BIOINFORMATICS ORIGINAL PAPER

Vol. 26 no. 7 2010, pages 919–924
doi:10.1093/bioinformatics/btq055

Systems biology

Advance Access publication February 21, 2010

Viewing cancer genes from co-evolving gene modules

Jing Zhu[1], Hui Xiao[1], Xiaopei Shen[1], Jing Wang[1], Jinfeng Zou[1], Lin Zhang[1], Da Yang[2], Wencai Ma[2], Chen Yao[1], Xue Gong[2], Min Zhang[2], Yang Zhang[1] and Zheng Guo[1,2,*]

[1]Bioinformatics Centre and Key Laboratory for NeuroInfomation of the Education Ministry of China, School of Life Science, University of Electronic Science and Technology of China, Chengdu 610054 and [2]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Studying the evolutionary conservation of cancer genes can improve our understanding of the genetic basis of human cancers. Functionally related proteins encoded by genes tend to interact with each other in a modular fashion, which may affect both the mode and tempo of their evolution.

Results: In the human PPI network, we searched for subnetworks within each of which all proteins have evolved at similar rates since the human and mouse split. Identified at a given co-evolving level, the subnetworks with non-randomly large sizes were defined as co-evolving modules. We showed that proteins within modules tend to be conserved, evolutionarily old and enriched with housekeeping genes, while proteins outside modules tend to be less-conserved, evolutionarily younger and enriched with genes expressed in specific tissues. Viewing cancer genes from co-evolving modules showed that the overall conservation of cancer genes should be mainly attributed to the cancer proteins enriched in the conserved modules. Functional analysis further suggested that cancer proteins within and outside modules might play different roles in carcinogenesis, providing a new hint for studying the mechanism of cancer.

Contact: guoz@ems.hrbmu.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on September 1, 2009; revised on January 12, 2010; accepted on February 4, 2010

1 INTRODUCTION

Studying the evolutionary conservation of disease/cancer genes can improve our understanding of the genetic basis of human diseases. However, current results on this topic are still inconsistent. Some studies suggested that disease genes related to morbid phenotypes evolve more slowly than other genes (Blekhman *et al.*, 2008; Furney *et al.*, 2006; Kondrashov *et al.*, 2004; Thomas *et al.*, 2003; Tu *et al.*, 2006). Other studies (Huang *et al.*, 2004; Smith and Eyre-Walker, 2003) indicated that disease genes may evolve faster than other genes and thus are under weaker negative selection. Notably, most of these studies treated disease/cancer genes and the rest genes as if they were completely isolated and uncorrelated. However, it is known that proteins encoded by genes tend to perform functions by interacting with each other in a modular fashion (Barabasi and Oltvai, 2004), which may affect both the mode and tempo of molecular evolution by reducing pleiotropy and enabling protein modules to be co-opted for new functions (Brown and Jurisica, 2007; Fraser, 2005, 2006; Gerhart, 1997; Hakes *et al.*, 2007; Hartwell *et al.*, 1999; Schlosser, 2002; Waxman and Peck, 1998). Thus, studying the evolutionary patterns of cancer genes in the context of protein (gene) modules may provide us new clues to understand the genetic basis of human cancer.

In this study, in the human PPI network, we firstly searched for PPI subnetworks within each of which all proteins have evolved at similar rates since the human and mouse split. At a given co-evolving level, the identified PPI subnetworks with significantly large sizes were selected as co-evolving modules, which could be highly cohesive fragments of some co-evolving functional units. Comparing proteins within and outside modules, we found that genes encoding proteins within modules tend to be conserved, evolutionarily old and enriched with housekeeping genes (Dezso *et al.*, 2008), while genes encoding proteins outside modules are less-conserved, evolutionarily younger and enriched with genes expressed in specific tissues.

Then, for proteins encoded by cancer genes collected in Cancer Gene Census database (Futreal *et al.*, 2004), we showed that the cancer proteins within modules tend to be evolutionarily old and conserved, while the cancer proteins outside modules tend to be evolutionarily younger and less conserved. Thus, the overall conservation of cancer proteins in comparison with other proteins (Furney *et al.*, 2006; Thomas *et al.*, 2003) should be mainly attributed to the cancer proteins enriched in co-evolving modules. These results suggested that cancer proteins with different evolutionary patterns may play different functional roles in carcinogenesis. In fact, we found that cancer proteins together with their co-evolving neighbors within modules tend to participate in some essential biological processes such as 'nuclear mRNA splicing, via spliceosome' and 'post-translational protein modification', while cancer proteins outside modules tend to be enriched in functions such as 'DNA damage response' and 'DNA repair'.

2 METHODS

2.1 The evolutionary rate and evolutionary age data

The synonymous substitution rate (K_s) and non-synomymous substitution rate (K_a) data for human genes and their mouse (gallus) homologs were downloaded from NCBI HomoloGene (ftp://ftp.ncbi.nlm.nih.gov/pub/HomoloGene/build58) (Wheeler *et al.*, 2007). The K_a and K_s data for pairs of homolog genes in a HomoloGene group are denoted by the symbols <HG-Stats_ka> and <HG-Stats_ks> in the homologene.xml.gz file (a XML dump

*To whom correspondence should be addressed.

of the HomoloGene build). To prevent possible contamination by paralogous genes, we only considered one-to-one mapped orthologous pairs as Tu *et al.* (2006) did.

The orthologs of human genes in 39 fully sequenced eukaryotes were obtained from the phylogenetic data by (Huerta-Cepas *et al.*, 2007). Here, we termed the proteins with fungi origin 'old', proteins with ortholgs in metazoan (*Anopheles gambiae*, *Drosophila melanogaster*, *Apis mellifera*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*) 'middle', and proteins specific to species emerged later than metazoan 'young'.

2.2 The GO annotation and PPI data

The GO annotation data was downloaded on August 16, 2008. The human PPI data was downloaded from MINT (Zanzoni *et al.*, 2002), BIND (Bader *et al.*, 2001), IntAct (Hermjakob *et al.*, 2004), HPRD (Peri *et al.*, 2004), MIPS (Mewes *et al.*, 2008), DIP (Salwinski *et al.*, 2004), KEGG (PPrel for protein–protein interactions and ECrel for enzymes involved in neighboring steps) (Kanehisa *et al.*, 2006) and Reactome protein pairs involved in a complex, reaction and neighboring reaction (Joshi-Tope *et al.*, 2005). We pooled together the eight PPI datasets as did in (Lage *et al.*, 2007) and compiled an integrated PPI network of 84 872 distinct interactions involving 10 863 connected human proteins with evolutionary rate data (K_a, K_a/K_s). Here, to reduce the false positive rate of the PPI data, we used the literature-based PPI data from multiple sources. Though the literature-based human PPI data might be biased to well-studied proteins (Oti *et al.*, 2006), some researchers argued that this bias is negligible (Rambaldi *et al.*, 2008). We note that even if this bias exists, our following results would be affected little because we focused on comparing the cancer proteins within and outside modules.

2.3 Cancer genes and housekeeping genes

From Cancer Gene Census database (Futreal *et al.*, 2004), we extracted 313 cancer genes with both K_a and PPI data. We also analyzed 340 candidate cancer genes with both K_a and PPI data, which were found in four somatic mutation screens of cancer genomes (Greenman *et al.*, 2007; Jones *et al.*, 2008; Parsons *et al.*, 2008; Wood *et al.*, 2007). Focusing on genes with both K_a and PPI data, 1856 housekeeping genes expressed in all of 31 human tissues and 598 genes uniquely expressed in a single tissue were extracted from (Dezso *et al.*, 2008).

2.4 Measures of the difference of proteins in evolutionary rates

The evolutionary rate of a protein is often measured by K_a defined as the number of non-synonymous substitutions per non-synonymous site of the protein since the split of two species. It is also often measured by K_a/K_s where K_s is the number of synonymous substitutions per synonymous site, which calculates the accepted non-synonymous substitutions normalized by chance if selection does not act on silent sites (Hurst, 2002; Wall *et al.*, 2005). However, K_s is also subject to selection in many species (Akashi, 2001). Thus, we used both K_a and K_a/K_s to measure the evolutionary rates of human proteins. Here, we only showed the results based on K_a. The results based on K_a/K_s were consistent.

The absolute difference of two proteins i and j in evolutionary rate since the human–mouse split was calculated as $\Delta K_{aij} = |K_{ai} - K_{aj}|$. If they have evolved at similar rates over this time range, then ΔK_{aij} should be small (Marino-Ramirez *et al.*, 2006). As a distance metrics, ΔK_{aij} has three main properties: (i) symmetry, $\Delta K_{aij} = \Delta K_{aji}$; (ii) non-negative, $\Delta K_{aij} \geq 0$; and (iii) delta non-negative, $\Delta K_{aij} \leq (\Delta K_{aik} + \Delta K_{ajk})$.

2.5 Finding co-evolving functional modules

In the human PPI network, we searched for locally maximal PPI subnetworks within each of which all proteins have evolved at similar rates since the human and mouse split. In each subnetwork, the absolute difference of the evolutionary rates for every two proteins was required within a given threshold, reflecting the co-evolving level of these proteins in the subnetwork. Here, the threshold was taken as the top 5, 2.5 or 1 percentile of the distribution of ΔK_{aij} for all-against-all proteins pairs.

Using the method proposed in Chuang *et al.* (2007), we performed a greedy search in the human PPI network to find co-evolving PPI subnetworks with locally maximal sizes. First, we took each protein as a seed. Then, from its interacting neighbors, we selected one protein whose evolutionary rate is similar to that of the seed. Iteratively, from proteins interacting with at least one of the pre-selected proteins, we selected one protein whose evolutionary rate was similar to all members of the pre-selected proteins. At each iteration step, when there were several proteins evolved at similar rates with all the pre-selected proteins, we randomly chose one. Such a random choice may lead to slightly different subnetworks (see details in Supplementary Data). The search stops when no proteins can be added.

The subnetworks selected above were considered as candidate subnetworks, from which we further selected significant subnetworks with non-randomly large sizes according to two random experiments (Chuang *et al.*, 2007). First, we produced 100 random networks by permuting the evolutionary rates of the proteins in the original PPI network and performed the same search procedure in each random network. The significance of the size of a candidate subnetwork was calculated as the percentage of the subnetworks, extracted from the 100 random networks, whose sizes were larger than the candidate subnetwork. Then, in each of the 100 random networks, we extracted subnetworks initialized from the same seed protein for producing the candidate subnetwork and estimated the percentage of the extracted subnetworks which were larger than the candidate subnetwork. Finally, from the candidate subnetworks, we selected the ones with significantly larger sizes according to both tests ($P \leq 0.05$) as co-evolving modules or modules for short.

The non-randomly large sizes of the modules suggested that the emergence of the modules could be introduced by the underlying evolutionary constraint on interacting proteins. Notably, when allowing a larger threshold for the absolute difference between the evolutionary rates of every two proteins in a subnetwork, we could identify larger subnetworks. The smaller the threshold is controlled, the more cohesive are the identified modules in co-evolution. Identified at a co-evolving level, the modules could be considered as highly cohesive fragments of some co-evolving functional units. At a given co-evolving level, we defined all proteins in the identified modules as proteins within modules and compared them with proteins outside modules.

Identified by the greedy search algorithm, some modules might share proteins and their boundaries could not be clearly defined. This is a common problem for algorithms for finding PPI subnetworks based on optimizing some given scores (Chuang *et al.*, 2007; Kelley and Ideker, 2005; Ulitsky and Shamir, 2007). Here, we did not try to delete or merge modules with overlapping proteins according to some empirical criterion as did in some researches (Kelley and Ideker, 2005; Ulitsky and Shamir, 2007). Instead, we simply treated all modules as cohesive fragments of some co-evolving functional units.

2.6 Difference between proteins within and outside modules in the evolutionary rate

To evaluate the significance of the difference between the median K_a of the proteins within and outside modules, we calculated the corresponding difference in each of 100 random networks produced by randomly permuting the K_a values of the proteins in the original PPI network. Then, the significance (P-value) of the real difference was calculated as the percentage of the random differences which were larger than the real difference. When evaluating the significance of the difference between the median K_a of the cancer proteins within and outside modules, we produced 100 random networks by separately permuting K_a for cancer proteins and other proteins in the original PPI network. By such a way, the K_a distribution of cancer proteins in each random network was the same as that in the original PPI network.

We note that when applying the Mann–Whitney U-test to examine the difference of proteins within and outside modules in K_a, the same significant results were found. However, the Mann–Whitney U-test might be biased to some extend since the evolutionary rates of interacting proteins are correlated.

2.7 Functional enrichment analysis

We identified biological process categories of GO enriched with proteins within (or outside) modules by Elim algorithm (Alexa *et al.*, 2006) which can remove some generic categories whose enrichment could be explained by their significant descendant categories. In a category of GO, the probability (P) of observing at least a given number of proteins within (or outside) modules by random chance was calculated by the hypergeometric test. The $P-$values were adjusted by the Bonferroni correction.

3 RESULTS

3.1 Co-evolving modules and proteins' conservation

We set the threshold for controlling the absolute difference of the evolutionary rates for each protein pair in a subnetwork as the top 5, 2.5 or 1 percentile cutoff of the distribution of ΔK_{aij} for all-against-all proteins pairs. When setting the threshold as the top 5 percentile cutoff, we found 382 co-evolving modules with sizes significantly larger than expected by random chance ($P \leq 0.05$) by the algorithm described in 'Methods' section. When setting the threshold as the top 2.5 or 1 percentile cutoff, though the identified modules became smaller, all our following conclusions were consistently obtained (Supplementary Data and Supplementary Figs S1 and S2).

As shown in Figure 1A, most proteins within modules are conserved with K_a ranging from 0 to 0.0261 (median 0.0103), while most proteins outside modules are not conserved with K_a ranging from 0 to 0.7038 (median 0.0644). However, due to some statistical and biological factors (see more in 'Discussion' section), a small fraction of the proteins outside modules are also conserved. The difference between the median K_a of proteins within and outside modules was not observed in 100 random networks with permuted K_a ($P < 0.01$). Furthermore, according to the human phylome (Huerta-Cepas *et al.*, 2007), 81.13% of the genes encoding proteins within modules have fungi origin (termed old, see 'Methods' section) or orthologs in metazoan (termed middle, see 'Methods' section) (Fig. 1B). Thus, the proteins within modules tend to be involved in essential functions for both single- and multi-cellular organisms. In comparison, only 55.93% of the proteins outside modules are of old or middle age ($P < 2.20e-16$, Fisher's exact test). Additionally, based on the gene expression profiles in 31 human tissues (Dezso *et al.*, 2008), we found that genes encoding proteins within modules are enriched with housekeeping genes ($P = 1.59e-11$, hypergeometric test), while genes encoding proteins outside modules are enriched with genes expressed in specific tissues ($P = 2.27e-06$, hypergeometric test). All together, these results supported the assumption that evolutionary innovations tend to occur by altering proteins outside modules (Fraser, 2005; Hartwell *et al.*, 1999).

Similar results were observed when focusing cancer proteins within and outside modules. The median K_a (0.01) of the cancer proteins within modules is smaller than the median K_a (0.0528) of the cancer proteins outside modules (Fig. 2A). By the randomized experiment as described in 'Methods' section, we found that the difference between the median K_a of cancer proteins within and outside modules could not be observed in any of the 100 random networks ($P < 0.01$). This result suggested that the previously reported overall conservation of cancer proteins in comparison with other proteins (Furney *et al.*, 2006; Thomas *et al.*, 2003) should be mainly attributed to the enrichment of cancer proteins in co-evolving modules ($P = 2.60e-08$, hypergeometric test). Additionally, 62.82% of the cancer proteins within modules are of old or middle age, while less (50.28%) cancer proteins outside modules are of old or middle age ($P = 0.0631$, Fisher's exact test) (Fig. 2B).

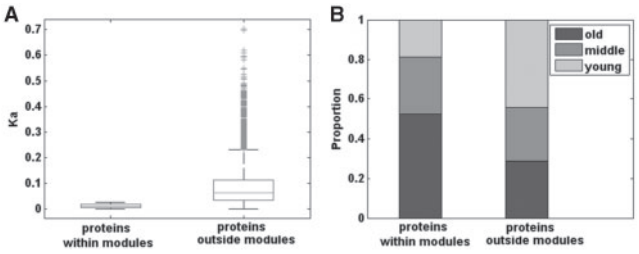

Fig. 1. Difference between proteins within and outside modules in the evolutionary rate and evolutionary age. (**A**) The box-and-whisker plots of K_a for proteins within and outside modules. The center mark of each of the boxes is the median of K_a and the edges of the box are the 25th and 75th percentiles. The whiskers extend to the furthest point within 1.5 times the inter-quartile range (75th–25th percentile). Beyond the whiskers, all outliers are shown in crosses. (**B**) Evolutionary ages of the proteins within and outside modules: proteins with fungi origin (old age, black); proteins with orthologs in metazoan (middle age, gray); proteins younger than metazoan (young age, white).

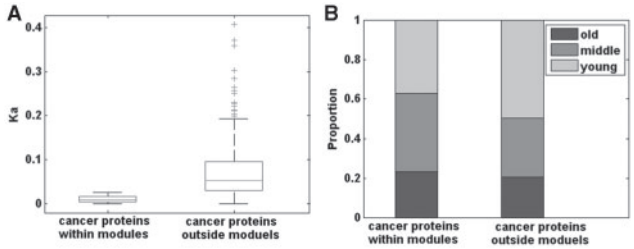

Fig. 2. Difference between cancer proteins within and outside modules in the evolutionary rate and evolutionary age. (**A**) The box-and-whisker plots of K_a for cancer proteins within and outside modules. The legend is the same as in Figure 1A. (**B**) Evolutionary ages of the cancer proteins within and outside modules. The evolutionary ages are colored as in Figure 1B.

3.2 Functional roles of different cancer proteins

Different conservation patterns of cancer proteins within and outside modules suggest that they may play distinct functional roles in carcinogenesis. By the hypergeometric test with 5% FDR control, we found that 139 of the 382 co-evolving module are enriched with cancer proteins. Additionally, we found that proteins encoded by the candidate cancer genes found from somatic mutation screens of cancer genomes (Greenman *et al.*, 2007; Jones *et al.*, 2008; Parsons *et al.*, 2008; Wood *et al.*, 2007) are also significantly enriched in these 139 co-evolving modules ($P = 0.0301$, hypergeometric test), suggesting that disruptions of these co-evolving modules might play essential roles in carcinogenesis.

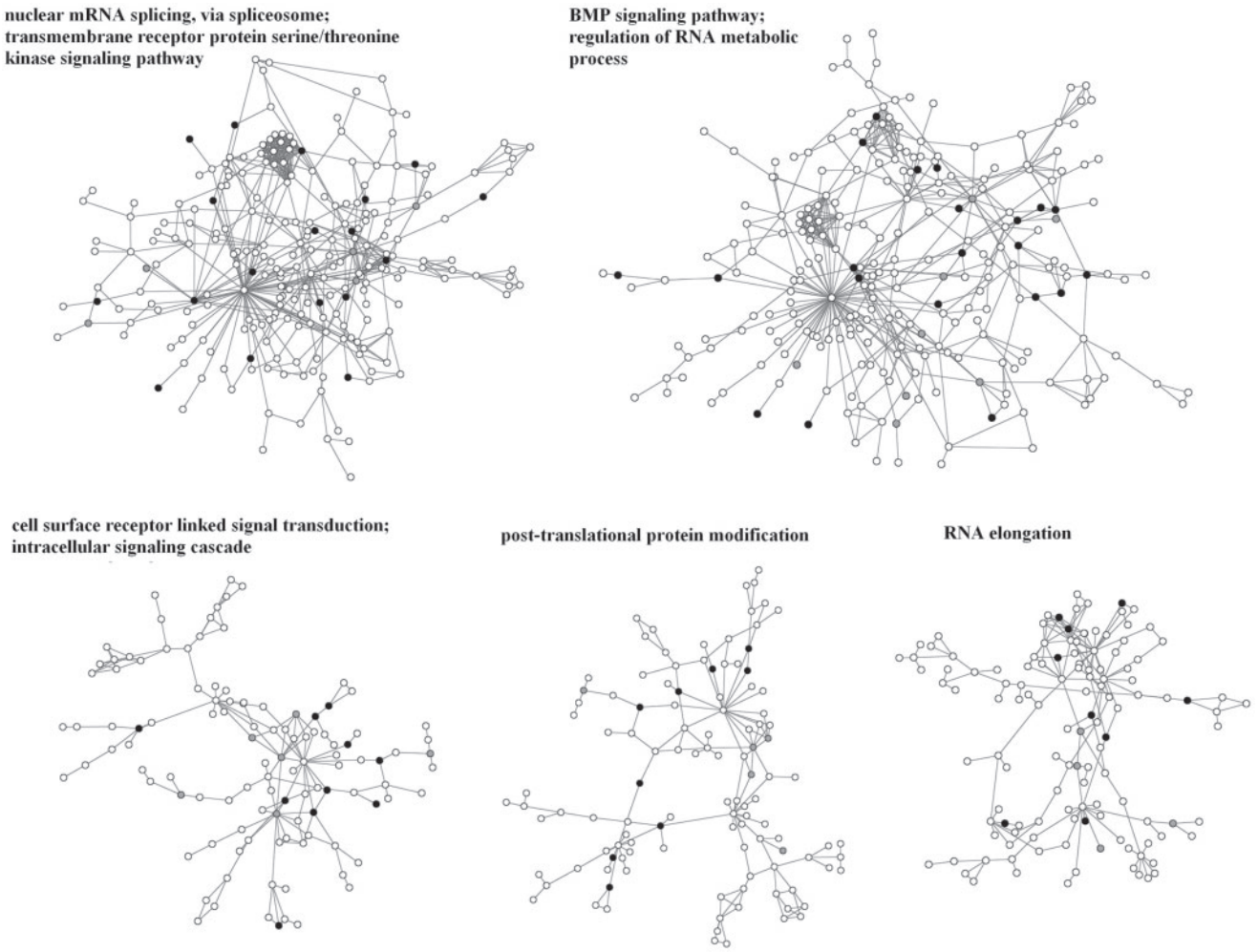

Fig. 3. Five modules enriched with cancer proteins. Black nodes represent cancer proteins, grey nodes represent candidate cancer proteins, white nodes represent other proteins, and grey lines represent PPI links. The biological processes enriched with proteins from a module are listed alongside the module.

Then, we performed functional enrichment analysis for each of the 139 modules, by the method described in 'Methods' section. The results showed that genes within these co-evolving modules are enriched in essential biological processes 'nuclear mRNA splicing, via spliceosome', 'transmembrane receptor protein serine/threonine kinase signaling pathway', 'RNA elongation', 'post-translational protein modification', 'regulation of RNA metabolic process' and 'BMP signaling pathway' (Fig. 3) (Supplementary Table S1). The cancer proteins outside modules are enriched in GO biological processes 'DNA damage response', 'signal transduction by p53 class mediator resulting in induction of apoptosis', 'DNA repair', 'mismatch repair', 'regulation of transcription, DNA-dependent', 'transcription from RNA polymerase II promoter' and 'peptidyl-tyrosine phosphorylation'. The results suggested that cancer proteins within modules tend to participate in essential biological processes and be conserved, while cancer proteins outside modules tend to participate in functions for keeping the stability of the genome and have a higher degree of evolutionary plasticity.

4 DISCUSSION AND CONCLUSION

Based on the assumption that genes/proteins tend to express and perform functions in a modular fashion, some methods (e.g. EM algorithm, simulated annealing and greedy search) have been proposed for finding different types of gene/protein modules such as modules of co-regulated genes (Segal *et al.*, 2003) and PPI subnetworks related to a specific condition(s) (Chuang *et al.*, 2007; Guo *et al.*, 2007; Ideker *et al.*, 2002; Kelley and Ideker, 2005; Ulitsky and Shamir, 2007). Here, similarly to algorithms searching for PPI subnetworks related to a specific condition (Chuang *et al.*, 2007), we performed a greedy search in the human PPI network to identify co-evolving PPI subnetworks at a given co-evolving level. The identified PPI subnetworks could be highly cohesive parts or fragments of some co-evolving functional units. Our results suggested that proteins within and outside modules have different evolutionary patterns and supported the assumption that evolutionary innovations tend to occur by altering proteins outside modules (Fraser, 2005; Hartwell *et al.*, 1999). Viewing cancer genes

from co-evolving modules suggested that the overall conservation of cancer genes (Furney *et al.*, 2006; Thomas *et al.*, 2003) should be mainly attributed to the enrichment of cancer proteins in the conserved modules. In addition, we found that different from cancer genes within modules participating in some essential biological processes, cancer genes outside modules tend to play specific roles in functions related to DNA damage response and DNA repair. All together, these results suggested that the dysregulation of cancer proteins with different evolutionary and functional patterns might play different roles in carcinogenisis.

Genes might have differentiated functionalities in different species, leading to changed co-evolving modules during different evolutionary periods. Here, we analyzed a case about the functional overlapping between modules co-evolving since the human–mouse split and modules co-evolving since the human–gallus split which took place much earlier than the time of the human–mouse split. Generally, the modules of genes co-evolving since the human–gallus split are smaller than the modules of genes co-evolving since the human–mouse split. The largest module for human–gallus contains 134 genes enriched in 'nuclear mRNA splicing, via spliceosome' and 'translational elongation' (by the Elim algorithm, Bonferroni correction $P < 0.05$). As shown in Supplementary Table S2, this module shares 68 genes with the largest module for human–mouse which contains 396 genes enriched in the above two functions as well as in another four functions. However, some other genes in the largest module for human-gallus are not included in the largest modules for human–mouse. These results indicated that some genes cohesively co-evolving during the time since both the human–mouse split and the human–gallus split might perform some basic functions and thus be subjected to higher evolutionary constraints, while some other genes might have differentiated functionalities and thus different co-evolving patterns since the human–mouse split. However, this result should be interpreted with great caution because the partial overlapping between co-evolving modules might be artificially introduced by the algorithm limitation that a co-evolving functional unit could be detected as several fragmental modules.

As shown in the 'Results' section, some proteins outside modules are conserved though most are not. The conservation of some proteins outside modules could be explained by some statistical and biological factors. First, with a threshold for controlling the absolute difference between the evolutionary rates of protein pairs in the PPI subnetworks, the modules identified at a co-evolving level may miss some proteins (or some parts) of co-evolving functional units. Second, because current PPI data may cover only a small fraction of all the human protein–protein interactions (Hart *et al.*, 2006), some proteins co-evolving with proteins in modules might be missed though we pooled eight literature-based PPI datasets together (Lage *et al.*, 2007). Also, the incomplete PPI data may lead to the presence of many small modules whose sizes were not significantly larger than expected by random chance but they could be biologically meaningful. This problem needs to be further addressed by integrating high-throughput PPI data and by assessing the significance of the modules on other aspects. Third, one hypothesis about how a PPI network and its constituent modules evolve indicates that proteins interacting with multiple modules might evolve more slowly owing to the high pleiotropic constraints on them (Fisher, 1930; Greenberg *et al.*, 2008; Hahn and Kern, 2005; Waxman and Peck, 1998). According to this hypothesis, some conserved proteins outside modules might regulate or participate in multiple functions and be subjected to high evolutionary constraints. In our previous work (Ma *et al.*, 2009), we found that the disruption of genes densely interacting with multiple functions associated with cancers may play pivot roles in carcinogenesis. In general, because genes play biological functions in a coordinated fashion in biological systems, studying the coordinated operations of genes within and outside modules can help us understand how these genes (and modules) play functional roles in inducing diseases. In our future work, we will further elucidate the coordinated roles of cancer proteins within and outside modules in inducing cancer by exploiting molecular data such as somatic mutations and epigenetic changes in cancer genomes (Chan *et al.*, 2008; Schuebel *et al.*, 2007).

ACKNOWLEDGEMENTS

Funding: National Natural Science Foundation of China (grant nos 30571034, 30970668, 30670539).

Conflict of Interest: none declared.

REFERENCES

Akashi,H. (2001) Gene expression and molecular evolution. *Curr. Opin. Genet. Dev.*, **11**, 660–666.

Alexa,A. *et al.* (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600–1607.

Bader,G.D. *et al.* (2001) BIND–the biomolecular interaction network database. *Nucleic Acids Res.*, **29**, 242–245.

Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.

Blekhman,R. *et al.* (2008) Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.*, **18**, 883–889.

Brown,K.R. and Jurisica,I. (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.*, **8**, R95.

Chan,T.A. *et al.* (2008) Convergence of mutation and epigenetic alterations identifies common genes in cancer that predict for poor prognosis. *PLoS Med.*, **5**, e114.

Chuang,H.Y. *et al.* (2007) Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.*, **3**, 140.

Dezso,Z. *et al.* (2008) A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol.*, **6**, 49.

Fisher,R.A. (1930) *The Genetical Theory of Natural Selection*. Dover, New York.

Fraser,H.B. (2005) Modularity and evolutionary constraint on proteins. *Nat. Genet.*, **37**, 351–352.

Fraser,H.B. (2006) Coevolution, modularity and human disease. *Curr. Opin. Genet. Dev.*, **16**, 637–644.

Furney,S.J. *et al.* (2006) Structural and functional properties of genes involved in human cancer. *BMC Genomics*, **7**, 3.

Futreal,P.A. *et al.* (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.

Gerhart,J.K.M. (1997) *Cells, Embryos, and Evolution*. Blackwell Science, Malden, Massachusetts.

Greenberg,A.J. *et al.* (2008) Evolutionary constraint and adaptation in the metabolic network of Drosophila. *Mol. Biol. Evol.*, **25**, 2537–2546.

Greenman,C. *et al.* (2007) Patterns of somatic mutation in human cancer genomes. *Nature*, **446**, 153–158.

Guo,Z. *et al.* (2007) Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network. *Bioinformatics*, **23**, 2121–2128.

Hahn,M.W. and Kern,A.D. (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.*, **22**, 803–806.

Hakes,L. *et al.* (2007) Specificity in protein interactions and its relationship with sequence diversity and coevolution. *Proc. Natl Acad. Sci. USA*, **104**, 7999–8004.

Hart,G.T. *et al.* (2006) How complete are current yeast and human protein-interaction networks? *Genome Biol.*, **7**, 120.

Hartwell,L.H. *et al.* (1999) From molecular to modular cell biology. *Nature*, **402**, C47–C52.

Hermjakob,H. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res.*, **32**, D452–D455.

Huang,H. *et al.* (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol.*, **5**, R47.
Huerta-Cepas,J. *et al.* (2007) The human phylome. *Genome Biol.*, **8**, R109.
Hurst,L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.*, **18**, 486.
Ideker,T. *et al.* (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18**(Suppl 1), S233–S240.
Jones,S. *et al.* (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, **321**, 1801–1806.
Joshi-Tope,G. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
Kanehisa,M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
Kelley,R. and Ideker,T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.*, **23**, 561–566.
Kondrashov,F.A. *et al.* (2004) Bioinformatical assay of human gene morbidity. *Nucleic Acids Res.*, **32**, 1731–1737.
Lage,K. *et al.* (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.*, **25**, 309–316.
Ma,W.C. *et al.* (2009) Finding disease-specific coordinated functions by multi-function genes: insight into the coordination mechanisms in diseases. *Genomics*, **94**, 94–100.
Marino-Ramirez,L. *et al.* (2006) Co-evolutionary rates of functionally related yeast genes. *Evol. Bioinform. Online*, **2**, 295–300.
Mewes,H.W. *et al.* (2008) MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res.*, **36**, D196–D201.
Oti,M. *et al.* (2006) Predicting disease genes using protein-protein interactions. *J. Med. Genet.*, **43**, 691–698.
Parsons,D.W. *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science*, **321**, 1807–1812.
Peri,S. *et al.* (2004) Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.*, **32**, D497–D501.
Rambaldi,D. *et al.* (2008) Low duplicability and network fragility of cancer genes. *Trends Genet.*, **24**, 427–430.
Salwinski,L. *et al.* (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
Schlosser,G. (2002) Modularity and the units of evolution. *Theory Biosci.*, **121**, 1–80.
Schuebel,K.E. *et al.* (2007) Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS Genet.*, **3**, 1709–1723.
Segal,E. *et al.* (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.
Smith,N.G. and Eyre-Walker,A. (2003) Human disease genes: patterns and predictions. *Gene*, **318**, 169–175.
Thomas,M.A. *et al.* (2003) Evolutionary dynamics of oncogenes and tumor suppressor genes: higher intensities of purifying selection than other genes. *Mol. Biol. Evol.*, **20**, 964–968.
Tu,Z. *et al.* (2006) Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics*, **7**, 31.
Ulitsky,I. and Shamir,R. (2007) Pathway redundancy and protein essentiality revealed in the Saccharomyces cerevisiae interaction networks. *Mol. Syst. Biol.*, **3**, 104.
Wall,D.P. *et al.* (2005) Functional genomic analysis of the rates of protein evolution. *Proc. Natl Acad. Sci. USA*, **102**, 5483–5488.
Waxman,D. and Peck,J.R. (1998) Pleiotropy and the preservation of perfection. *Science*, **279**, 1210–1213.
Wheeler,D.L. *et al.* (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **35**, D5–D12.
Wood,L.D. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108–1113.
Zanzoni,A. *et al.* (2002) MINT: a Molecular INTeraction database. *FEBS Lett.*, **513**, 135–140.