-36.14752183	-45.96449121	5
422	Sqstm1	Map2k3	-36.14752183	-45.22161003	5

423	Sqstm1	Uhmk1	-36.14752183	-46.45676576	5
424	Map2k3	Uhmk1	-36.14752183	-44.4926005	5
425	Uhmk1	Rhod	-36.14752183	-45.28406991	5
426	FYN	CRKL	-36.09941856	-144.5021595	13
427	PTK2	PDGFRB	-36.09474132	-104.3050692	11
428	TGFBI	C1QR1	-36.06354996	-73.83296072	6
429	SYK	PLCG1	-35.94622955	-125.7579613	12
430	TRAF2	DIPA	-35.90885608	-283.1926128	22
431	GTF2E2	GTF2E1	-35.7842048	-85.02246909	7
432	ESR1	HNF4A	-35.76439277	-119.2556816	10
433	IKBKB	RIPK1	-35.73511568	-90.29451606	8
434	CDC2	CDK2	-35.72524914	-125.8585826	10
435	PARD6G	PARD6A	-35.70623253	-57.17807582	5
436	TNFRSF19	TNFRSF11A	-35.70623253	-53.47486974	5
437	BAK1	PMAIP1	-35.70419407	-60.61054889	5
438	TFDP1	TFDP2	-35.70419407	-66.5474495	5
439	Rasd2	Rab38	-35.70419407	-44.04891418	5
440	Rasd2	Rhobtb1	-35.70419407	-45.96449121	5
441	Rasd2	Rhoj	-35.70419407	-45.28406991	5
442	Rasd2	Smad3	-35.70419407	-45.86548616	5
443	Rasd2	Rras2	-35.70419407	-47.3351771	5
444	Rasd2	Stat1	-35.70419407	-45.92794604	5
445	COL4A2	COL1A2	-35.68840138	-82.69612751	7
446	VIL2	MSN	-35.68291003	-106.6973835	8
447	PTPRC	PAG1	-35.68291003	-79.63800386	8
448	DCN	HABP2	-35.61010368	-76.79095267	6
449	PTPN6	LYN	-35.58471818	-149.3415021	13
450	RELA	JUN	-35.582208	-156.1129099	15
451	INHBA	INHBB	-35.58125809	-78.6355371	6
452	GRB2	INSR	-35.57112982	-171.0766973	15
453	PIK3R1	PTK2B	-35.53017551	-135.8205498	14
454	RAD9A	HUS1	-35.49931139	-75.89204162	6
455	JAK1	PTPN11	-35.48636942	-131.213017	12
456	CD22	CBL	-35.43098247	-77.42968386	8
457	CCND1	CCND2	-35.41841277	-90.72353943	7
458	YWHAH	SMARCE1	-35.41841277	-72.80221266	7
459	NFKB1	NFKB2	-35.38845238	-101.7668362	8
460	DNM1	WASL	-35.29172923	-85.57242498	7
461	IKBKB	IKBKG	-35.28503947	-97.95715952	8
462	SMAD4	SMURF1	-35.19121876	-263.1069878	17
463	RXRA	PPARA	-35.1706672	-124.9050345	10
464	LYN	HCK	-35.13675288	-127.3651279	10
465	CASP3	CASP7	-35.07561256	-107.0499034	8

466	FBLN2	SERPINE2	-35.06018402	-76.79095267	6
467	GRB2	GRAP2	-35.03829526	-159.0576748	13
468	RARA	ESR1	-35.0124807	-144.9254486	12
469	MEP50	DDX20	-34.99365295	-81.31323519	6
470	GAB2	GAB3	-34.98794177	-55.64576093	6
471	ACVR1B	INHBB	-34.96181113	-77.19315327	6
472	C1QR1	OSM	-34.96181113	-76.79095267	6
473	DAZAP2	RNF11	-34.9335871	-126.1961635	9
474	CCL8	CCL7	-34.89489481	-63.50804216	5
475	ARHGEF12	ARHGEF11	-34.89489481	-63.18994706	5
476	LILRB2	LILRB1	-34.89421532	-68.16645686	5
477	GGA3	GGA1	-34.89421532	-68.53818941	5
478	TOB1	ZNF8	-34.89258419	-48.97444407	5
479	USP7	TNFRSF19	-34.89258419	-53.47486974	5
480	PIK3R1	ZAP70	-34.77856073	-128.0593263	12
481	RAF1	BRAF	-34.72406764	-92.8809925	8
482	KIAA0980	HOOK2	-34.71540429	-102.1030953	8
483	JAK2	SHC1	-34.6932362	-154.7227013	14
484	PECAM1	TRPV4	-34.68001391	-60.94693657	6
485	BLNK	CBLB	-34.66747603	-73.09234363	7
486	TF	IGF1	-34.66741533	-88.94045496	6
487	KIAA0408	TU3A	-34.66461869	-72.34650036	7
488	PTPN6	PLCG1	-34.65915131	-139.2295527	13
489	CSK	FYN	-34.57068818	-126.9290328	11
490	PLCG1	FYN	-34.56321283	-168.0280133	15
491	ZBTB16	HDAC3	-34.52712025	-127.3560678	11
492	RAP1A	MRAS	-34.51739407	-78.32921827	6
493	SMAD2	SMAD1	-34.39386332	-145.4455218	12
494	FN1	COL1A1	-34.38279739	-131.4316458	10
495	SIN3A	HDAC9	-34.38188955	-90.63924088	8
496	RB1	SP1	-34.38086977	-160.3971522	14
497	SLC9A3R1	SLC9A3R2	-34.3603185	-111.7973943	8
498	SRC	RASA1	-34.34859508	-143.5616914	13
499	MMP9	SERPINE2	-34.32265839	-76.79095267	6
500	BGN	DCN	-34.31263744	-85.13127648	7
501	CCR5	CCR3	-34.30406426	-102.8772919	7
502	PECAM1	CBL	-34.29840109	-88.24072021	9
503	POU2F1	POU2F2	-34.22793856	-69.96311341	6
504	BLNK	LCP2	-34.20903362	-78.37354623	7
505	TBL1X	TBL1XR1	-34.19875724	-62.58438979	5
506	CDC5L	CCDC5	-34.19875724	-60.05789863	5
507	RAC1	RAC2	-34.11034585	-112.3971561	8
508	S100A1	S100B	-34.0926579	-102.3084425	7

509	ORC4L	ORC2L	-34.0888758	-76.25991841	6
510	EGFR	ERBB3	-34.07603135	-121.7110921	10
511	CCND1	CCND3	-34.06321245	-100.0779632	8
512	FOS	CEBPB	-34.05116493	-114.8295819	10
513	CCR3	CCR1	-34.04647284	-89.34592007	6
514	PTPN11	SOCS1	-34.01472826	-103.2900617	10
515	MLLT4	RASIP1	-33.98595884	-76.14064973	6
516	NEDD4	NEDD4L	-33.98550548	-83.6034998	6
517	ACVR1	ACVR2A	-33.95679771	-109.0517684	8
518	LYN	CRKL	-33.9558387	-125.7169321	11
519	MAPK1	MAPK14	-33.92044979	-160.6026348	11
520	ORC4L	ORC5L	-33.91284246	-61.96298679	5
521	HUS1	RAD1	-33.91284246	-71.1701573	5
522	SCNN1B	SCNN1G	-33.91284246	-66.6742012	5
523	MAP2K3	MAP2K6	-33.91284246	-67.35821776	5
524	ZNFN1A1	ZNFN1A4	-33.89055406	-75.36823352	6
525	MBD2	MBD3	-33.87814276	-85.45156742	6
526	NUP153	RANBP2	-33.87814276	-73.80201592	6
527	FXR2	FLJ32855	-33.87642856	-98.99838339	9
528	GRB2	PTK2B	-33.83177756	-160.7430491	16
529	GHR	IGF1R	-33.82987341	-88.56633783	8
530	HDAC4	HDAC9	-33.77457789	-81.91910026	7
531	YWHAZ	YWHAQ	-33.69046478	-129.652023	9
532	DCN	SERPINE2	-33.66637099	-76.79095267	6
533	NCOA1	JUN	-33.65774353	-140.5353335	13
534	CREBBP	JUN	-33.63178649	-165.8707259	15
535	CBX1	CBX5	-33.60652711	-93.69535104	7
536	PTPN12	PTK2B	-33.5598132	-85.8268805	8
537	ESR1	PPARG	-33.54224535	-124.295377	10
538	APP	C1QR1	-33.52055199	-102.9786486	8
539	ZAP70	LCP2	-33.51980547	-83.83132016	8
540	PIK3R1	PTK2	-33.45295382	-143.4695103	14
541	PTPN11	CSK	-33.45225403	-107.972887	10
542	NCOR2	HDAC3	-33.43088621	-123.7613897	10
543	DIPA	MDFI	-33.41648469	-275.3632229	19
544	LAT	PLCG1	-33.39324909	-104.2851291	9
545	SMARCA4	SMARCA2	-33.36340926	-73.89034773	6
546	BCL6	ZBTB16	-33.29316751	-92.97564178	8
547	NCK1	CRKL	-33.27477166	-115.4856006	10
548	JAK2	SYK	-33.25824935	-111.7408158	11
549	RARA	THRA	-33.23045779	-98.03845869	8
550	CCNA2	CCNE1	-33.20098475	-86.19745052	7
551	PTK2B	PDGFRB	-33.18243816	-94.18335239	10

552	JAK2	TYK2	-33.15466294	-97.43908829	9
553	GRB2	PLCG2	-33.15358051	-133.4569799	11
554	LAT	LCP2	-33.104406	-78.46792829	7
555	INHBC	INHBB	-33.1034073	-61.13765404	5
556	TNFRSF10A	TNFRSF10B	-33.10136856	-60.52826152	5
557	REM1	CDC25B	-33.05702283	-54.28692389	5
558	RABEP1	AFTIPHILIN	-33.05226513	-68.10316201	5
559	DIPA	KRTAP4-12	-33.02838293	-233.4692543	17
560	GRB2	ERBB2	-33.02799709	-150.9252139	14
561	PIK3R1	TYK2	-32.99424352	-107.4069855	10
562	APEX2	RIBC2	-32.94806323	-61.83105909	6
563	LCK	ZAP70	-32.90135464	-116.0301879	10
564	COL2A1	COL1A1	-32.87620443	-100.3510895	8
565	SYK	PTPN6	-32.86554283	-118.7756582	11
566	RIPK1	RIPK2	-32.79579325	-77.69960059	7
567	SHC1	CBL	-32.75931493	-144.5900294	13
568	FN1	MMP9	-32.73817565	-111.3577322	9
569	GNAS	GNAQ	-32.73806062	-113.6733988	8
570	SIN3A	HDAC2	-32.71826037	-120.9104291	10
571	YWHAZ	YWHAH	-32.6937727	-130.1203729	10
572	ATF3	FOSL2	-32.66048741	-57.88802068	5
573	SMN1	LSM2	-32.65337462	-87.60111767	7
574	FYN	LCP2	-32.6525298	-113.9976768	10
575	KRT15	NDP52	-32.6442493	-143.3894736	11
576	BMPR1B	SMURF1	-32.63634858	-194.6474246	13
577	NCOA3	NCOA2	-32.61165933	-81.32546031	7
578	NCOA1	TIF1	-32.59360525	-97.73605227	9
579	SHC1	STAT5A	-32.59325557	-115.735683	11
580	OSM	FBLN2	-32.58588828	-76.79095267	6
581	NUP98	NUP153	-32.56677358	-76.40998266	6
582	CD40	TNFRSF19	-32.56607728	-53.47486974	5
583	CD4	LCK	-32.55970867	-122.5366485	10
584	FN1	TGFBI	-32.55353135	-85.32745066	7
585	NOTCH1	NOTCH3	-32.53688296	-85.4482858	6
586	CD9	CD81	-32.5052127	-82.62493258	6
587	NCOA1	PPARGC1A	-32.50422281	-94.50659117	8
588	ZBTB16	NRIP1	-32.49577075	-87.90110711	9
589	EMP3	KCNQ2	-32.4374787	-47.02795198	4
590	CCL11	CCL13	-32.4374787	-50.19981384	4
591	RPP14	RPP21	-32.4374787	-55.74247697	4
592	Arhgef7	Trim35	-32.4374787	-35.40988928	4
593	Pftk1	Nat5	-32.4374787	-36.83099199	4
594	MAFF	MAFG	-32.41039613	-72.91127304	5

595	BMP7	BMP6	-32.40767763	-60.47934284	5
596	TNFRSF14	TNFRSF17	-32.40767763	-58.12704277	5
597	SYK	PLCG2	-32.37197104	-95.18028599	8
598	MLLT4	RALGDS	-32.365557	-92.15692824	7
599	RASA1	SHC1	-32.28121773	-125.3254865	12
600	PTPN11	PLCG1	-32.20979568	-141.9619983	13
601	MAP2K4	MAP2K7	-32.18825561	-79.73741389	6
602	TU3A	KIAA1267	-32.18825561	-58.04879702	6
603	MMP9	FBLN2	-32.18142273	-86.60267342	7
604	RARA	PPARG	-32.18018211	-95.65850212	8
605	STAT3	STAT1	-32.15265248	-113.9395696	10
606	PTK2B	CBL	-32.14192738	-103.8014498	11
607	BCL6	HDAC3	-32.13667985	-92.38746555	8
608	HSPA5	CANX	-32.12555011	-112.4662423	8
609	PLCG1	LCP2	-32.01650369	-100.9770405	9
610	BCAR1	CRK	-32.00183808	-104.8603711	9
611	ORC4L	MCM2	-31.99167383	-74.39574636	6
612	CHUK	MAP3K14	-31.96295261	-83.05833272	7
613	CSF3R	PTK2B	-31.95792521	-66.47528197	7
614	HNF4A	THRB	-31.88925458	-87.63468639	7
615	OSM	MMP9	-31.85054094	-76.79095267	6
616	TLR2	TLR4	-31.85050877	-74.22758315	5
617	SMAD7	SMAD6	-31.84156256	-79.39663943	6
618	ADAM15	CBL	-31.80737585	-70.1869827	7
619	MAP1A	CRIP1	-31.800998	-54.98866917	5
620	BLNK	CBL	-31.78599342	-84.86274172	8
621	PTPN11	GRB10	-31.77781584	-95.46808621	9
622	CANX	CALR	-31.77660889	-103.5384027	8
623	RBBP7	RBBP4	-31.72421716	-57.07161006	5
624	KCNJ12	ATP2B4	-31.71741906	-54.58546523	5
625	MCM10	MCM3	-31.70587608	-91.39203747	7
626	CASP8	RIPK1	-31.70415659	-103.5002758	8
627	AFAP	HABP4	-31.68006612	-57.53713921	5
628	SNRPD3	SNRPD1	-31.68006612	-66.19077455	5
629	PTK2B	LCK	-31.67173124	-115.2096598	11
630	FBLN2	FBLN1	-31.66388593	-73.54489411	6
631	GRB2	CRK	-31.6597022	-167.808353	14
632	COL4A2	COL1A1	-31.64751933	-85.75866846	7
633	COL4A1	COL1A1	-31.64751933	-86.78317277	7
634	VIM	KRT15	-31.63073262	-126.2457415	10
635	TRAF2	IKBKKG	-31.63064588	-138.4271118	11
636	BRCA1	RB1	-31.61476877	-156.3847421	14
637	PTK2B	FYN	-31.59120717	-140.520247	13

638	JAK2	VAV1	-31.54245229	-116.4323945	11
639	TRAF1	TRAF6	-31.49650211	-111.9558318	9
640	PTPN11	EPOR	-31.48659412	-105.5320496	10
641	PIK3R1	EGFR	-31.47494251	-162.1650224	16
642	GRB2	PTK2	-31.44723677	-166.9434391	16
643	PTPN11	CRK	-31.42206663	-117.0050637	11
644	PRKAR2A	PRKAR2B	-31.41622061	-88.78857443	6
645	TNFAIP3	MAP3K14	-31.39002439	-64.21578463	6
646	SOS1	PDGFRB	-31.38561486	-78.17246333	8
647	REM1	BAD	-31.34952371	-54.28692389	5
648	EPOR	PAG1	-31.33706949	-66.01767275	7
649	PTPRC	CD5	-31.32026466	-75.49957905	7
650	JAK3	IL2RB	-31.29465117	-77.59721877	7
651	CD82	CD81	-31.28412446	-80.29480992	6
652	IGSF1	INHBA	-31.2655359	-61.74378984	5
653	INHBC	INHBA	-31.2655359	-61.74378984	5
654	CCL5	CCL8	-31.26145759	-62.40204976	5
655	RABEP1	APIGBP1	-31.26145759	-68.10316201	5
656	JAK1	STAT3	-31.23779968	-114.540012	10
657	FYN	CBL	-31.22107532	-147.1857299	13
658	PIK3R1	RASA1	-31.19756881	-138.9056496	12
659	PTK2	SHC1	-31.19518894	-134.7707299	13
660	DCN	OSM	-31.18582016	-76.79095267	6
661	HXA1	CATSPER1	-31.15559461	-54.05257076	5
662	NOTCH2	NOTCH3	-31.15559461	-73.86678449	5
663	PTPN11	TYK2	-31.14416544	-97.9043831	9
664	PRKDC	XRCC6	-31.12227761	-95.81309889	8
665	SOS1	CBLB	-31.03169926	-70.98504121	7
666	RBMX	KHDRBS3	-31.02959518	-75.51382838	6
667	BIRC2	BIRC4	-31.02959518	-79.79049898	6
668	FGF1	FGF2	-31.02959518	-79.49214238	6
669	BAD	BAX	-31.02633054	-75.6748186	6
670	LMO2	LMO1	-30.99305818	-74.08992804	5
671	SIN3A	SIN3B	-30.97066433	-84.69341608	7
672	SFRS1	U2AF2	-30.93942936	-100.3646354	7
673	TIF1	PPARBP	-30.93657046	-75.10238772	7
674	DLG4	GRIN2B	-30.93432139	-114.6437909	9
675	DVL2	FXR2	-30.8953341	-116.6308025	10
676	COPG	COPB	-30.88277562	-80.97005152	6
677	RBBP4	DNMT1	-30.87851954	-66.54218575	6
678	RASA1	FYN	-30.87845917	-134.937216	12
679	IL16	INADL	-30.87706211	-88.44059837	6
680	BCL2A1	MCL1	-30.86954393	-70.96251794	5

681	HLA-B	HLA-G	-30.86954393	-68.39451381	5
682	MEP50	GEMIN5	-30.86954393	-66.90639459	5
683	RGS16	RGS19	-30.86954393	-61.16342782	5
684	LIN7B	LIN7A	-30.86954393	-57.12851243	5
685	PER3	PER1	-30.86886426	-70.13323858	5
686	USP7	TNFRSF11A	-30.86886426	-53.47486974	5
687	CASP8	CASP10	-30.85105869	-94.78382841	7
688	GLI1	GLI2	-30.82804079	-58.93718969	4
689	Rheb1l	Rhobtb1	-30.82804079	-36.83099199	4
690	Rheb1l	Rhoj	-30.82804079	-36.83099199	4
691	Arhgef7	Smad3	-30.82804079	-35.40988928	4
692	1200013B22 Rik	Rhoj	-30.82804079	-35.47234916	4
693	Trim35	Smad3	-30.82804079	-35.40988928	4
694	Smad2	Smad3	-30.82804079	-36.73198694	4
695	KCNQ5	EMP3	-30.82749731	-47.02795198	4
696	KCNQ5	KCNQ2	-30.82749731	-47.02795198	4
697	ZNF8	ZFHX1B	-30.82749731	-40.52136615	4
698	PTCH	PTCH2	-30.82749731	-67.36095093	4
699	POP1	RPP14	-30.82749731	-55.74247697	4
700	POP1	RPP21	-30.82749731	-55.74247697	4
701	Rab38	Arhgef6	-30.82749731	-35.59583626	4
702	Rhobtb1	Pftk1	-30.82749731	-36.83099199	4
703	Rhobtb1	Nat5	-30.82749731	-36.83099199	4
704	Rhoj	Ilkap	-30.82749731	-36.15057069	4
705	JAK2	PTK2B	-30.77981862	-108.2485729	11
706	ACVR1B	INHBC	-30.7754058	-61.91084393	5
707	GGA2	GGA1	-30.7754058	-72.97296661	5
708	SYK	PTK2B	-30.66180183	-95.26766814	10
709	E2F1	E2F2	-30.65974521	-78.48187091	6
710	GRB2	BLNK	-30.65670975	-120.3327043	10
711	TCF4	TCF3	-30.64875469	-83.94407337	7
712	EPOR	PTK2B	-30.60277689	-84.59039467	9
713	RPA2	RPA1	-30.59509422	-79.54375183	6
714	ORC1L	ORC2L	-30.59509422	-77.83249473	6
715	PTPN11	PTPN12	-30.59238456	-78.21043474	8
716	IRS2	VAV3	-30.59101314	-56.64860035	6
717	CEBPG	BATF	-30.59101314	-77.28930497	6
718	PML	JUN	-30.59048666	-127.5630773	12
719	EP300	JUN	-30.58288398	-154.1086094	14
720	ITK	PLCG1	-30.57818452	-86.38077167	8
721	TNFRSF1A	TRPC4AP	-30.57769724	-66.07590436	6
722	IPO7	RANBP5	-30.55176989	-71.48477226	5

723	STAM	STAM2	-30.55176989	-60.90738284	5
724	SNTB2	SNTB1	-30.54905102	-71.55623122	5
725	SP1	SP3	-30.53379597	-82.46918274	7
726	ZAP70	CD3Z	-30.5130943	-96.66278047	8
727	PPARG	THRB	-30.50973846	-87.58117972	7
728	CD22	LCP2	-30.50396706	-59.23466246	6
729	PPARBP	NR0B2	-30.49743689	-62.16398723	6
730	FLJ32855	KIAA1267	-30.49743689	-60.12905523	6
731	FLJ32855	C16orf48	-30.49743689	-57.60455668	6
732	HSPCA	TEBP	-30.47408992	-98.7493154	8
733	EPOR	IGF1R	-30.46509596	-87.499932	8
734	RARA	PPARA	-30.42990791	-99.84815686	8
735	LCK	CD3Z	-30.39640064	-112.7030258	9
736	PTK2	CRKL	-30.38170107	-100.8116102	10
737	NRBF2	PPARBP	-30.37041296	-55.92972196	5
738	CGI-63	PPARBP	-30.37041296	-55.92972196	5
739	PTPN6	PTK2B	-30.36253445	-107.3659262	11
740	VAMP1	VAMP2	-30.33846747	-69.95625593	5
741	ZNF250	TU3A	-30.32094537	-63.78927639	6
742	FYN	YES1	-30.26928892	-111.3274872	9
743	PAG1	ZAP70	-30.26729977	-67.9525623	7
744	PTPN11	PTK2	-30.24728526	-121.3937455	12
745	THBS1	COL1A1	-30.22677991	-97.73549068	8
746	CD22	VAV1	-30.22331259	-69.70956063	7
747	BMP7	BMP4	-30.1754454	-66.2271422	5
748	SNRPB	SNRPD1	-30.1754454	-67.28938684	5
749	SRC	YES1	-30.16948538	-117.6526729	9
750	ABLIM1	RIBC2	-30.14797327	-63.84199498	6
751	PTPN6	CBL	-30.12052074	-114.4451687	11
752	JAK2	PIK3R1	-30.1170636	-132.7598553	13
753	DDIT3	ATF4	-30.10441242	-83.51794802	7
754	STX1A	SNAP23	-30.0802505	-113.8608355	8
755	JAK2	INSR	-30.07291882	-113.1983839	10
756	RELA	RXRA	-30.06364042	-143.0928839	12
757	NFKB1	JUN	-30.05027524	-131.8268917	12
758	VAV1	LCP2	-30.02714709	-87.35803829	8
759	NCOR1	ZBTB16	-30.0252852	-99.81703578	9
760	PTK2B	SHB	-30.02217216	-79.34269335	8
761	LRPAP1	DAB1	-30.0130455	-62.39035372	5
762	GRB10	SHC1	-29.98785821	-99.95373595	9
763	USP2	TU3A	-29.93313847	-50.42322479	5
764	AP1G1	AP1G2	-29.92701976	-74.17693941	5
765	BMPR1B	BMPR2	-29.85857895	-96.61389367	7

766	ACVR1	SMURF1	-29.85557579	-180.8283706	12
767	PAG1	PLCG1	-29.82246951	-79.60351768	8
768	USP7	TRIM37	-29.8203985	-66.23305171	6
769	FYN	NCK1	-29.81354004	-139.4346562	12
770	PAG1	BCR	-29.81052911	-55.89621302	6
771	EPOR	CSF3R	-29.80569961	-55.91574806	6
772	PLSCR1	RBPMS	-29.78827891	-148.1500656	11
773	GTF2A1	TBP	-29.76841192	-86.51154003	7
774	SRC	SHC1	-29.76590365	-168.7906788	15
775	EGFR	SHC1	-29.76590365	-162.1615187	15
776	TSC22D4	EFCBP2	-29.74677946	-81.79213535	7
777	GDF9	BMP6	-29.7294285	-48.50611559	4
778	Rheb1l	Rps27a	-29.7294285	-36.83099199	4
779	Rheb1l	Fbxo3	-29.7294285	-36.83099199	4
780	Rheb1l	Uhmk1	-29.7294285	-36.83099199	4
781	Rheb1l	Rhod	-29.7294285	-36.83099199	4
782	Arhgef7	Sqstm1	-29.7294285	-35.40988928	4
783	Arhgef7	Map2k3	-29.7294285	-35.40988928	4
784	1200013B22 Rik	Uhmk1	-29.7294285	-35.47234916	4
785	1200013B22 Rik	Rhod	-29.7294285	-35.47234916	4
786	GDF5	GDF9	-29.72834147	-48.50611559	4
787	CCL8	CCL11	-29.72834147	-50.19981384	4
788	CCL8	CCL13	-29.72834147	-50.19981384	4
789	POP7	RPP14	-29.72834147	-55.74247697	4
790	POP7	RPP21	-29.72834147	-55.74247697	4
791	POP4	RPP14	-29.72834147	-55.74247697	4
792	POP4	RPP21	-29.72834147	-55.74247697	4
793	Rps27a	1200013B22R ik	-29.72834147	-35.47234916	4
794	Rps27a	Ilkap	-29.72834147	-36.15057069	4
795	Rps27a	Pftk1	-29.72834147	-36.83099199	4
796	Rps27a	Nat5	-29.72834147	-36.83099199	4
797	Fbxo3	Arhgef6	-29.72834147	-35.59583626	4
798	Fbxo3	Pftk1	-29.72834147	-36.83099199	4
799	Fbxo3	Nat5	-29.72834147	-36.83099199	4
800	Sqstm1	Trim35	-29.72834147	-35.40988928	4
801	Map2k3	Arhgef6	-29.72834147	-35.59583626	4
802	Map2k3	Trim35	-29.72834147	-35.40988928	4
803	Uhmk1	Ilkap	-29.72834147	-36.15057069	4
804	Uhmk1	Pftk1	-29.72834147	-36.83099199	4
805	Uhmk1	Nat5	-29.72834147	-36.83099199	4

806	Rhod	Ilkap	-29.72834147	-36.15057069	4
807	Rhod	Nkiras1	-29.72834147	-38.15308965	4
808	GTF2A1	GTF2B	-29.69263171	-76.96296024	6
809	YWHAH	YWHAG	-29.67995499	-85.85211212	7
810	DIPA	LDOC1	-29.6567833	-138.1661495	11
811	FN1	HABP2	-29.62070003	-76.79095267	6
812	HDAC1	SIN3B	-29.58543353	-97.96864765	8
813	C1QR1	FBLN2	-29.57222511	-76.79095267	6
814	CASP7	CASP9	-29.57222511	-81.62883586	6
815	PRKCABP	SDCBP	-29.57222511	-85.31041601	6
816	CSNK1D	CSNK1E	-29.57107716	-60.77524715	5
817	SNRPD2	SNRPD1	-29.56903787	-67.28938684	5
818	GTF2H1	MNAT1	-29.56814348	-66.48111952	6
819	NCOA1	NR0B2	-29.56441359	-74.76801853	7
820	SERPINA5	SERPINB6	-29.55178211	-65.91715059	5
821	PTK2B	SHC1	-29.55122404	-118.3624281	12
822	JAK2	EGFR	-29.5418775	-135.6048075	13
823	POLR2A	BAZ1B	-29.53666482	-96.3703725	7
824	CDH15	CDH2	-29.52780983	-62.62390552	5
825	CD40	TNFRSF8	-29.52373097	-56.69707335	5
826	PTK2B	PECAM1	-29.48122689	-77.0966084	8
827	PTK2B	BCAR1	-29.4761638	-94.4072495	9
828	RXRA	HNF4A	-29.45798692	-104.4827955	8
829	ITK	PLCG2	-29.41780277	-66.48780753	6
830	PTK2B	PLCG1	-29.40724611	-105.7825471	11
831	RELA	SP1	-29.39197455	-130.3367837	12
832	COL4A1	COL1A2	-29.34175122	-69.93794553	6
833	RELA	MYOD1	-29.34017324	-101.8729666	10
834	PIK3R1	VAV1	-29.32182843	-124.2346563	12
835	GHR	KIT	-29.29760047	-69.09055156	7
836	TYK2	EPOR	-29.29760047	-71.47875929	7
837	JAK2	KIT	-29.27943552	-93.34691818	9
838	DEF6	ARHGDIA	-29.21819531	-42.74755273	4
839	DSC1	DSC2	-29.21819531	-54.41651876	4
840	ATP2B4	GUCY1A2	-29.21819531	-43.61261183	4
841	Rab38	Rhoj	-29.21819531	-35.35910128	4
842	Rhobtb1	Rhoj	-29.21819531	-36.83099199	4
843	B2M	LILRB1	-29.20293115	-68.16645686	5
844	TCF3	ID2	-29.19535356	-78.21011612	6
845	LSM1	LSM6	-29.18301252	-74.46341483	6
846	PLCG1	CBL	-29.1448595	-112.4999166	11
847	CDC7	MCM7	-29.12551869	-80.01032244	6
848	BMPR1B	SMAD2	-29.12328583	-175.3069097	12

849	LRPAP1	APOE	-29.08624926	-62.39035372	5
850	MAP3K14	TRPC4AP	-29.08149027	-54.49440304	5
851	SRC	GRAP2	-29.04477411	-114.6672196	10
852	RFC4	RFC1	-29.03248885	-62.09133682	5
853	BMP2	BMP4	-29.02976961	-66.2271422	5
854	SNRPE	SNRPD3	-29.02976961	-65.48657754	5
855	PPARGC1A	NCOA2	-29.01342743	-61.67688606	6
856	GTF2H1	ERCC3	-29.0101619	-72.97748307	6
857	KHDRBS1	WAS	-29.00092927	-69.56914187	7
858	YWHAB	YWHAG	-28.96498241	-85.85211212	7
859	PLCG1	GRAP2	-28.96453203	-108.8299725	9
860	PTPN6	EPOR	-28.94143404	-92.89211295	9
861	SYK	ITK	-28.91871394	-72.61243706	7
862	NR3C1	JUN	-28.9077122	-140.5444232	13
863	HDAC9	HDAC5	-28.90765905	-67.84873189	6
864	MCM10	MYST2	-28.89519476	-69.54711026	6
865	CCL7	CCL11	-28.88049998	-50.19981384	4
866	CCL7	CCL13	-28.88049998	-50.19981384	4
867	STRAP	Wwp2	-28.88049998	-37.65198224	4
868	COPB	COPB2	-28.8769644	-67.34822734	5
869	MAP2K1	MAP2K2	-28.8769644	-57.13037623	5
870	KIAA0408	USP2	-28.8769644	-50.42322479	5
871	COL4A3	COL1A2	-28.87229329	-69.93794553	6
872	FN1	MATN2	-28.85352829	-88.12201829	7
873	EPOR	CRKL	-28.83777334	-80.09127891	8
874	PTPN6	RASA1	-28.82583931	-106.2772981	10
875	TRAF6	MAP3K7	-28.76267001	-108.2190772	8
876	WWP2	WWP1	-28.75892387	-62.78592424	5
877	SRC	ERBB2	-28.74717966	-109.6039678	11
878	EGFR	PDGFRA	-28.74612024	-103.6483368	9
879	SRC	LYN	-28.74469116	-155.2766279	13
880	ERBB2	MUC1	-28.742113	-68.91182163	7
881	NCOR1	HDAC3	-28.72024739	-110.4278692	9
882	BRCA1	JUN	-28.71278637	-141.3348215	13
883	GGA3	AP1G1	-28.68105637	-69.74216221	5
884	KCNJ12	KCNJ4	-28.67561715	-56.39492428	5
885	BCL2L1	BCL2A1	-28.66261416	-84.67621136	6
886	FOS	ATF3	-28.66261416	-69.65580396	6
887	HDAC1	DNMT1	-28.64072856	-93.84842714	8
888	PRKCB1	PRKCD	-28.58268394	-93.97435546	7
889	NSD1	PPARBP	-28.57987886	-50.70810029	5
890	RELA	RB1	-28.57115228	-132.4988695	13
891	PIK3R1	EPOR	-28.56861525	-102.2538134	10

892	STAT3	PIK3R1	-28.56007532	-132.981611	12
893	CD40	TNFRSF11A	-28.54399	-53.47486974	5
894	SMARCB1	SMARCC1	-28.54399	-60.1994413	5
895	JUN	TBP	-28.53192292	-136.9601126	12
896	MAPK8	MAPK9	-28.52616406	-86.06256784	7
897	PLCG1	CRK	-28.51318208	-108.3651143	10
898	GAB2	SHB	-28.50342002	-59.79636895	6
899	ORC1L	MCM2	-28.50116267	-77.44608277	6
900	LSM1	LSM5	-28.49218063	-74.64573639	6
901	HRAS	RRAS2	-28.48684069	-81.74668972	7
902	VAV1	LCK	-28.47502339	-102.7353504	10
903	CHUK	RIPK1	-28.4591813	-77.69048476	7
904	PLCG1	PLCG2	-28.45595749	-99.3640846	8
905	SIT1	PTK2B	-28.4489583	-57.14192258	6
906	PML	NR3C1	-28.43470265	-108.7244988	11
907	TNFRSF19	TRIM37	-28.43271869	-53.47486974	5
908	STAT5A	STAT1	-28.43244139	-86.62968762	8
909	ESR1	VDR	-28.40291637	-90.68348975	8
910	PSMA1	LNK	-28.39916854	-77.56049878	7
911	SRC	CRKL	-28.39406721	-116.7709607	11
912	NRIP1	NR0B2	-28.35812221	-68.87628895	6
913	HSPA1A	HSPA8	-28.35095766	-111.364164	8
914	KHDRBS1	FASLG	-28.33195004	-68.99242733	7
915	SIT1	CBL	-28.32126168	-59.36900012	6
916	BIRC3	BIRC4	-28.30663518	-64.46736764	5
917	TNPO1	IPO7	-28.30051564	-71.48477226	5
918	SYK	GAB2	-28.29460641	-69.91291521	7
919	SYK	PAG1	-28.29460641	-67.87961111	7
920	PRKCZ	PRKCI	-28.2891537	-81.46463484	6
921	NCOA1	PELP1	-28.28609734	-57.25043877	6
922	RIBC2	USP2	-28.28281176	-50.42322479	5
923	VAV1	CRKL	-28.24067621	-99.54411743	9
924	GNB1	GNB5	-28.21979171	-81.58546998	5
925	ORC1L	MCM10	-28.21429848	-76.72650233	6
926	GHR	EGFR	-28.20961701	-95.51617822	9
927	RGS18	RIC8	-28.18898346	-44.28768091	4
928	RGS16	RGS18	-28.1868091	-44.28768091	4
929	RGS16	RGS5	-28.1868091	-45.27029429	4
930	RIC8	RGS5	-28.1868091	-45.27029429	4
931	RAB11FIP2	RAB11FIP5	-28.1868091	-59.25927319	4
932	ACTR2	ARPC5	-28.1868091	-57.62736242	4
933	NRP1	NRP2	-28.1868091	-59.95971967	4
934	Rasd2	Rhebl1	-28.1868091	-36.83099199	4

935	Rasd2	Arhgef7	-28.1868091	-35.40988928	4
936	Rasd2	1200013B22R ik	-28.1868091	-35.47234916	4
937	Rasd2	Arhgef6	-28.1868091	-35.59583626	4
938	Rasd2	Trim35	-28.1868091	-35.40988928	4
939	Rasd2	Smad2	-28.1868091	-36.73198694	4
940	Rasd2	Ilkap	-28.1868091	-36.15057069	4
941	Rasd2	Nkiras1	-28.1868091	-38.15308965	4
942	Rasd2	Pftk1	-28.1868091	-36.83099199	4
943	Rasd2	Nat5	-28.1868091	-36.83099199	4
944	Rasd2	Smad7	-28.1868091	-37.36162024	4
945	GRB2	PIK3R2	-28.18542496	-102.9243264	9
946	PTK2	BCAR1	-28.18271888	-102.4010602	9
947	VAV1	WAS	-28.17807724	-81.67199171	8
948	PAG1	PECAM1	-28.12515664	-58.96763868	6
949	GHR	BCR	-28.12189067	-57.57909023	6
950	TYK2	IL4R	-28.12189067	-68.83726301	6
951	FLJ32855	TU3A	-28.12189067	-60.3720014	6
952	CD22	EPOR	-28.11973949	-55.28788369	6
953	SAA1	HABP2	-28.11971896	-50.17449604	4
954	ASIP	POMC	-28.11971896	-63.55428844	4
955	POP1	POP7	-28.11971896	-55.74247697	4
956	POP1	POP4	-28.11971896	-55.74247697	4
957	Rab38	Rps27a	-28.11971896	-35.35910128	4
958	Rab38	Fbxo3	-28.11971896	-35.59583626	4
959	Rab38	Sqstm1	-28.11971896	-36.08811081	4
960	Rab38	Uhmk1	-28.11971896	-35.35910128	4
961	Rab38	Rhod	-28.11971896	-35.35910128	4
962	Rhobtb1	Sqstm1	-28.11971896	-38.00368784	4
963	Rhobtb1	Map2k3	-28.11971896	-36.03952258	4
964	Rhobtb1	Rhod	-28.11971896	-36.83099199	4
965	CCL8	CCL14	-28.11917541	-50.19981384	4
966	MAP2K2	KSR2	-28.11917541	-44.02281861	4
967	SCNN1G	SCNN1A	-28.11917541	-52.72839556	4
968	TAF1A	TAF1B	-28.11917541	-46.92642106	4
969	TAF1C	TAF1B	-28.11917541	-46.92642106	4
970	CRIP1	ATP2B4	-28.11917541	-43.61261183	4
971	CRIP1	GUCY1A2	-28.11917541	-43.61261183	4
972	CDC45L	CCDC5	-28.11917541	-47.13266257	4
973	SFRS2IP	SFRS11	-28.11917541	-48.7106621	4
974	AQP1	FLJ22746	-28.11917541	-39.12516735	4
975	Rps27a	Stat1	-28.11917541	-35.47234916	4
976	Fbxo3	Rhoj	-28.11917541	-36.83099199	4

977	Sqstm1	Rhoj	-28.11917541	-37.32326654	4
978	Sqstm1	Smad3	-28.11917541	-35.40988928	4
979	Sqstm1	Rras2	-28.11917541	-36.87958022	4
980	Map2k3	Rhoj	-28.11917541	-35.35910128	4
981	Map2k3	Smad3	-28.11917541	-35.40988928	4
982	Uhmkl	Stat1	-28.11917541	-35.47234916	4
983	Rhod	Rras2	-28.11917541	-38.6453642	4
984	Rhod	Stat1	-28.11917541	-36.79444682	4
985	EPOR	PRLR	-28.10667203	-71.25452147	6
986	SMN1	LSM6	-28.10667203	-74.79081157	6
987	VAMP8	VAMP2	-28.10596483	-61.57591392	5
988	SNX6	SNX4	-28.10188542	-57.44171401	5
989	MAPK8IP3	DUSP16	-28.10188542	-63.3004028	5
990	MAP3K14	TNFRSF11A	-28.10188542	-52.45233975	5
991	CFLAR	DEDD	-28.10188542	-60.34274537	5
992	PROC	F10	-28.06727287	-68.03132419	5
993	SNRPB	SNRPD2	-28.06727287	-67.28938684	5
994	GRB2	SHB	-28.03828855	-107.7701219	10
995	SPTAN1	SPTA1	-28.01033347	-68.63134539	5
996	EXOC8	RALA	-28.00370146	-91.69729533	6
997	B2M	LILRB2	-27.95180083	-68.16645686	5
998	NCOA1	NR3C1	-27.94520126	-120.6194393	11
999	RAPGEF1	CRKL	-27.94002377	-61.59795291	6
1000	RIPK1	MAP3K14	-27.93834498	-61.73017949	6

Table S2 Predictions of 466 KEGG pathways for 274 proteins and 123 GO annotations for 114 proteins. The 2nd column is the predicted KEGG and GO IDs for proteins in the 1st column, with 3rd column as corresponding KEGG pathway names and GO terms. Ratio is the number of significant partners with the assigned annotation(s) divided by the total number of significant partners. P values were first calculated by Fisher's exact test and then adjusted by "BH" method [81]. Table S2 was partially reproduced from REF. 76.

Protein	KEGG ID	KEGG Pathway Name	Ratio	P value
DNASE1	hsa04810	Regulation of actin cytoskeleton	2/2	0.00302
DLG4	hsa04080; hsa04530; hsa04720	Neuroactive ligand-receptor interaction; Tight junction; Long-term potentiation	2/11	0.144; 0.0366; 0.0129
GRIN2B	hsa04020	Calcium signaling pathway	3/6	0.00169
PML	hsa05220	Chronic myeloid leukemia	4/30	0.00204
CREBBP	hsa04010; hsa04620; hsa04660; hsa04662	MAPK signaling pathway; Toll-like receptor signaling pathway; T cell receptor signaling pathway; B cell receptor signaling pathway	3/12	0.0276; 0.00278; 0.00321; 0.00144
NCOR1	hsa04330	Notch signaling pathway	2/9	0.00458
NCOR2	hsa04110; hsa05220	Cell cycle; Chronic myeloid leukemia	3/14	0.00565; 0.00208
RARA	hsa03320; hsa04080	PPAR signaling pathway; Neuroactive ligand-receptor interaction	3/9	0.000607; 0.0165
BRCA1	hsa04110; hsa05212; hsa05220	Cell cycle; Pancreatic cancer; Chronic myeloid leukemia	4/22	0.00282; 0.000722; 0.000821
TGFBR3	hsa04010; hsa04060; hsa04350; hsa05210; hsa05212; hsa05220	MAPK signaling pathway; Cytokine-cytokine receptor interaction; TGF-beta signaling pathway; Colorectal cancer; Pancreatic cancer; Chronic myeloid leukemia	2/6	0.0414; 0.0422; 0.00607; 0.00583; 0.00442; 0.0047
GDF9	hsa04350	TGF-beta signaling pathway	5/5	3.91E-08
IGSF1	hsa04060; hsa04350	Cytokine-cytokine receptor interaction; TGF-beta signaling pathway	6/6	3.46e-07; 1.29e-09
NCOA1	hsa04920	Adipocytokine signaling pathway	5/38	0.000551
NR3C1	hsa04620; hsa04920	Toll-like receptor signaling pathway; Adipocytokine signaling pathway	4/22	0.00181; 6e-04
STAT1P1	hsa04060; hsa04630	Cytokine-cytokine receptor interaction; Jak-STAT signaling pathway	2/3	0.0109; 0.00422
MET	hsa04810	Regulation of actin cytoskeleton	3/6	0.00253
KHDRBS1	hsa04650; hsa04660	Natural killer cell mediated cytotoxicity; T cell receptor signaling pathway	5/18	0.000396; 0.000144
AR	hsa04110;	Cell cycle;	3/19	0.012;

	hsa05212;	Pancreatic cancer;		0.00392;
	hsa05220	Chronic myeloid leukemia		0.00426
PTPN11	hsa04510	Focal adhesion	16/50	3.81E-09
RET	hsa04510	Focal adhesion	7/15	6.87E-06
JUN	hsa05220	Chronic myeloid leukemia	7/35	7.60E-06
TRAF3	hsa04010	MAPK signaling pathway	2/4	0.019
LRPAP1	hsa05010	Alzheimer's disease	2/4	0.00865
APH1B	hsa04330;	Notch signaling pathway;	2/2	0.000308;
	hsa05010	Alzheimer's disease		0.0021
CD19	hsa04650	Natural killer cell mediated cytotoxicity	5/10	2.40E-05
CD22	hsa04650;	Natural killer cell mediated cytotoxicity;	8/23	1.43e-06;
	hsa04660	T cell receptor signaling pathway		2.97e-07
ELA2	hsa04080	Neuroactive ligand-receptor interaction	2/4	0.0225
CD3E	hsa04630;	Jak-STAT signaling pathway;	2/6	0.0157;
	hsa04910;	Insulin signaling pathway;		0.0128;
	hsa05220	Chronic myeloid leukemia		0.0047
GNB2L1	hsa04630;	Jak-STAT signaling pathway;	2/3	0.00422;
	hsa04650;	Natural killer cell mediated cytotoxicity;		0.00362;
	hsa04670;	Leukocyte transendothelial migration;		0.00281;
	hsa04910;	Insulin signaling pathway;		0.00351;
	hsa04920;	Adipocytokine signaling pathway;		0.00121;
	hsa04930;	Type II diabetes mellitus;		0.000711;
	hsa05220	Chronic myeloid leukemia		0.0014
TRAF2	hsa04620	Toll-like receptor signaling pathway	4/17	0.000848
TRAF1	hsa04010;	MAPK signaling pathway;	4/10	0.00253;
	hsa04210	Apoptosis		9.35e-05
CSK	hsa04650	Natural killer cell mediated cytotoxicity	10/15	1.32E-10
TNK2	hsa04630;	Jak-STAT signaling pathway;	2/3	0.00422;
	hsa04810;	Regulation of actin cytoskeleton;		0.00713;
	hsa04910	Insulin signaling pathway		0.00351
KIT	hsa04630	Jak-STAT signaling pathway	11/38	2.82E-07
SIT1	hsa04670	Leukocyte transendothelial migration	5/8	3.53E-06
BLNK	hsa04650;	Natural killer cell mediated cytotoxicity;	7/15	8.33e-07;
	hsa04660;	T cell receptor signaling pathway;		2.32e-07;
	hsa04664	Fc epsilon RI signaling pathway		3.91e-08
GAB1	hsa04510	Focal adhesion	11/27	7.15E-08
IRS2	hsa04630	Jak-STAT signaling pathway	9/27	8.70E-07
GAB2	hsa04650	Natural killer cell mediated cytotoxicity	11/34	3.99E-08
IRS4	hsa04060	Cytokine-cytokine receptor interaction	4/7	0.000751
PTPN12	hsa04510	Focal adhesion	9/15	3.25E-08
ADAM15	hsa04670	Leukocyte transendothelial migration	3/10	0.00253
MUC1	hsa04510	Focal adhesion	5/9	6.00E-05
PTPRC	hsa04650	Natural killer cell mediated cytotoxicity	7/15	8.33E-07
TEK	hsa04060	Cytokine-cytokine receptor interaction	4/5	0.000177

CD2AP	hsa04670	Leukocyte transendothelial migration	3/10	0.00253
DNM1	hsa04810	Regulation of actin cytoskeleton	2/3	0.00713
ABL1	hsa04510	Focal adhesion	5/9	6.00E-05
PAG1	hsa04660	T cell receptor signaling pathway	13/45	8.51E-10
CD28	hsa04640	Hematopoietic cell lineage	4/12	0.000177
MAPK14	hsa04510; hsa04910; hsa04930; hsa05210; hsa05212	Focal adhesion; Insulin signaling pathway; Type II diabetes mellitus; Colorectal cancer; Pancreatic cancer	3/3	0.000231; 9.48e-05; 6.83e-06; 2.8e-05; 1.96e-05
RAPGEF1	hsa04670; hsa04810	Leukocyte transendothelial migration; Regulation of actin cytoskeleton	5/10	1.28e-05; 0.000137
BCL2A1	hsa01510; hsa04210; hsa05030	Neurodegenerative Diseases; Apoptosis; Amyotrophic lateral sclerosis (ALS)	2/3	NA; 0.0018; NA
GAB3	hsa04060	Cytokine-cytokine receptor interaction	5/21	0.00696
CSF1R	hsa04630; hsa04670	Jak-STAT signaling pathway; Leukocyte transendothelial migration	4/17	0.00278;0.0 0132
GRB7	hsa04650; hsa05214	Natural killer cell mediated cytotoxicity; Glioma	6/11	1.98e-06; 5.2e-08
TUB	hsa04510	Focal adhesion	3/8	0.00459
FRS2	hsa04650; hsa05214	Natural killer cell mediated cytotoxicity; Glioma	2/4	0.00627; 0.00188
ITK	hsa04664	Fc epsilon RI signaling pathway	8/13	4.73E-10
MST1R	hsa04510; hsa04810	Focal adhesion; Regulation of actin cytoskeleton	4/10	0.00108; 0.00131
PTK2B	hsa04630	Jak-STAT signaling pathway	21/68	1.39E-12
EPHA2	hsa04920	Adipocytokine signaling pathway	3/6	0.000185
SCAP1	hsa04650; hsa04670	Natural killer cell mediated cytotoxicity; Leukocyte transendothelial migration	4/14	0.00118; 0.000719
CALD1	hsa04080	Neuroactive ligand-receptor interaction	2/15	0.233
CBLB	hsa04650	Natural killer cell mediated cytotoxicity	8/19	3.44E-07
SYNJ1	hsa04810	Regulation of actin cytoskeleton	2/3	0.00713
SH3KBP1	hsa04010; hsa04510; hsa04910; hsa05220	MAPK signaling pathway; Focal adhesion; Insulin signaling pathway; Chronic myeloid leukemia	2/4	0.019; 0.0114; 0.00608; 0.00238
INPP5D	hsa04630	Jak-STAT signaling pathway	6/11	3.23E-06
RASA1	hsa04510	Focal adhesion	11/22	7.50E-09
SHB	hsa04650; hsa04660	Natural killer cell mediated cytotoxicity; T cell receptor signaling pathway	8/25	2.84e-06; 4.91e-07
BCR	hsa04630	Jak-STAT signaling pathway	9/17	1.62E-08
SYN1	hsa04020; hsa04510;	Calcium signaling pathway; Focal adhesion;	2/6	0.0199; 0.0244;

	hsa04670;	Leukocyte transendothelial migration;		0.0101;
	hsa04810	Regulation of actin cytoskeleton		0.0275
BCL3	hsa04620;	Toll-like receptor signaling pathway;	2/8	0.0128;
	hsa05120	Epithelial cell signaling in Helicobacter pylori infection		0.00677
GIT2	hsa04810	Regulation of actin cytoskeleton	3/3	0.000276
HLA-F	hsa04650	Natural killer cell mediated cytotoxicity	4/4	5.10E-06
CTLA4	hsa04670	Leukocyte transendothelial migration	2/7	0.0132
SOSTDC1	hsa04350	TGF-beta signaling pathway	3/3	2.93E-05
PCTK1	hsa04010	MAPK signaling pathway	2/6	0.0414
REM1	hsa04010;	MAPK signaling pathway;	3/11	0.0219;
	hsa04510;	Focal adhesion;		0.0107;
	hsa04910;	Insulin signaling pathway;		0.00438;
	hsa05210;	Colorectal cancer;		0.00153;
	hsa05212;	Pancreatic cancer;		0.00109;
	hsa05218;	Melanoma;		0.00107;
	hsa05220	Chronic myeloid leukemia		0.00119
E2F4	hsa05220	Chronic myeloid leukemia	3/10	0.000945
RBBP4	hsa05220	Chronic myeloid leukemia	5/17	2.62E-05
ERBB2IP	hsa05217	Basal cell carcinoma	2/4	0.00149
KRT18	hsa04370;	VEGF signaling pathway;	2/7	0.00627;
	hsa04510	Focal adhesion		0.0326
CTNNB1	hsa04514	Cell adhesion molecules (CAMs)	2/4	0.00581
RICS	hsa04510	Focal adhesion	8/15	4.32E-07
CTNNA1	hsa05216	Thyroid cancer	2/2	0.000147
GRIK2	hsa00230	Purine metabolism	2/7	0.0199
BAD	hsa04010	MAPK signaling pathway	2/12	0.143
MYF6	hsa04080	Neuroactive ligand-receptor interaction	2/18	0.299
ASCL2	hsa04080	Neuroactive ligand-receptor interaction	2/18	0.299
PPEF2	hsa04080	Neuroactive ligand-receptor interaction	2/18	0.299
KCNQ5	hsa04080	Neuroactive ligand-receptor interaction	2/15	0.233
PPEF1	hsa04080	Neuroactive ligand-receptor interaction	2/18	0.299
ESR1	hsa04110	Cell cycle	4/21	0.00249
GRM7	hsa04310	Wnt signaling pathway	2/11	0.0451
EMP3	hsa04080	Neuroactive ligand-receptor interaction	2/17	0.279
CALM3	hsa04350	TGF-beta signaling pathway	2/4	0.00295
STX1A	hsa04130	SNARE interactions in vesicular transport	5/6	6.01E-09
ITGB1	hsa04640	Hematopoietic cell lineage	4/6	1.07E-05
SPIB	hsa04010	MAPK signaling pathway	2/2	0.00436
FOS	hsa04310	Wnt signaling pathway	4/19	0.00367
PKP2	hsa01430	Cell Communication	3/5	NA
PECAM1	hsa04510;	Focal adhesion;	4/26	0.0223;
	hsa04640	Hematopoietic cell lineage		0.00207
SERPINB13	hsa04610	Complement and coagulation cascades	2/2	0.000565

PCAF	hsa01510; hsa04110; hsa04310; hsa04350; hsa04520; hsa04630; hsa04720; hsa04916; hsa05040	Neurodegenerative Diseases; Cell cycle; Wnt signaling pathway; TGF-beta signaling pathway; Adherens junction; Jak-STAT signaling pathway; Long-term potentiation; Melanogenesis; Huntington's disease	2/2	NA; 0.00122; 0.00178; 0.000776; 0.000643; 0.0018; 0.000591; 0.00101; NA
MYC	hsa04110	Cell cycle	2/5	0.00712
CEBPB	hsa04620	Toll-like receptor signaling pathway	3/10	0.00178
CEBPG	hsa04010	MAPK signaling pathway	4/9	0.00178
CKS1B	hsa04110	Cell cycle	2/2	0.00122
BIRC5	hsa04210; hsa04510	Apoptosis; Focal adhesion	3/5	0.000208;0.00132
LYN	hsa04650	Natural killer cell mediated cytotoxicity	12/22	4.55E-11
MCM10	hsa04110	Cell cycle	12/26	5.61E-11
AKAP8	hsa04110; hsa05212; hsa05214; hsa05218; hsa05220	Cell cycle; Pancreatic cancer; Glioma; Melanoma; Chronic myeloid leukemia	2/3	0.00278; 0.00132; 0.00116; 0.00131; 0.0014
CDC25A	hsa04010	MAPK signaling pathway	2/3	0.0107
CDC42	hsa04310	Wnt signaling pathway	3/5	0.00071
KTN1	hsa04810	Regulation of actin cytoskeleton	2/4	0.0128
ACTR3	hsa04810	Regulation of actin cytoskeleton	2/3	0.00713
RAC2	hsa05120	Epithelial cell signaling in Helicobacter pylori infection	2/2	0.00056
SUV39H1	hsa00271; hsa04110	Methionine metabolism; Cell cycle	2/13	NA; 0.0398
PHB	hsa04110	Cell cycle	2/7	0.0131
RBBP8	hsa04110	Cell cycle	2/3	0.00278
BAK1	hsa04210	Apoptosis	3/7	0.000565
PXN	hsa04650	Natural killer cell mediated cytotoxicity	6/16	2.12E-05
COL4A3	hsa01430; hsa04510; hsa04512	Cell Communication; Focal adhesion; ECM-receptor interaction	6/7	NA; 3.46e-07; 5.74e-09
COL7A1	hsa04350	TGF-beta signaling pathway	2/3	0.00177
SPARC	hsa01430; hsa04350; hsa04510; hsa04512; hsa04810	Cell Communication; TGF-beta signaling pathway; Focal adhesion; ECM-receptor interaction; Regulation of actin cytoskeleton	2/5	NA; 0.00436; 0.0174; 0.0042; 0.0194
CD36	hsa04670	Leukocyte transendothelial migration	3/5	0.000432

COL4A5	hsa01430; hsa04510; hsa04512	Cell Communication; Focal adhesion; ECM-receptor interaction	6/7	NA; 3.46e-07; 5.74e-09
CDK3	hsa04110	Cell cycle	2/3	0.00278
DSP	hsa01430	Cell Communication	2/2	NA
JUND	hsa04912	GnRH signaling pathway	2/4	0.0037
ERCC3	hsa03022	Basal transcription factors	2/5	0.0013
SUMO1	hsa04010; hsa04110; hsa05220	MAPK signaling pathway; Cell cycle; Chronic myeloid leukemia	2/6	0.0414; 0.00999; 0.0047
GNAZ	hsa04916	Melanogenesis	3/3	4.59E-05
ENG	hsa04060; hsa04350	Cytokine-cytokine receptor interaction; TGF-beta signaling pathway	3/7	0.00683; 0.00056
GRB10	hsa04650; hsa04910	Natural killer cell mediated cytotoxicity; Insulin signaling pathway	7/11	7.97e-08; 7.15e-08
RGS16	hsa04080	Neuroactive ligand-receptor interaction	5/14	0.00191
TRIP4	hsa04920	Adipocytokine signaling pathway	4/22	6.00E-04
POU2F1	hsa04620; hsa05120	Toll-like receptor signaling pathway; Epithelial cell signaling in Helicobacter pylori infection	3/14	0.00399; 0.0017
PELP1	hsa04110; hsa04310; hsa04510; hsa05210	Cell cycle; Wnt signaling pathway; Focal adhesion; Colorectal cancer	2/8	0.0167; 0.0255; 0.0414; 0.00975
MNAT1	hsa03022	Basal transcription factors	2/4	0.000882
NRIP1	hsa04010; hsa04920; hsa05220	MAPK signaling pathway; Adipocytokine signaling pathway; Chronic myeloid leukemia	2/11	0.123; 0.0114; 0.0137
BTK	hsa04660	T cell receptor signaling pathway	3/3	5.17E-05
CASP8AP2	hsa04210	Apoptosis	4/4	9.74E-07
GRAP2	hsa04664	Fc epsilon RI signaling pathway	5/7	2.97E-07
FOSL1	hsa04010	MAPK signaling pathway	2/5	0.0293
TRPV4	hsa04670	Leukocyte transendothelial migration	3/10	0.00253
NEDD9	hsa04510	Focal adhesion	7/13	2.27E-06
CD2	hsa04660	T cell receptor signaling pathway	5/6	3.46E-07
CASP7	hsa05210	Colorectal cancer	2/3	0.0017
INSR	hsa04630	Jak-STAT signaling pathway	8/17	2.82E-07
HCK	hsa04650; hsa04660; hsa04664	Natural killer cell mediated cytotoxicity; T cell receptor signaling pathway; Fc epsilon RI signaling pathway	3/4	0.000302; 0.000163; 7.19e-05
RAP1A	hsa04530; hsa04810	Tight junction; Regulation of actin cytoskeleton	3/5	0.00056; 0.00152
YES1	hsa04510; hsa04664	Focal adhesion; Fc epsilon RI signaling pathway	3/4	0.00068; 7.19e-05

NGB	hsa04080	Neuroactive ligand-receptor interaction	2/12	0.165
RGS14	hsa04080	Neuroactive ligand-receptor interaction	2/9	0.103
RIC8	hsa04080	Neuroactive ligand-receptor interaction	6/15	0.000497
UNC119	hsa04514	Cell adhesion molecules (CAMs)	2/3	0.00337
LILRB2	hsa04612	Antigen processing and presentation	2/3	0.0018
RHOA	hsa04010; hsa04370; hsa05120; hsa05212	MAPK signaling pathway; VEGF signaling pathway; Epithelial cell signaling in Helicobacter pylori infection; Pancreatic cancer	2/4	0.019; 0.00238; 0.00204; 0.00222
IGF1R	hsa04630	Jak-STAT signaling pathway	7/12	2.97E-07
SHC1	hsa04630	Jak-STAT signaling pathway	15/50	1.29E-09
SOCS2	hsa04920	Adipocytokine signaling pathway	3/3	1.66E-05
SOCS3	hsa04650	Natural killer cell mediated cytotoxicity	6/10	9.74E-07
PRKCD	hsa04010; hsa04020; hsa04070; hsa04310; hsa04370; hsa04510; hsa04540; hsa04650; hsa04670; hsa04720; hsa04730; hsa04916; hsa05214	MAPK signaling pathway; Calcium signaling pathway; Phosphatidylinositol signaling system; Wnt signaling pathway; VEGF signaling pathway; Focal adhesion; Gap junction; Natural killer cell mediated cytotoxicity; Leukocyte transendothelial migration; Long-term potentiation; Long-term depression; Melanogenesis; Glioma	3/3	0.000518; 0.000177; 2.15e-05; 0.000121; 2.12e-05; 0.000231; 3.3e-05; 9.95e-05; 6.5e-05; 1.96e-05; 1.94e-05; 4.59e-05; 1.54e-05
SERPINB6	hsa04610	Complement and coagulation cascades	2/2	0.000565
BIK	hsa04210	Apoptosis	3/5	0.000208
BCL2L11	hsa01510; hsa04210; hsa05030; hsa05210	Neurodegenerative Diseases; Apoptosis; Amyotrophic lateral sclerosis (ALS); Colorectal cancer	2/4	NA; 0.00306; NA; 0.00285
PMAIP1	hsa04210	Apoptosis	3/7	0.000565
TRAF5	hsa04010	MAPK signaling pathway	2/4	0.019
CD5	hsa04660	T cell receptor signaling pathway	6/10	2.97E-07
ASIP	hsa04080; hsa04920	Neuroactive ligand-receptor interaction; Adipocytokine signaling pathway	2/2	0.0051; 0.00056
MAP1A	hsa04020	Calcium signaling pathway	2/7	0.0265
CCBP2	hsa04060	Cytokine-cytokine receptor interaction	4/6	0.000432
JUNB	hsa04010	MAPK signaling pathway	3/4	0.00132
SP1	hsa05220	Chronic myeloid leukemia	4/15	0.000231
SKI	hsa04110; hsa04350	Cell cycle; TGF-beta signaling pathway	3/3	6.45e-05; 2.93e-05
MAP3K8	hsa04210;	Apoptosis;	2/2	0.000806;

	hsa04620; hsa04662; hsa04920; hsa05120; hsa05212; hsa05220	Toll-like receptor signaling pathway; B cell receptor signaling pathway; Adipocytokine signaling pathway; Epithelial cell signaling in Helicobacter pylori infection; Pancreatic cancer; Chronic myeloid leukemia		0.000974; 0.000613; 0.00056; 0.00056; 0.000591; 0.000613
BRAF	hsa04110; hsa04370	Cell cycle; VEGF signaling pathway	2/7	0.0131; 0.00627
RAP2A	hsa04010	MAPK signaling pathway	4/5	0.000171
BATF	hsa04010	MAPK signaling pathway	3/10	0.0172
TOB1	hsa05217	Basal cell carcinoma	2/5	0.00219
RASIP1	hsa04670; hsa05210; hsa05212	Leukocyte transendothelial migration; Colorectal cancer; Pancreatic cancer	2/6	0.0101; 0.00583; 0.00442
CRKL	hsa04650	Natural killer cell mediated cytotoxicity	14/43	8.51E-10
SLA	hsa04650	Natural killer cell mediated cytotoxicity	4/5	1.98E-05
UBE2L3	hsa04120	Ubiquitin mediated proteolysis	2/5	0.00918
HDAC3	hsa04330	Notch signaling pathway	3/16	0.00108
KCNA4	hsa04020	Calcium signaling pathway	3/8	0.00353
KSR2	hsa04010	MAPK signaling pathway	4/5	0.000171
PRKCG	hsa04664; hsa04912	Fc epsilon RI signaling pathway; GnRH signaling pathway	3/3	2.26e-05; 4.43e-05
RHOG	hsa04360; hsa04510; hsa04520; hsa04670; hsa04810	Axon guidance; Focal adhesion; Adherens junction; Leukocyte transendothelial migration; Regulation of actin cytoskeleton	3/3	7.91e-05; 0.000231; 2.2e-05; 6.5e-05; 0.000276
NAPA	hsa04130	SNARE interactions in vesicular transport	4/6	6.16E-07
ERBB3	hsa04320; hsa04510	Dorso-ventral axis formation; Focal adhesion	3/3	NA; 0.000231
STRAP	hsa04350	TGF-beta signaling pathway	2/5	0.00436
TNFAIP3	hsa04010	MAPK signaling pathway	5/15	0.0018
ZNF8	hsa05217	Basal cell carcinoma	2/5	0.00219
ALS2CR2	hsa04010; hsa04620	MAPK signaling pathway; Toll-like receptor signaling pathway	2/2	0.00436; 0.000974
CASP9	hsa01510; hsa05010; hsa05050	Neurodegenerative Diseases; Alzheimer's disease; Dentatorubropallidoluysian atrophy (DRPLA)	2/3	NA; 0.00483; NA
DIABLO	hsa04210	Apoptosis	3/4	9.95E-05
KCNJ12	hsa04020	Calcium signaling pathway	5/17	0.000699
MAP3K14	hsa04060	Cytokine-cytokine receptor interaction	6/17	0.000608
TRPC4AP	hsa04210	Apoptosis	8/13	8.51E-10
TANK	hsa04060	Cytokine-cytokine receptor interaction	5/9	0.000208
SH3BP5	hsa04010	MAPK signaling pathway	2/3	0.0107

SIN3A	hsa04110; hsa04330; hsa05220	Cell cycle; Notch signaling pathway; Chronic myeloid leukemia	3/16	0.00782; 0.00108; 0.00285
E2F2	hsa04110; hsa05212; hsa05214; hsa05218; hsa05220	Cell cycle; Pancreatic cancer; Glioma; Melanoma; Chronic myeloid leukemia	2/3	0.00278; 0.00132; 0.00116; 0.00131; 0.0014
PLCG2	hsa04660	T cell receptor signaling pathway	6/8	5.76E-08
MCL1	hsa01510; hsa04210; hsa05030	Neurodegenerative Diseases; Apoptosis; Amyotrophic lateral sclerosis (ALS)	2/3	NA; 0.0018; NA
AHR	hsa03320; hsa04010; hsa04620; hsa04660; hsa04662; hsa04920; hsa05210	PPAR signaling pathway; MAPK signaling pathway; Toll-like receptor signaling pathway; T cell receptor signaling pathway; B cell receptor signaling pathway; Adipocytokine signaling pathway; Colorectal cancer	2/13	0.0162; 0.161; 0.0303; 0.034; 0.0185; 0.0154; 0.0226
HNF4A	hsa04080	Neuroactive ligand-receptor interaction	2/7	0.0657
PPARBP	hsa04920	Adipocytokine signaling pathway	3/23	0.00519
DUSP22	hsa04010	MAPK signaling pathway	2/3	0.0107
ID2	hsa04020; hsa04070; hsa04720; hsa04740; hsa04910; hsa04912; hsa04916; hsa05040; hsa05214	Calcium signaling pathway; Phosphatidylinositol signaling system; Long-term potentiation; Olfactory transduction; Insulin signaling pathway; GnRH signaling pathway; Melanogenesis; Huntington's disease; Glioma	2/6	0.0199; 0.00481; 0.00442; 0.0769; 0.0128; 0.00766; 0.00791; NA; 0.00384
PPARD	hsa04920	Adipocytokine signaling pathway	2/5	0.00295
SH3BP2	hsa04660	T cell receptor signaling pathway	7/12	3.99E-08
KCNJ4	hsa04020	Calcium signaling pathway	3/10	0.00627
BIRC7	hsa04210; hsa04510	Apoptosis; Focal adhesion	3/4	9.95e-05; 0.00068
SNIP1	hsa05217	Basal cell carcinoma	2/3	0.000896
CRIP1	hsa04020	Calcium signaling pathway	4/14	0.00224
SEMA4C	hsa00230	Purine metabolism	2/7	0.0199
GDA	hsa04020	Calcium signaling pathway	3/8	0.00353
KCNJ10	hsa04020; hsa04080; hsa04720	Calcium signaling pathway; Neuroactive ligand-receptor interaction; Long-term potentiation	3/5	0.00103; 0.00322; 0.00013
PGF	hsa04060	Cytokine-cytokine receptor interaction	2/3	0.0109

SIN3B	hsa04330; hsa05220	Notch signaling pathway; Chronic myeloid leukemia	3/4	2.14e-05; 6.57e-05
ZNFN1A1	hsa04330; hsa05220	Notch signaling pathway; Chronic myeloid leukemia	3/8	0.000185; 0.00056
THRB	hsa03320	PPAR signaling pathway	3/8	0.000476
USP7	hsa04060	Cytokine-cytokine receptor interaction	11/15	9.44E-10
TRIM37	hsa04060	Cytokine-cytokine receptor interaction	5/13	0.00109
MAP4K1	hsa04910	Insulin signaling pathway	3/5	0.000588
DEDD	hsa04210	Apoptosis	4/5	4.22E-06
CDC5L	hsa04110	Cell cycle	4/5	1.09E-05
KIF1C	hsa04010	MAPK signaling pathway	2/7	0.0553
DAB1	hsa05010	Alzheimer's disease	2/4	0.00865
RIPK2	hsa04210	Apoptosis	4/12	0.000182
ACTR2	hsa04810	Regulation of actin cytoskeleton	5/6	6.83E-06
TP73	hsa04110; hsa05212; hsa05214; hsa05218; hsa05220	Cell cycle; Pancreatic cancer; Glioma; Melanoma; Chronic myeloid leukemia	2/2	0.00122; 0.000591; 0.000542; 0.000588; 0.000613
DMTF1	hsa04110; hsa05220	Cell cycle; Chronic myeloid leukemia	3/5	0.000425;0. 000144
DIPA	hsa01430	Cell Communication	2/19	NA
USHBP1	hsa01430	Cell Communication	3/4	NA
CCNH	hsa00500; hsa00790	Starch and sucrose metabolism; Folate biosynthesis	2/4	0.00138; 0.000147
HOOK2	hsa01430	Cell Communication	2/9	NA
ESR2	hsa03320; hsa05216	PPAR signaling pathway; Thyroid cancer	2/8	0.00691; 0.00186
IGF2	hsa04150	mTOR signaling pathway	2/3	0.000829
ATF3	hsa04010	MAPK signaling pathway	3/8	0.00955
FOSL2	hsa04010	MAPK signaling pathway	3/8	0.00955
UCP2	hsa04010; hsa04510; hsa04910; hsa05210; hsa05212; hsa05218; hsa05220	MAPK signaling pathway; Focal adhesion; Insulin signaling pathway; Colorectal cancer; Pancreatic cancer; Melanoma; Chronic myeloid leukemia	2/5	0.0293; 0.0174; 0.00925; 0.00422; 0.00327; 0.00322; 0.0035
UCP3	hsa04010; hsa04510; hsa04910; hsa05210; hsa05212; hsa05218;	MAPK signaling pathway; Focal adhesion; Insulin signaling pathway; Colorectal cancer; Pancreatic cancer; Melanoma;	2/5	0.0293; 0.0174; 0.00925; 0.00422; 0.00327; 0.00322;

	hsa05220	Chronic myeloid leukemia		0.0035
TEF	hsa04010	MAPK signaling pathway	3/7	0.00662
HLF	hsa04010	MAPK signaling pathway	2/8	0.0705
KCNQ2	hsa04080	Neuroactive ligand-receptor interaction	2/17	0.279
NID	hsa04512	ECM-receptor interaction	2/3	0.00169
IRS1	hsa04510	Focal adhesion	8/21	7.37E-06
CBL	hsa04650	Natural killer cell mediated cytotoxicity	17/62	2.09E-10
RALGDS	hsa04670	Leukocyte transendothelial migration	2/5	0.00713
TRAF3IP2	hsa04010; hsa04210; hsa04660; hsa05120	MAPK signaling pathway; Apoptosis; T cell receptor signaling pathway; Epithelial cell signaling in Helicobacter pylori infection	2/8	0.0705; 0.0105; 0.0143; 0.00677
LILRB1	hsa04612	Antigen processing and presentation	2/3	0.0018
NOG	hsa04060	Cytokine-cytokine receptor interaction	2/3	0.0109
CDC25B	hsa04910	Insulin signaling pathway	4/13	0.000868
CDH2	hsa04520; hsa05130; hsa05131; hsa05216	Adherens junction; Pathogenic Escherichia coli infection - EHEC; Pathogenic Escherichia coli infection - EPEC; Thyroid cancer	2/3	0.00145; 0.000851; NA; 0.00036
SKIL	hsa04110	Cell cycle	3/8	0.00141
GUCY1A2	hsa04020	Calcium signaling pathway	4/15	0.00281
NFE2L1	hsa04010	MAPK signaling pathway	2/6	0.0414
ZIC1	hsa04340	Hedgehog signaling pathway	3/3	1.09E-05
ZIC2	hsa05217	Basal cell carcinoma	2/3	0.000896
GMFB	hsa04010	MAPK signaling pathway	6/6	3.44E-07
STXBP6	hsa04130	SNARE interactions in vesicular transport	2/2	0.000219
CCL3L1	hsa04060	Cytokine-cytokine receptor interaction	4/4	4.85E-05
Rasd2 [¶]	mmu05212	Pancreatic cancer	4/24	NA
Rps27a [¶]	mmu04620	Toll-like receptor signaling pathway	2/15	NA
Sqstm1 [¶]	mmu04010; mmu04810	MAPK signaling pathway; Regulation of actin cytoskeleton	2/13	NA; NA
Map2k3 [¶]	mmu04810; mmu05212	Regulation of actin cytoskeleton; Pancreatic cancer	2/13	NA; NA
Uhmk1 [¶]	mmu04620	Toll-like receptor signaling pathway	2/15	NA
Rhod [¶]	mmu04010; mmu04620	MAPK signaling pathway; Toll-like receptor signaling pathway	2/15	NA; NA
CCDC5	hsa04110	Cell cycle	5/6	4.91E-07
Protein	GO ID	GO Term	Ratio	P value
APP	GO:0005509	calcium ion binding	4/9	0.00148
DLG4	GO:0004385	guanylate kinase activity	4/11	1.13E-09
NCOA1	GO:0008270	zinc ion binding	13/38	0.00148
KHDRBS1	GO:0005524	ATP binding	5/18	0.0151

RELA	GO:0008270	zinc ion binding	11/31	0.00239
PTPN11	GO:0005524	ATP binding	16/50	5.53E-06
GHR	GO:0005524	ATP binding	14/33	6.30E-07
IL2RB	GO:0005524	ATP binding	8/17	7.44E-05
JUN	GO:0008270	zinc ion binding	12/35	0.00217
SAA1	GO:0005509	calcium ion binding	4/11	0.00307
CD19	GO:0005524	ATP binding	4/10	0.00839
CD22	GO:0005524	ATP binding	9/23	0.000133
PIK3R1	GO:0005524	ATP binding	21/66	2.50E-07
PTPN6	GO:0005524	ATP binding	11/35	0.000201
TRAF1	GO:0005524; GO:0043123	ATP binding; positive regulation of I-kappaB kinase/NF-kappaB cascade	4/10	0.00839; 7.74e-07
PTPN1	GO:0005159	insulin-like growth factor receptor binding	4/14	4.29E-09
GRB2	GO:0005524	ATP binding	23/61	2.24E-09
BLNK	GO:0005070	SH3/SH2 adaptor activity	4/15	4.29E-07
IRS2	GO:0005524	ATP binding	7/27	0.00746
EPOR	GO:0005524	ATP binding	12/38	9.70E-05
GAB2	GO:0005070; GO:0005524	SH3/SH2 adaptor activity; ATP binding	6/34	5.44e-09; 0.0587
PTPN12	GO:0005524	ATP binding	7/15	0.000228
PTPRC	GO:0005524	ATP binding	4/15	0.0317
CD2AP	GO:0005524	ATP binding	4/10	0.00839
VAV1	GO:0005524	ATP binding	14/43	1.63E-05
PAG1	GO:0005524	ATP binding	10/45	0.00495
CD28	GO:0005070	SH3/SH2 adaptor activity	4/12	1.72E-07
RAPGEF1	GO:0005524	ATP binding	5/10	0.00144
GAB3	GO:0005524	ATP binding	6/21	0.0079
TUB	GO:0005524	ATP binding	4/8	0.00402
SOS1	GO:0005524	ATP binding	6/19	0.00501
PLCG1	GO:0005524	ATP binding	13/49	0.000309
BCAR1	GO:0005524	ATP binding	6/28	0.0266
RASA1	GO:0005524	ATP binding	6/22	0.00938
SOCS1	GO:0005524	ATP binding	6/23	0.0115
VAV3	GO:0005524	ATP binding	5/17	0.0122
SHB	GO:0005524	ATP binding	7/25	0.00501
CRK	GO:0005524	ATP binding	4/13	0.0198
BCR	GO:0005524	ATP binding	6/17	0.00307
TBP	GO:0008270	zinc ion binding	7/20	0.0125
REM1	GO:0005524	ATP binding	4/11	0.0116
RBBP4	GO:0008270	zinc ion binding	7/17	0.00555
RICS	GO:0005524	ATP binding	5/15	0.00784
GNAI1	GO:0003924	GTPase activity	4/4	9.45E-08
FOS	GO:0008270	zinc ion binding	6/19	0.0313

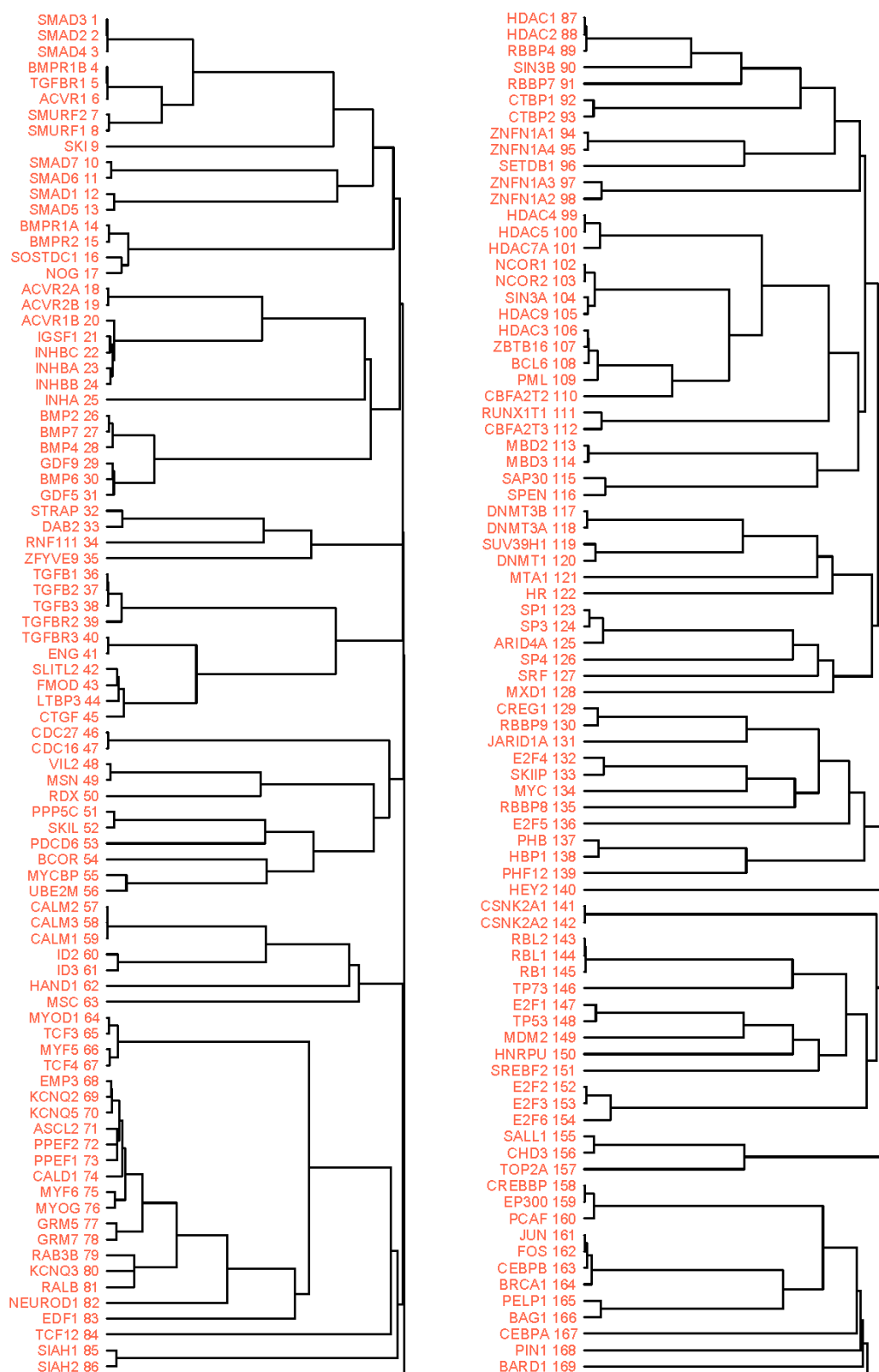
PECAM1	GO:0005524	ATP binding	9/26	0.00036
FN1	GO:0005509	calcium ion binding	6/18	0.000478
RB1	GO:0008270	zinc ion binding	6/16	0.0144
GADD45G	GO:0030521	androgen receptor signaling pathway	4/8	1.22E-08
MCM10	GO:0008270	zinc ion binding	10/26	0.00206
ORC2L	GO:0005524	ATP binding	8/11	1.25E-06
ASK	GO:0005524	ATP binding	6/8	2.24E-05
GTF2H1	GO:0005524	ATP binding	4/9	0.00602
HDAC1	GO:0008270	zinc ion binding	8/24	0.011
PXN	GO:0005524	ATP binding	5/16	0.00966
DCN	GO:0005509	calcium ion binding	5/15	0.00148
TGFB1	GO:0005509	calcium ion binding	4/13	0.00523
OSM	GO:0005509	calcium ion binding	4/12	0.00402
ANTXR2	GO:0005509	calcium ion binding	4/10	0.00221
HABP2	GO:0005509	calcium ion binding	4/12	0.00402
COL1A1	GO:0005587	collagen type IV	6/12	8.49E-16
COL1A2	GO:0005587	collagen type IV	6/9	1.16E-16
POU2F1	GO:0008270	zinc ion binding	5/14	0.0296
FAS	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	5/6	6.74E-10
CASP10	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	4/6	8.47E-08
TIF1	GO:0003714	transcription corepressor activity	4/12	5.50E-06
SHC1	GO:0005524	ATP binding	16/50	5.53E-06
SOCS3	GO:0005159	insulin-like growth factor receptor binding	4/10	1.17E-09
IL4R	GO:0005524; GO:0043560	ATP binding; insulin receptor substrate binding	4/12	0.0151; 9.39e-10
PIK3R2	GO:0005524	ATP binding	4/10	0.00839
LCP2	GO:0005524	ATP binding	7/25	0.00501
GTF2E2	GO:0005524	ATP binding	6/9	6.09E-05
CRKL	GO:0005524	ATP binding	16/43	6.86E-07
HDAC3	GO:0008270	zinc ion binding	5/16	0.0494
SERPINE2	GO:0005509	calcium ion binding	4/12	0.00402
ARID4A	GO:0008270	zinc ion binding	4/5	0.00239
TNFAIP3	GO:0005524	ATP binding	7/15	0.000228
KCNJ12	GO:0005509; GO:0005524	calcium ion binding; ATP binding	5/17	0.00248; 0.0122
MAP3K14	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	7/17	8.18E-11
SIN3A	GO:0003714; GO:0008270	transcription corepressor activity; zinc ion binding	4/16	1.57e-05; 0.146
NCOA2	GO:0008270; GO:0030521	zinc ion binding; androgen receptor signaling pathway	5/12	0.0151; 7.33e-10
AHR	GO:0008270	zinc ion binding	7/13	0.00118

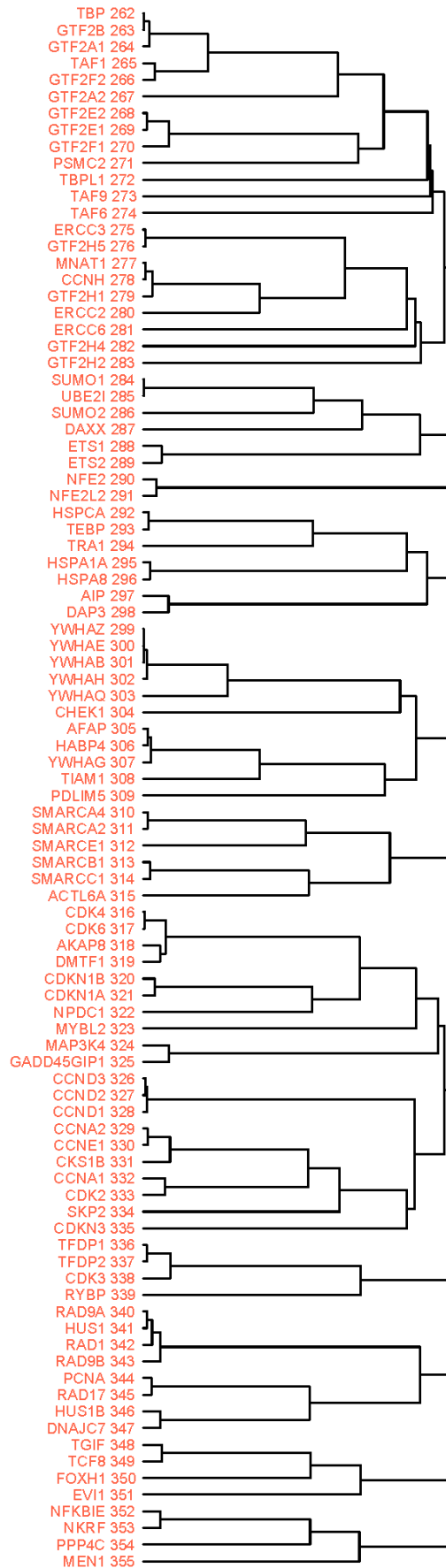
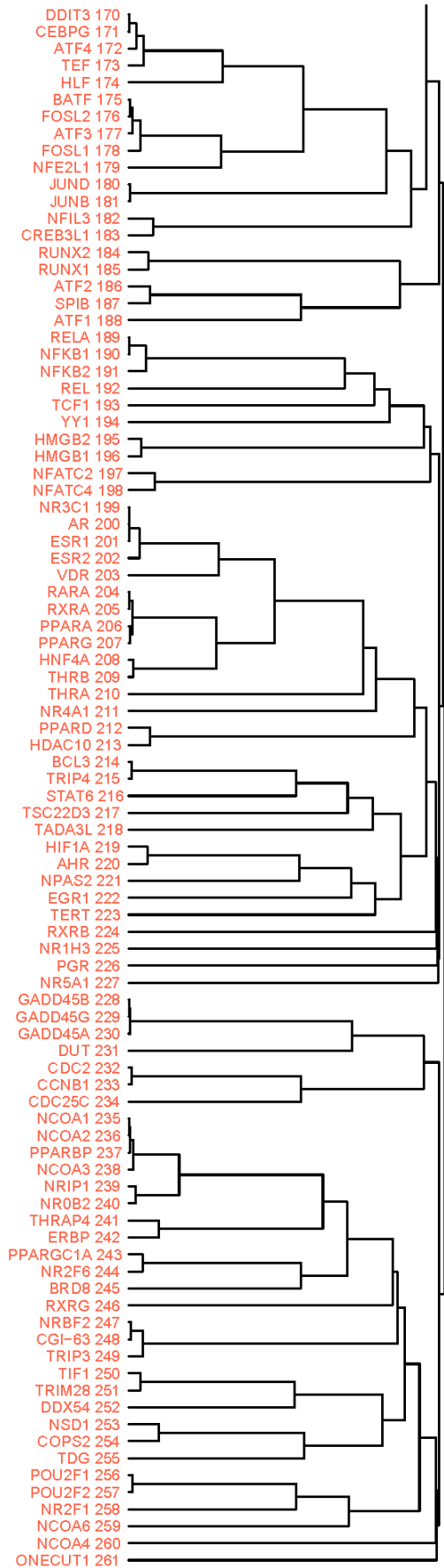
PPARBP	GO:0008270	zinc ion binding	9/23	0.00285
KCNJ4	GO:0000287; GO:0005509	magnesium ion binding; calcium ion binding	4/10	0.000201; 0.00221
CRIP1	GO:0005509	calcium ion binding	4/14	0.00678
CDC45L	GO:0005524	ATP binding	4/8	0.00402
ORC3L	GO:0005524	ATP binding	4/5	0.000483
HDAC9	GO:0008270	zinc ion binding	5/12	0.0151
NR0B2	GO:0030521; GO:0050681	androgen receptor signaling pathway; androgen receptor binding	4/10	3.44e-08; 1.02e-08
USP7	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	4/15	4.29E-06
TNFRSF19	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	4/19	1.00E-05
ACTR2	GO:0005885	Arp2/3 protein complex	6/6	5.13E-18
RIBC2	GO:0008270	zinc ion binding	4/14	0.101
FLJ32855	GO:0008270	zinc ion binding	6/19	0.0313
TU3A	GO:0008270	zinc ion binding	9/27	0.00784
UTP14A	GO:0008270	zinc ion binding	4/7	0.00937
C1orf65	GO:0008270	zinc ion binding	4/8	0.0151
WAS	GO:0005524	ATP binding	7/23	0.00333
IRS1	GO:0005524	ATP binding	8/21	0.000381
CBL	GO:0005524	ATP binding	18/62	6.08E-06
TNFRSF17	GO:0005031	tumor necrosis factor receptor activity	4/11	7.33E-10
ATP2B4	GO:0030955	potassium ion binding	4/16	1.52E-05
TRAF3IP2	GO:0005524	ATP binding	4/8	0.00402
CDC25B	GO:0005524	ATP binding	4/13	0.0198
GUCY1A2	GO:0005509	calcium ion binding	4/15	0.00838
TNFRSF8	GO:0005031	tumor necrosis factor receptor activity	4/14	1.38E-09
TNFRSF11 A	GO:0005524; GO:0043123	ATP binding; positive regulation of I-kappaB kinase/NF-kappaB cascade	4/17	0.0481; 6.39e-06
Rps27a [¶]	GO:0005525	GTP binding	6/15	1.47E-06
Fbxo3 [¶]	GO:0005525	GTP binding	6/13	6.30E-07
Sqstm1 [¶]	GO:0005525	GTP binding	6/13	6.30E-07
Map2k3 [¶]	GO:0005525	GTP binding	5/13	1.45E-05
Uhmk1 [¶]	GO:0005525	GTP binding	6/15	1.47E-06
KIAA1267	GO:0008270	zinc ion binding	4/15	0.123
CCDC5	GO:0005524	ATP binding	4/6	0.00131

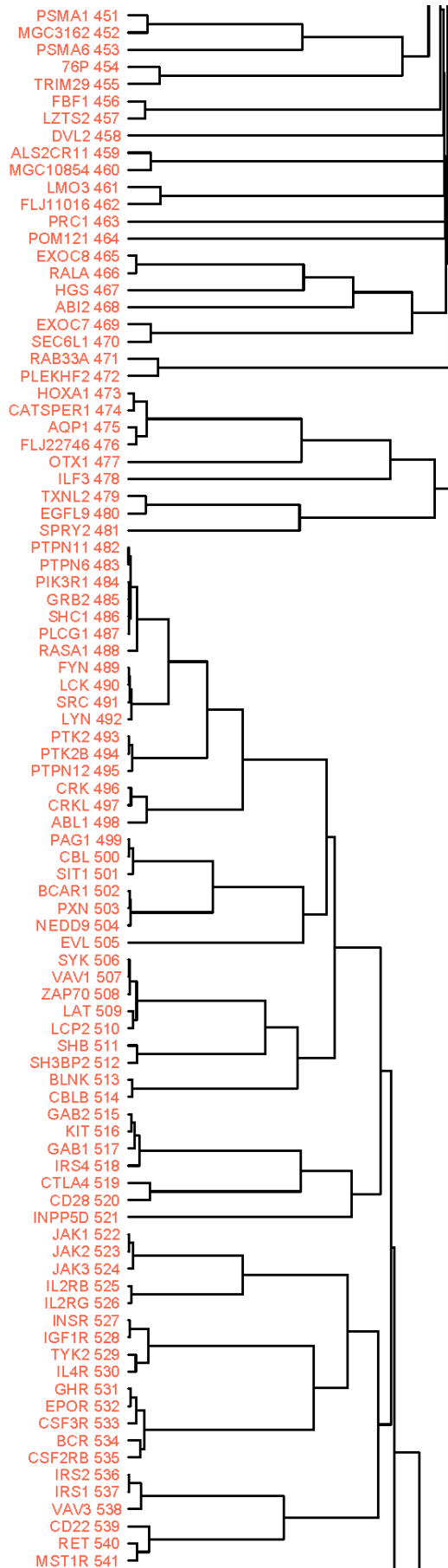
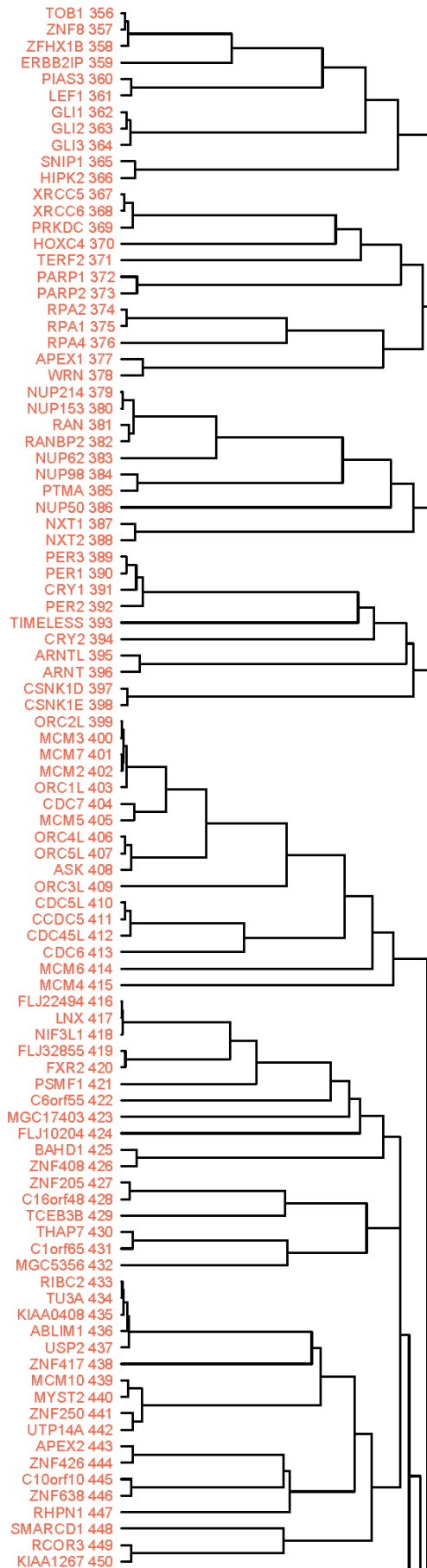
[¶] indicates mouse proteins.

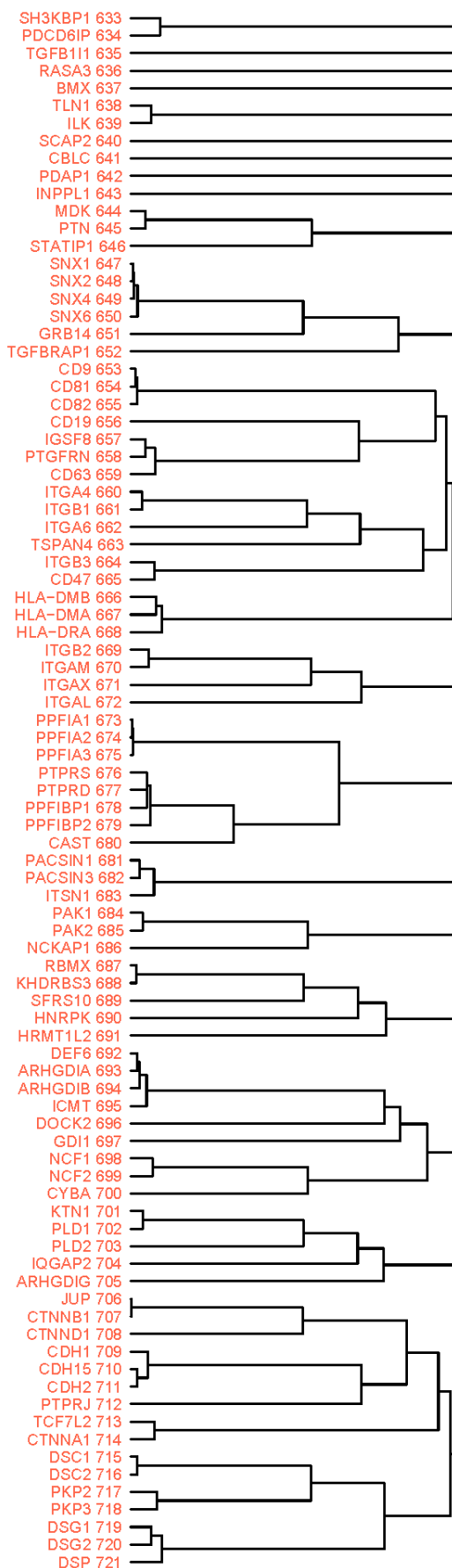
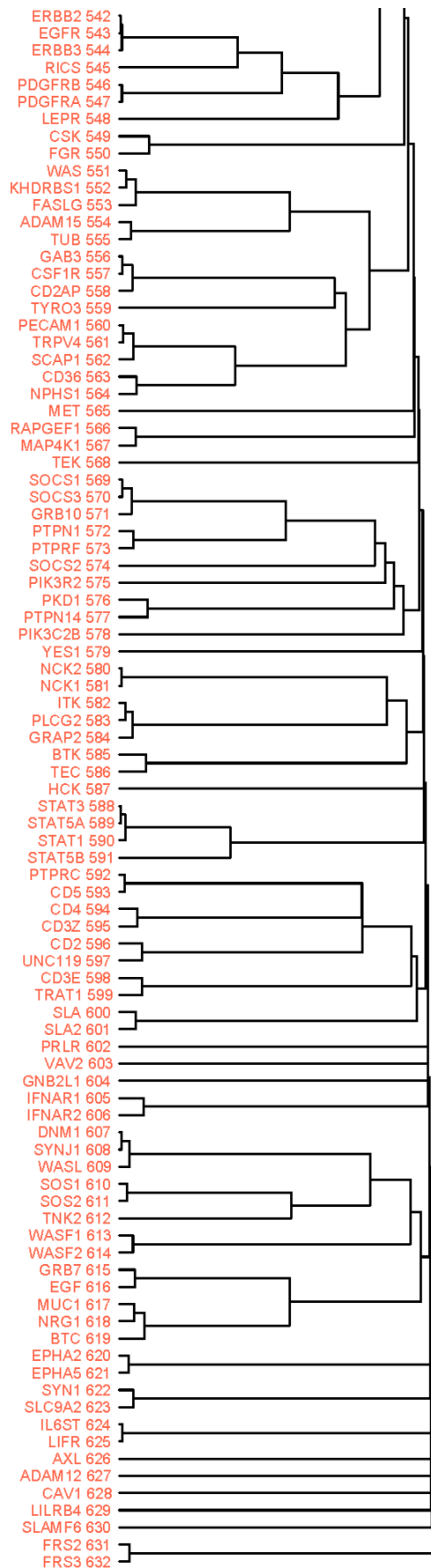
Appendix C

Figure S1 The whole cluster that consists of 1,729 human proteins. Indices above protein names are their coordinates in this cluster. Fig. S1 was reproduced from REF. 76.



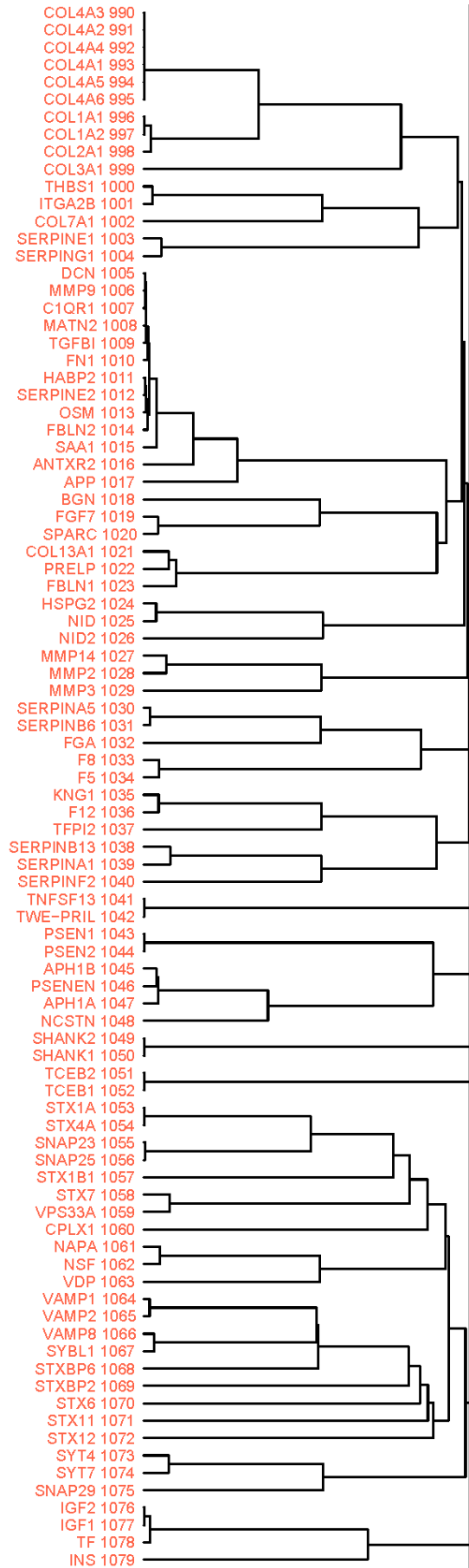
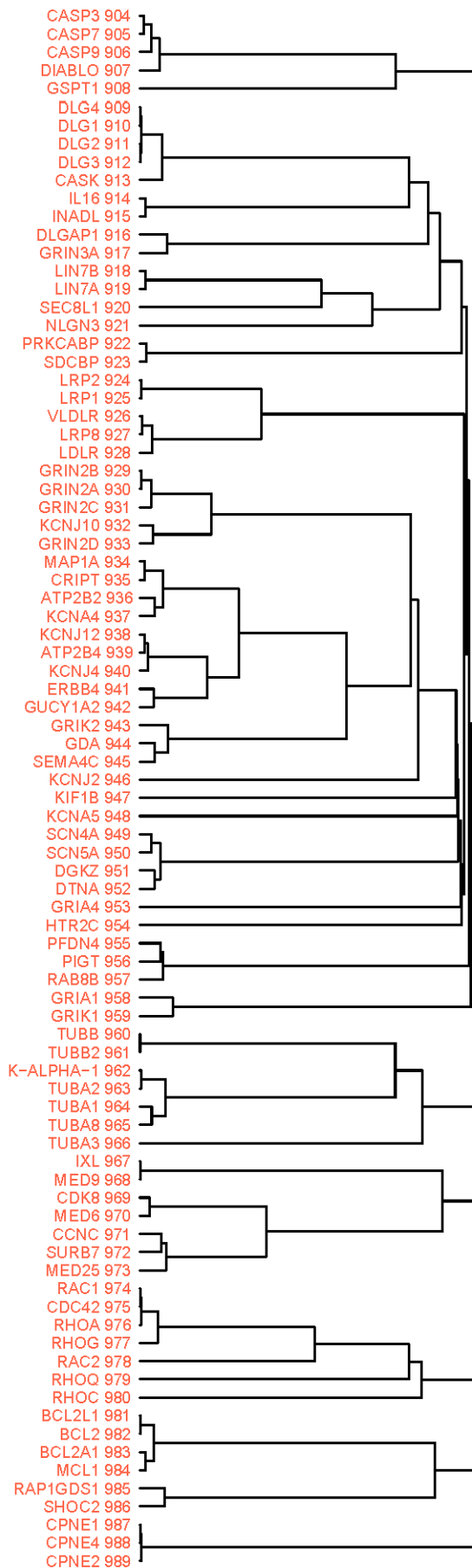


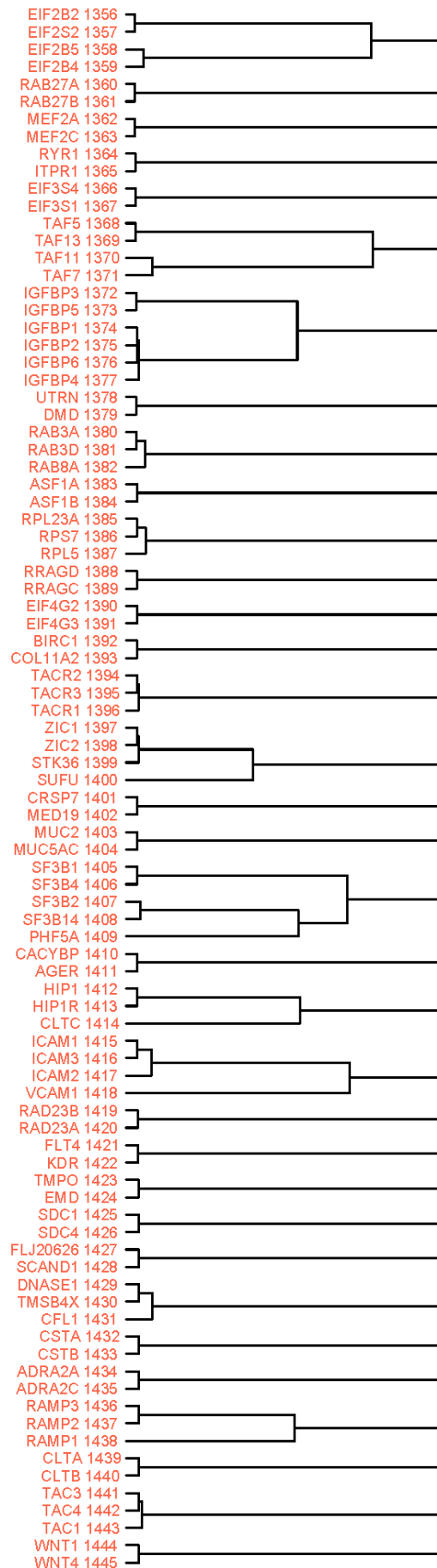
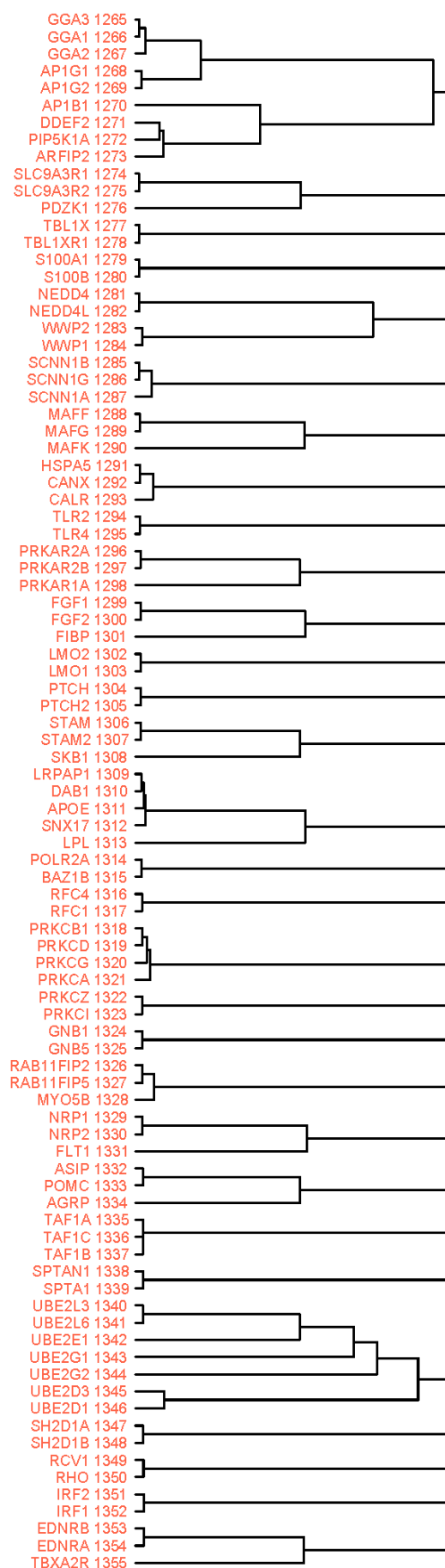


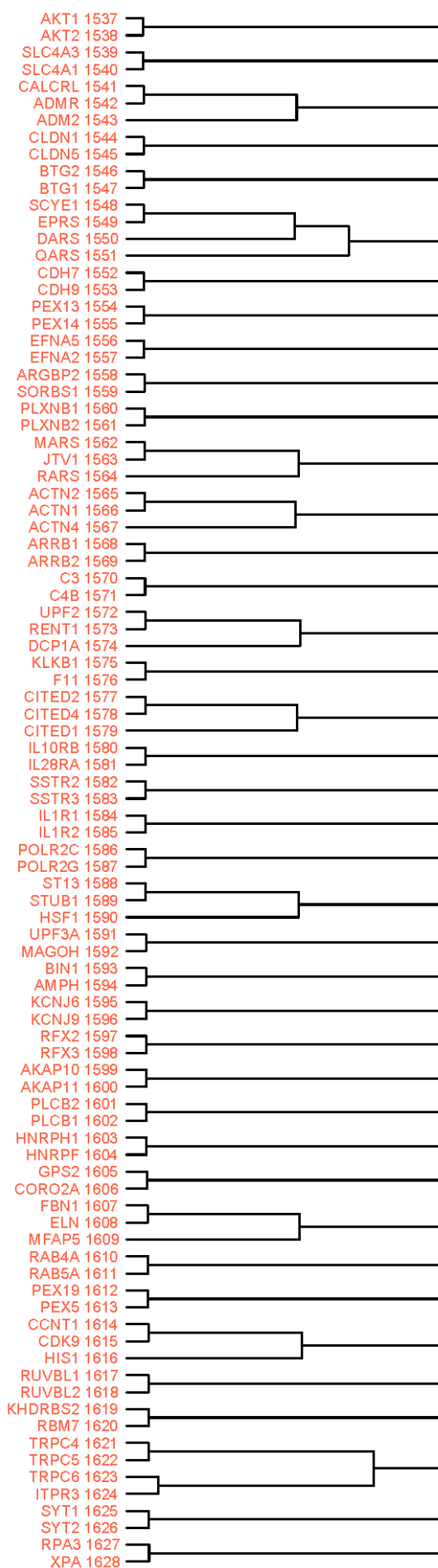
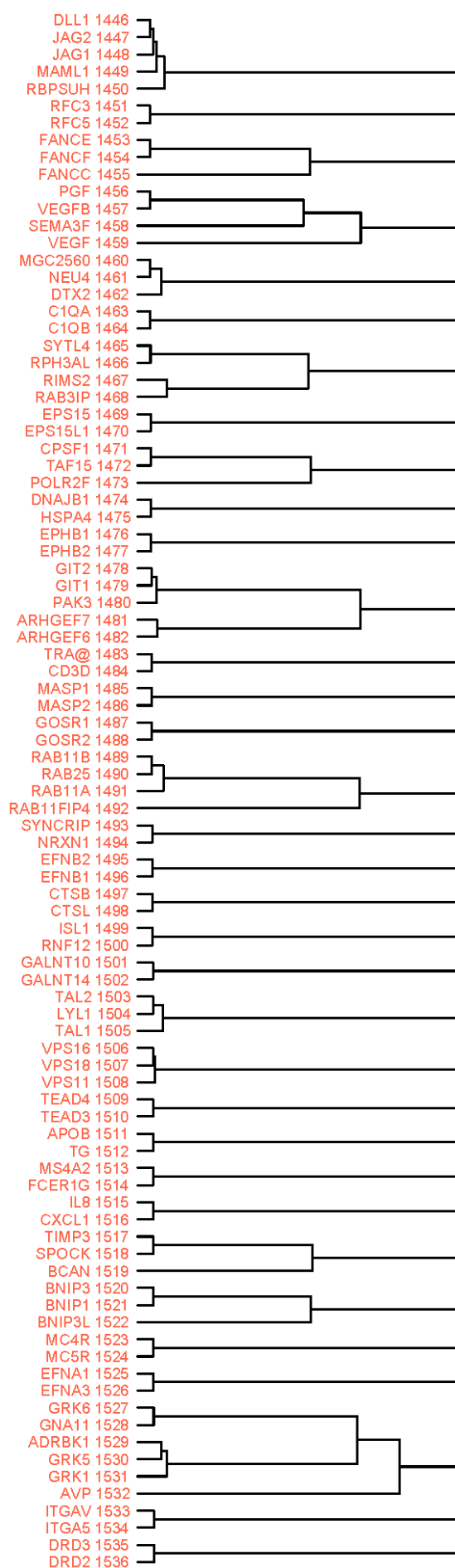


MAPK1 722
 MAPK3 723
 MAPK14 724
 MAPK8 725
 MAPK9 726
 MAPK10 727
 MAP2K4 728
 MAP2K7 729
 MAPK8IP3 730
 DUSP16 731
 DUSP10 732
 SH3BP5 733
 PRKD1 734
 MAP3K1 735
 MAP3K2 736
 MAP3K11 737
 DUSP22 738
 SPAG9 739
 MAP2K1 740
 MAP2K2 741
 KSR2 742
 MAP2K1IP1 743
 RPS6KA2 744
 PBP 745
 PTPN7 746
 DUSP4 747
 GMFB 748
 DUSP1 749
 ELK1 750
 RPS6KA4 751
 TNFSF11 752
 SNCG 753
 MRAS 754
 RAP2A 755
 RRAS2 756
 RRAS 757
 HRAS 758
 RAP1A 759
 RAP2B 760
 NRAS 761
 KRAS 762
 RASIP1 763
 RALGDS 764
 MLLT4 765
 RASSF5 766
 RAPGEF5 767
 RAPGEF6 768
 BAK1 769
 PMAIP1 770
 BAD 771
 BAX 772
 BCL2L1 773
 BIK 774
 BID 775
 MOAP1 776
 RAF1 777
 BRAF 778
 REM1 779
 CDC25B 780
 UCP2 781
 UCP3 782
 CDC25A 783
 RIN1 784
 KRT18 785
 ZFP36 786
 PCTK1 787
 KIF1C 788
 EPB41L3 789
 MLLT7 790
 FOXO1A 791
 TRAF2 792
 TRAF1 793
 TRAF3 794
 TRAF5 795
 TRAF6 796
 IKBKB 797
 CHUK 798
 IKBKG 799
 MAP3K8 800
 MAP3K14 801
 TRAF3IP2 802
 TNFAIP3 803
 TANK 804
 TNFRSF1A 805
 TRPC4AP 806
 TTRAP 807
 RIPK2 808

RIPK1 809
 TRADD 810
 CASP8AP2 811
 CASP8 812
 CFLAR 813
 FADD 814
 CASP10 815
 DEDD 816
 TNFRSF10A 817
 TNFRSF10B 818
 FAS 819
 DEDD2 820
 TNFRSF19 821
 TNFRSF11A 822
 USP7 823
 TRIM37 824
 CD40 825
 TNFRSF8 826
 LTBR 827
 TNFRSF4 828
 TNFRSF7 829
 TNFRSF14 830
 TNFRSF17 831
 TNFRSF18 832
 TNFRSF12A 833
 TNFRSF9 834
 TNFRSF13B 835
 TNFRSF1B 836
 MAP3K5 837
 NR2C1 838
 IRAK1 839
 CASP4 840
 CASP2 841
 RIPK3 842
 CASP1 843
 TAB3 844
 CASP14 845
 CARD4 846
 ZNF297B 847
 Cep72 848
 TRIP 849
 HIVEP3 850
 MAP2K3 851
 MAP2K6 852
 ALS2CR2 853
 MAP3K7IP2 854
 MAP3K7IP1 855
 MAP3K7 856
 BIRC3 857
 BIRC2 858
 BIRC4 859
 BIRC5 860
 BIRC7 861
 GORASP1 862
 MDF1 863
 KRTAP4-12 864
 PLSCR1 865
 TSR1 866
 EFEMP2 867
 PEPP-2 868
 RBPM5 869
 KRT15 870
 DIPA 871
 USHBP1 872
 VIM 873
 KRT19 874
 KIAA0980 875
 HOOK2 876
 KIAA1446 877
 NDP52 878
 Cep70 879
 LDOC1 880
 DPPA2 881
 ZBTB8 882
 THAP1 883
 TSC2D4 884
 EFCBP2 885
 RAD54B 886
 SSSCA1 887
 GFAP 888
 KRT20 889
 RABAC1 890
 RPIA 891
 PCBD1 892
 DCTD 893
 PAICS 894
 CUTC 895
 KCTD17 896
 TIFA 897
 CHML 898
 RTN4 899
 RTN3 900
 MGC2749 901
 LOC138046 902
 HNRPC 903







EIF3S12	1629	
EIF3S7	1630	
CUL1	1631	
SKP1A	1632	
TUBG1	1633	
TUBG2	1634	
PTPRZ1	1635	
PTPRB	1636	
SH3GL3	1637	
SH3GL2	1638	
SH3GL1	1639	
CCT5	1640	
CCT4	1641	
JM11	1642	
FLJ38984	1643	
NTRK1	1644	
NTRK2	1645	
ADRB2	1646	
OPRK1	1647	
CFTR	1648	
CLCN3	1649	
RNPS1	1650	
SFRS4	1651	
MKNK1	1652	
DYRK1B	1653	
RNF8	1654	
UBOX5	1655	
YAP1	1656	
VGLL1	1657	
PDGFA	1658	
PDGFB	1659	
PARC	1660	
UBE3A	1661	
TRIM39	1662	
RNF126	1663	
FANCA	1664	
FANCD2	1665	
AKAP5	1666	
AKAP12	1667	
MYOZ2	1668	
MYOZ1	1669	
PDLIM1	1670	
SIGIRR	1671	
TRIAD3	1672	
PNKP	1673	
ZNF655	1674	
PFN1	1675	
GSN	1676	
KIF13A	1677	
AP1M2	1678	
TES	1679	
ZYX	1680	
CTNND2	1681	
PKP4	1682	
CHEK2	1683	
CLSPN	1684	
TNNT1	1685	
MBIP	1686	
ITGA2	1687	
ITGA1	1688	
ARL2	1689	
ARL3	1690	
EPB41L1	1691	
EPB41L2	1692	
KIAA1217	1693	
FLJ12529	1694	
MAGEA11	1695	
RP11-301117.1	1696	
HOMER2	1697	
HOMER1	1698	
RAD51	1699	
RAD52	1700	
RYR2	1701	
HOMER3	1702	
CSPG3	1703	
CNTN1	1704	
MAPT	1705	
MAP4	1706	
CDH5	1707	
DSC3	1708	
REPS1	1709	
RALBP1	1710	
EIF3S8	1711	
EIF3S10	1712	
ZNF24	1713	
ZNF165	1714	
KCNE2	1715	
KCNE1	1716	
IHH	1717	
DHH	1718	
SMO	1719	
UCN3	1720	
CRH	1721	
PRKAB1	1722	
PRKAB2	1723	
TRIM54	1724	
USP13	1725	
ERO1L	1726	
ERO1LB	1727	
COG2	1728	
COG1	1729	

VITA

Hua Li, the son of Mengjun Li and Min Hua, was born in 1982 in the People's Republic of China. He attended Suzhou Middle School, Suzhou, Jiangsu Province, China. In 2001 he entered Shanghai Jiao Tong University, Shanghai, China. He received the degree of Bachelor of Science from Shanghai Jiao Tong University in June, 2005. He entered the Graduate School of Biomedical Sciences, the University of Texas Health Science Center at Houston and The University of Texas M.D. Anderson Cancer Center in August, 2005.

Permanent Address:

NO1 DongYi Rd, Suzhou, Jiangsu Province, China, 215007.

Bldg 1, Rm 103.

This manuscript was typed by the author.