Materials and Methods

PPI, DDI datasets and network construction

Human protein-protein interaction(Ashburner et al.) data were taken from a previous study(Wang et al., 2012) which integrated PPIs from the Human Protein Reference Database (HPRD, release 9)(Keshava Prasad et al., 2009), BioGrid (version 3.0.66)(Breitkreutz et al., 2008), IntAct (2010)(Ceol et al., 2010), Molecular Interaction Database (MINT, 2010)(Ceol et al., 2010), VisANT (2010)(Ceol et al., 2010), iRefWeb (version 3.9)(Turner et al., 2010) and a Y2H interactions(Rual et al., 2005; Stelzl et al., 2005; Venkatesan et al., 2009; Yu et al., 2011). After removing the self- and duplicated interactions, 19,139 PPIs were included in the network.

Pfam domains for all proteins were assigned using Pfam HMM models. Domain-domain interactions (DDIs) were retrieved from iPfam(Finn et al., 2005), 3DID(Stein et al., 2005) and DOMINE(Yellaboina et al., 2011) database. DDIs from iPfam and 3DID are based on resolved 3D structural information in Protein DataBank(Berman et al., 2000), while those from DOMINE database are predicted using 13 computational methods. We used all DDIs in iPfam and 3DID database, but for DOMINE database we only used the high-confidence DDIs predicted by more than two methods. The combination of these three data sources gave 7,324 domains involved in 11,911 DDIs. Then Domains and their DDIs were incorporated into the PPI network to construct DDI network. The intra-DDI within an individual protein was also counted as elaborated in the results part.

Heterogeneity of domains and frequency of DDIs

Heterogeneity (H) measures the number of different interacting partners of each domain. It is defined as $H = \frac{N_{dn} - 1}{N_n}$, where N_{dn} is the number of different types of neighbors of a domain, and N_n is the number of total neighbors of a domain. Domains only having one neighbor are excluded from H analysis.

The frequency of DDIs were calculating as: $DDI\ frequency = \frac{N_{AB}}{\sqrt{N_A * N_B}}$, A and B are any two interacting domains in the DDI network, N_{AB} is the number of interactions between A and B, N_A and N_B are total numbers of A and B, respectively.

Network centrality measurement

The degree of a node in the network is the number of its direct neighbors(Tanaka et al., 2012). The betweenness of a node indicates its frequency of localization on the shortest paths between any other two nodes(Joy et al., 2005). It is defined as $B(x) = \sum_{u \neq v} \frac{\sigma_{uv}(x)}{\sigma_{uv}}$. σ_{uv} is the number of shortest paths between node u and v. $\sigma_{uv}(x)$ is the number of shortest paths between node u and v which pass node x . The clustering coefficient (CC) of a node indicates the density of its direct neighbors(Wuchty

et al., 2003). It is defined as $CC(x) = \frac{2e}{k_x(k_x - 1)}$. e is the number of edges between the direct neighbors of node x, k_x is the number of neighbors of x.

Evolutionary rate and co-evolution

Evolutionary rate of a specific domain was calculated by comparing the number of the domain in lower organism (yeast) vs. higher organism (human). The domain numbers of different species were retrieved from previous publication (Zmasek and Godzik, 2011). For each domain, the domain numbers of human, mouse, frog, fish, fly, worm and yeast were transformed into a vector. And the co-evolution of each interacting domain pair was calculated using Jensen Shannon Divergence (JSD) of the two vectors. $JSD(A,B) = H(\frac{A+B}{2}) - \frac{H(A)+H(B)}{2}$, where H denotes the Shannon entropy. $JSD^* = \sqrt{JSD}$ is a metric.

Mutation of domains

Mutation data was retrieve from a previous study(Kan et al., 2010). It contains 2,576 somatic mutations from 441 tumors comprising breast, lung, ovarian and prostate cancer types and subtypes. The amino acid positions of mutation were mapped to domains in the protein.

Network attack

Network attack is a simulation of network disruption by removing edges (or nodes) from the network. The characteristic path length (CPL) and the size of largest component of the network will be changed in the attack. CPL denotes the average shortest distance of any two nodes in the network. And the largest component denotes the largest connected subnetwork in the whole network. The more these two measures change, the more important the removed edges (or nodes) are for the network stability(Han et al., 2004).

In this work, we attacked the DDIs with high frequency (frequent DDIs) and low frequency (rare DDIs) from the network, respectively. And we compared the CPL of attacking them with that of attacking the same number of randomly picked DDIs. For these three types of DDIs, the attacking was done in a sequence from high- to low-betweenness DDIs.

Function combination of DDIs

The function of domains was assigned using Pfam2GO (downloaded on July 30, 2012) (Forslund and Sonnhammer, 2008). The GO term combination of interacting domains was counted. If a domain has *n* (*n>1*) GO terms, and its partner has *m* (*m>1*), then there are *n*m* combinations. DDIs between the same domains were not included. The frequency of a function combination was defined as: $Function\ combination\ frequency = \frac{Number\ of\ combination\ AB}{\sqrt{A \times B}}$. A and B are GO terms.

Gene ontology and subcellular localization analysis

All the gene oncology (GO) (Ashburner et al., 2000) enrichment analysis was done by using Ontologizer 2.0 (Bauer et al., 2008). The GO ontology (v1.2) was downloaded from GO database. The GO annotation was made based on the Pfam2GO (downloaded on July 30, 2012)(Forslund and Sonnhammer, 2008). The GO annotations of both domains in each DDI were used in GO analysis. The background dataset was all domains in the network. The protein subcellular localization information is retrieved from HPRD (072010) (Keshava Prasad et al., 2009). Domain subcellular localization is annotated according to the protein localization.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics *25*, 25-29.

Bauer, S., Grossmann, S., Vingron, M., and Robinson, P.N. (2008). Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics *24*, 1650-1651.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res *28*, 235-242.

Breitkreutz, B.J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D.H., Bahler, J., Wood, V.*, et al.* (2008). The BioGRID Interaction Database: 2008 update. Nucleic Acids Res *36*, D637-640.

Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., Castagnoli, L., and Cesareni, G. (2010). MINT, the molecular interaction database: 2009 update. Nucleic Acids Res *38*, D532-539.

Finn, R.D., Marshall, M., and Bateman, A. (2005). iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. Bioinformatics *21*, 410-412.

Forslund, K., and Sonnhammer, E.L. (2008). Predicting protein function from domain content. Bioinformatics *24*, 1681-1687.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P.*, et al.* (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88-93.
Joy, M.P., Brock, A., Ingber, D.E., and Huang, S. (2005). High-betweenness proteins in the yeast protein interaction network. Journal of biomedicine & biotechnology *2005*, 96-103.
Kan, Z., Jaiswal, B.S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H.M., Yue, P., Haverty, P.M., Bourgon, R., Zheng, J.*, et al.* (2010). Diverse somatic mutation patterns and pathway alterations in human cancers. Nature *466*, 869-873.
Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A.*, et al.* (2009). Human Protein Reference Database--2009 update. Nucleic Acids Res *37*, D767-772.
Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N.*, et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.
Stein, A., Russell, R.B., and Aloy, P. (2005). 3did: interacting protein domains of known three-dimensional structure. Nucleic Acids Res *33*, D413-417.
Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S.*, et al.* (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell *122*, 957-968.
Tanaka, G., Morino, K., and Aihara, K. (2012). Dynamical robustness in complex networks: the crucial role of low-degree nodes. Scientific reports *2*, 232.
Turner, B., Razick, S., Turinsky, A.L., Vlasblom, J., Crowdy, E.K., Cho, E., Morrison, K., Donaldson, I.M., and Wodak, S.J. (2010). iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database : the journal of biological databases and curation *2010*, baq023.
Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I.*, et al.* (2009). An empirical framework for binary interactome mapping. Nature methods *6*, 83-90.
Wang, X., Wei, X., Thijssen, B., Das, J., Lipkin, S.M., and Yu, H. (2012). Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol *30*, 159-164.
Wuchty, S., Oltvai, Z.N., and Barabasi, A.L. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. Nature genetics *35*, 176-179.
Yellaboina, S., Tasneem, A., Zaykin, D.V., Raghavachari, B., and Jothi, R. (2011). DOMINE: a comprehensive collection of known and predicted domain-domain interactions. Nucleic Acids Res *39*, D730-735.
Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X.*, et al.* (2011). Next-generation sequencing to generate interactome datasets. Nature methods *8*, 478-480.
Zmasek, C.M., and Godzik, A. (2011). Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. Genome Biol *12*, R4.