Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Theoretical Biology 287 (2011) 100–108

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Disease embryo development network reveals the relationship between disease genes and embryo development genes

Binsheng Gong[a,1], Tao Liu[a,1], Xiaoyu Zhang[b], Xi Chen[a], Jiang Li[a], Hongchao Lv[a], Yi Zou[c], Xia Li[a,c,*], Shaoqi Rao[a,**]

[a] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China
[b] Second Affiliated Hospital, Harbin Medical University, Harbin 150081, China
[c] Biomedical Engineering Institute of CUMS, Beijing 100054, China

ARTICLE INFO

Article history:
Received 8 March 2011
Received in revised form
15 June 2011
Accepted 22 July 2011
Available online 3 August 2011

Keywords:
Disease process
Embryo development
Human PPI network
Network properties

ABSTRACT

A basic problem for contemporary biology and medicine is exploring the correlation between human disease and underlying cellular mechanisms. For a long time, several efforts were made to reveal the similarity between embryo development and disease process, but few from the system level. In this article, we used the human protein–protein interactions (PPIs), disease genes with their classifications and embryo development genes and reconstructed a human disease-embryo development network to investigate the relationship between disease genes and embryo development genes. We found that disease genes and embryo development genes are prone to connect with each other. Furthermore, diseases can be categorized into three groups according to the closeness with embryo development in gene overlapping, interacting pattern in PPI network and co-regulated by microRNAs or transcription factors. Embryo development high-related disease genes show their closeness with embryo development at least in three biological levels. But it is not for embryo development medium-related disease genes and embryo development low-related disease genes. We also found that embryo development high-related disease genes are more central than other disease genes in the human PPI network. In addition, the results show that embryo development high-related disease genes tend to be essential genes compared with other diseases' genes. This network-based approach could provide evidence for the intricate correlation between disease process and embryo development, and help to uncover potential mechanisms of human complex diseases.

1. Introduction

One of the important challenges for contemporary biology and medicine is to understand the relationship between human diseases and underlying cellular mechanisms (Argmann et al., 2005; Giallourakis et al., 2005; Kann, 2007; Lage et al., 2007; Lamb et al., 2006; Loscalzo et al., 2007; Oti et al., 2008; Schadt et al., 2005). During the past decade, many efforts have been dedicated to identifying disease-related genes, proteins and metabolites, which directly or indirectly interact through genetic or physical behavior (Albert, 2005; Almaas, 2007; Alon, 2007; Barabasi and Oltvai, 2004; Basso et al., 2005; Gerstein et al., 2002; Yildirim et al., 2007). In recent years, researchers concentrate on the embryo development in disease process (Groenendijk et al., 2008; Relaix, 2006; Sharma et al., 2006; Tutarel et al., 2005). Some evidences suggest that cellular mechanisms during the embryogenesis are similar with what occur in some disease process such as cancer (Davidson, 2009; Giuffrida et al., 2009; Ma et al., 2010; Navarro and Monzo, 2010; Ruiz-Vela et al., 2009). For instance, Navarro et al. reviewed some complex relationships between embryo development and cancer. However, it is increasingly comprehended that those gene-by-gene or protein-by-protein approaches, even genome-wide studies, although tremendously successful, are far from achievement, because the complicated cellular mechanisms are executed through an intricate network comprising regulatory and protein interaction (Albert, 2005; Barabasi and Oltvai, 2004; Cusick et al., 2005; Zhu et al., 2007). Exploring such systematic relationship among cellular networks and disease patterns could possibly open a brand-new boulevard for understanding the interactome, and may help discover underlying mechanisms of diseases (Braun et al., 2008; Ergun et al., 2007; Goh et al., 2007; Lee et al., 2008; Loscalzo et al., 2007).

More investigations have focused on network-based approaches to study human diseases with the increasing high-throughput data (Feldman et al., 2008; Friedman and Perrimon, 2007; Goh et al.,

* Corresponding author at: College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Tel./fax: +86 045186615922.
** Corresponding author.
E-mail addresses: lixia@hrbmu.edu.cn (X. Li), shaoqirao@yahoo.com (S. Rao).
[1] Joint first authors.

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.07.018

2007; Ideker and Sharan, 2008; Lee et al., 2008; Li and Agarwal, 2009; Tu et al., 2006; Wang et al., 2009). For instance, Goh et al. (2007) constructed the disease phenome network (human disease network, HDN) and disease genome network (disease gene network, DGN), using disease-gene association pairs extracted from the Online Mendelian Inheritance in Man (OMIM) database. Thus, Lee et al. (2008) created a metabolic disease network (MDN), and explained that metabolic diseases trend to be co-morbid in the population if the enzymes and their related diseases are linked through metabolic pathways. Recently, Wang et al. (2009) demonstrated a close association between aging and diseases.

Using graphical approaches to study biological problems can provide an intuitive picture or useful insights for helping analyzing complicated mechanisms in these systems, as demonstrated by many previous studies on a series of important biological topics, such as enzyme-catalyzed reactions (Chou, 1981, 1989; Chou and Forsen, 1980; Zhou and Deng, 1984), protein folding kinetics (Chou, 1990), inhibition of HIV-1 reverse transcriptase (Althaus et al., 1993a, 1993b, 1993c), inhibition kinetics of processive nucleic acid polymerases and nucleases (Chou et al., 1994), drug metabolism systems (Chou, 2010), analysis of DNA sequence (Xie and Mo, 2011; Yu et al., 2009) and protein sequence evolution (Wu et al., 2010). Meanwhile, the 'wenxiang diagram' (Chou et al., 1997) has been used to investigate protein–protein interactions(Chou and Cai, 2006; Smith and Nilar, 2010), and the graphical representation utilized to identify the hub proteins from complicated network systems (Gonzalez-Diaz et al., 2009; Hu et al., 2011; Shen et al., 2010). Recently, the 'cellular automaton image' (Wolfram, 1984) has also been applied to study hepatitis B viral infections (Xiao et al., 2006), HBV virus gene missense mutation (Xiao et al., 2005b), and visual analysis of SARS-CoV (Wang et al., 2005), as well as representing complicated biological sequences (Xiao et al., 2005a), and helping to identify various protein attributes (Xiao et al., 2008, 2009).

In this paper we focus on embryo development, which is one of the important processes of human being and previous studies have investigated relationship between embryo development and disease such as cancer (Doi et al., 2009; Lalli and Alonso, 2010; Postovit et al., 2008; Takebe and Ivy, 2010; Ullmann, 2010). Research on embryo development is beneficial to comprehend the nature of disease by integrating disease and embryo development information at a network level. We note that their relationships have been pointed out for a long time, but seldom been researched from the systems level. Here we emphasize the intricate relationships between embryo development and disease since the embryo development process is gradual development process from zygote to adult cell and this process is similarity with some of disease such as cancer, in which the process is reversed from adult cell to cancer stem cell (Takebe and Ivy, 2010; Ullmann, 2010). We tried to use the embryo development genes and disease genes to highlight the intricate correlations between embryo development and disease process systematically. We analyzed the association between embryo development genes and disease genes from PPI network and reconstructed a disease-embryo development network (DEN) and analyzed its topological properties. Then according to the relationship between embryo development genes and disease genes in different biological levels, we divided the diseases into three groups: embryo development high-related disease (EHD), embryo development medium-related disease (EMD) and embryo development low-related disease (ELD). In addition, we compared the differences between the three groups of disease genes from network centrality, functional enrichment and the proportion of essential genes. And we also compared the closeness centrality of embryo development genes in the network of three groups of diseases.

2. Materials and methods

2.1. Datasets

2.1.1. The human embryo development genes

A list of human embryo development genes was extracted from two sources. Firstly, we obtained the gene set involving the word "embryo" and "development" in the Entrez Gene description field (downloaded from the Entrez Gene website ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/mammalia/Homo_sapiens.gene_info.gz on June 30, 2010). Secondly, we obtained the gene set annotated the Gene Ontology functional terms which include the descriptive word "embryo" and "development" (downloaded from http://www.geneontology.org on June 30, 2010). We finally got 643 genes in which 516 genes are contained in the human PPI network.

2.1.2. Disease genes and classification of diseases

The disease genes and their classification were obtained from previous study (Goh et al., 2007), which were manually classified into 22 disease classes based on the physiological system affected. Totally, there are 1777 disease genes in which 1345 are involved in the human PPI network. Particularly, some diseases, which have multiple clinical features, were marked "multiple", and some other diseases were labeled "unclassified" if they could not be assigned to a clear disease class. "Multiple" class disease represents that the disease has distinct multiple clinical features or has no evidence to show which systems the disease affect. And "unclassified" class disease represent that the disease cannot be classified clearly. Then we constructed the DEN with all 22 diseases class, but did not consider the "multiple" class and "unclassified" class in the following discussion.

2.1.3. Protein–protein network and essential genes

Human protein–protein interaction (PPI) network is obtained from HPRD Release 9 (http://www.hprd.org, downloaded on June 30, 2010). Until the download date, the network involved 9611 proteins and 39,118 interactions. Essential genes were downloaded from the Mouse Genome Informatics (http://www.informatics.jax.org, downloaded on August 2, 2010). We got 2661 essential genes in which 2191 are in HPRD.

2.2. Measures

2.2.1. Closeness centrality of nodes in network

We used closeness centrality (CC) to measure the centrality of nodes in a given network. CC is defined by

$$c_v = \frac{N-1}{\sum_{u \in v} d(v,u)} \tag{1}$$

where v is the node in a given network, N is the total number of nodes in the network and $d(v,u)$ represents the shortest path between node v and u.

2.2.2. p-Value by overlapping

We used this formula to calculate the significant p-value of gene sharing, interaction partner of PPI network, co-regulated by microRNAs and TFs. p-values were calculated by the following formula:

$$P(X=k|N,m,n) = 1 - \sum_{i=0}^{k} \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \tag{2}$$

Consider that a set containing M elements has two subsets S1 and S2 with K and N elements, respectively. We calculate the probability that there are x overlapping elements with hypergeometric distribution.

2.2.3. Degree-conserved network permutation

To prove whether embryo development genes tend to connect with disease genes, we performed degree-conserved network permutation for embryo development genes as following steps:

Step 1: We generated pseudo-embryo development gene sets for every embryo development genes, where the gene in the pseudo-embryo development gene sets has the same degree with corresponding embryo development gene in the PPI network.
Step 2: For the set of 516 embryo development genes, we randomly chose a gene from its corresponding pseudo-embryo development gene sets as pseudo-embryo development gene. Finally 516 pseudo-embryo development genes were generated. We used these pseudo-embryo development genes and disease genes to calculate the max connected components of pseudo-disease embryo development network.
Step 3: We repeated Step 2 1000 times and generated the degree-conserved pseudo-disease embryo development network.

We used the 1000 sets of pseudo-disease-embryo development network to substitute the disease-embryo development network as random control and calculated the significance of max connected components.

2.2.4. Disease co-morbidity index

In this paper, we use Φ index to measure disease co-morbidity and try to find out what the role of embryo development genes in disease co-morbidity. For disease i and disease j, the disease co-morbidity index was defined as

$$\Phi_{ij} = \frac{NN_{ij} - N_i N_j}{\sqrt{N_i N_j (N - N_i)(N - N_j)}} \quad (3)$$

where Φ_{ij} is co-morbidity index of disease i and j. $N = 13{,}039{,}018$ elderly patients and N_i represent the number of patients, who catch disease i.

3. Results

3.1. Construction of embryo development-related human disease network

We used embryo development genes and disease genes to construct a disease-embryo development network (DEN). The association pairs between diseases and genes were from the Online Mendelian Inheritance in Man (OMIM) (Hamosh et al., 2000, 2002; Supplementary Table S1). Human embryo development genes were extracted from Gene Ontology (GO) (Ashburner et al., 2000) and Entrez Gene (Maglott et al., 2005) by the keywords "embryo" and "development" (Supplementary Table S2). We generated the network, in which nodes represent known embryo development genes and disease genes, and two genes are connected if they associate with each other in Human Protein Reference Database (HPRD; Peri et al., 2004), and then extracted the maximum connected component of network as DEN (Fig. 1). In the DEN, nodes with black border represented embryo development genes, and the color of the nodes represents the disease class (Goh et al., 2007). In addition, if a disease gene belongs to more than one category, it will be labeled as "MD".

If the embryo development and disease processes are related to the cellular mechanisms, then embryo development genes and disease genes should be connected with each other in the network.

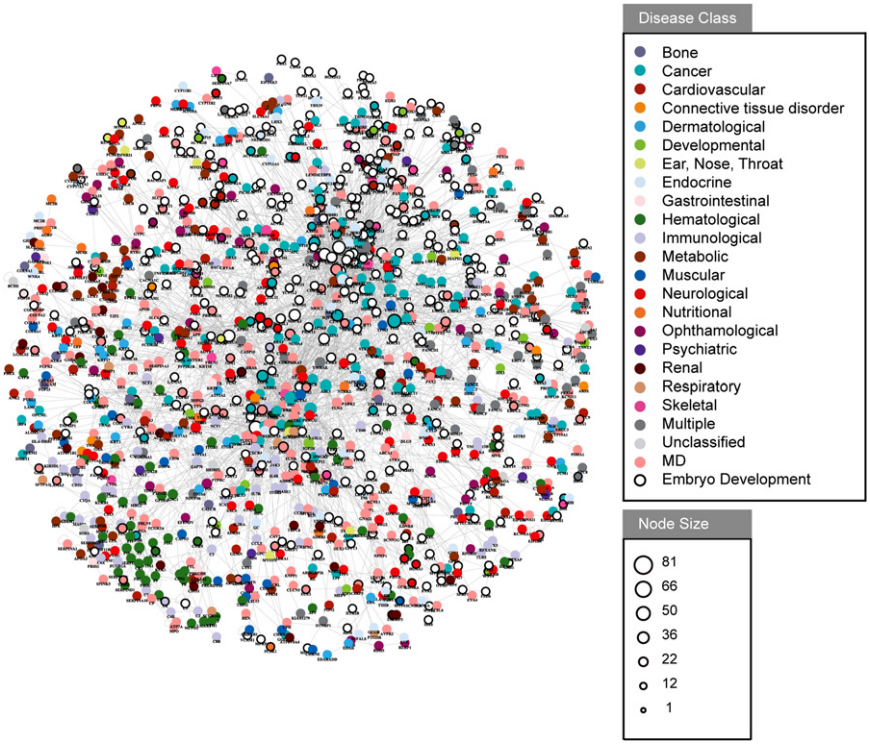


Fig. 1. The disease embryo development network (DEN). A protein–protein interacting network connects the disease genes and embryo development genes. In the DEN, disease genes are colored based on the disease categories to which it belongs, and embryo development genes are colored in white with blank border. A gene belonging to multi-disease class is labeled by "MD", and colored by pink. The size of each node is proportional to the number of interactions in the DEN.

As shown in Fig. 1, DEN contains 1257 nodes and 3617 edges, and the number of nodes and edges are significantly larger than random (p-values, 3.5e−3 for nodes and 3.6e−7 for edges, were calculated by 1000 degree-conserved network permutations, as shown in Fig. 2D and E). This indicates that embryo development genes and disease genes tend to connect with each other.

In the DEN, the degree of nodes followed power-law distribution ($P(k) \propto k^{-1.66}$; Fig. 2A), which indicates that the DEN is a scale free network, and can still maintain its connectivity, when a gene is removed randomly. Moreover, the average degree of embryo development genes is 7, which is larger than the average degree of disease genes 5.6 with a p-value 0.02 in the DEN (Wilcoxon rank sum test). Also, the average degree of genes, which are related to both embryo development and diseases, is 8.9, and higher than that of all genes 5.8 with a p-value 1.4e−3 in the network (Wilcoxon rank sum test). The results suggest that the embryo development genes may play an important role in the DEN.

Additionally, we calculated the clustering coefficient and topological coefficient of the DEN. In the theory of Barabasi AL and Oltvai ZN (Barabasi and Oltvai, 2004), the clustering coefficient measured the tendency of nodes to form clusters or groups in the network, and the topological coefficient is to describe the extent to which the node shares interaction with the other in the network. As shown in Fig. 2B, clustering coefficient of the node in DEN reduces with the increase the degree of node, implying the neighbors of high degree node do not tend to connect with each other, and DEN is a hierarchical network. In addition, the topological coefficient decreases with the link of nodes (see Fig. 2C). It obviously shows that either embryo development or disease hub genes do not share common interaction partners. We then used closeness centrality to measure which type of nodes are more 'central' in the network and found that the mean closeness centrality value of embryo development genes is larger than that of disease genes (p-value 8.0e−6).

3.2. Closeness between embryo development and diseases in many biological levels

To explore the relationship between embryo development and diseases, we firstly calculated the number of overlapping genes between them. 153 out of 516 embryo development genes in the PPI network are verified to be associated with some disease classes, significantly higher than expected (p-value $<$ 1.0e−20), indicating the close connection between embryo development and diseases (Supplementary Figure S1A). We then chose all human genes as a background set to recalculate the intersection between embryo development and diseases genes, and found that the count of overlapping genes between them are obviously higher than expected with p-value less than 1.0e−30 (Supplementary Figure S1B). In addition, we checked the link between embryo development and diseases genes. In total 39,118 edges in the PPI network, there are 858 edges belong to embryo development genes interactions, and 2304 edges belong to diseases gene interactions. The number of interactions between nodes which belong to both embryo development and disease is 152, three times as many as expected 50 (Supplementary Figure S1D).

The result of above shows indeed the close relationship between embryo development and disease, but is that all types

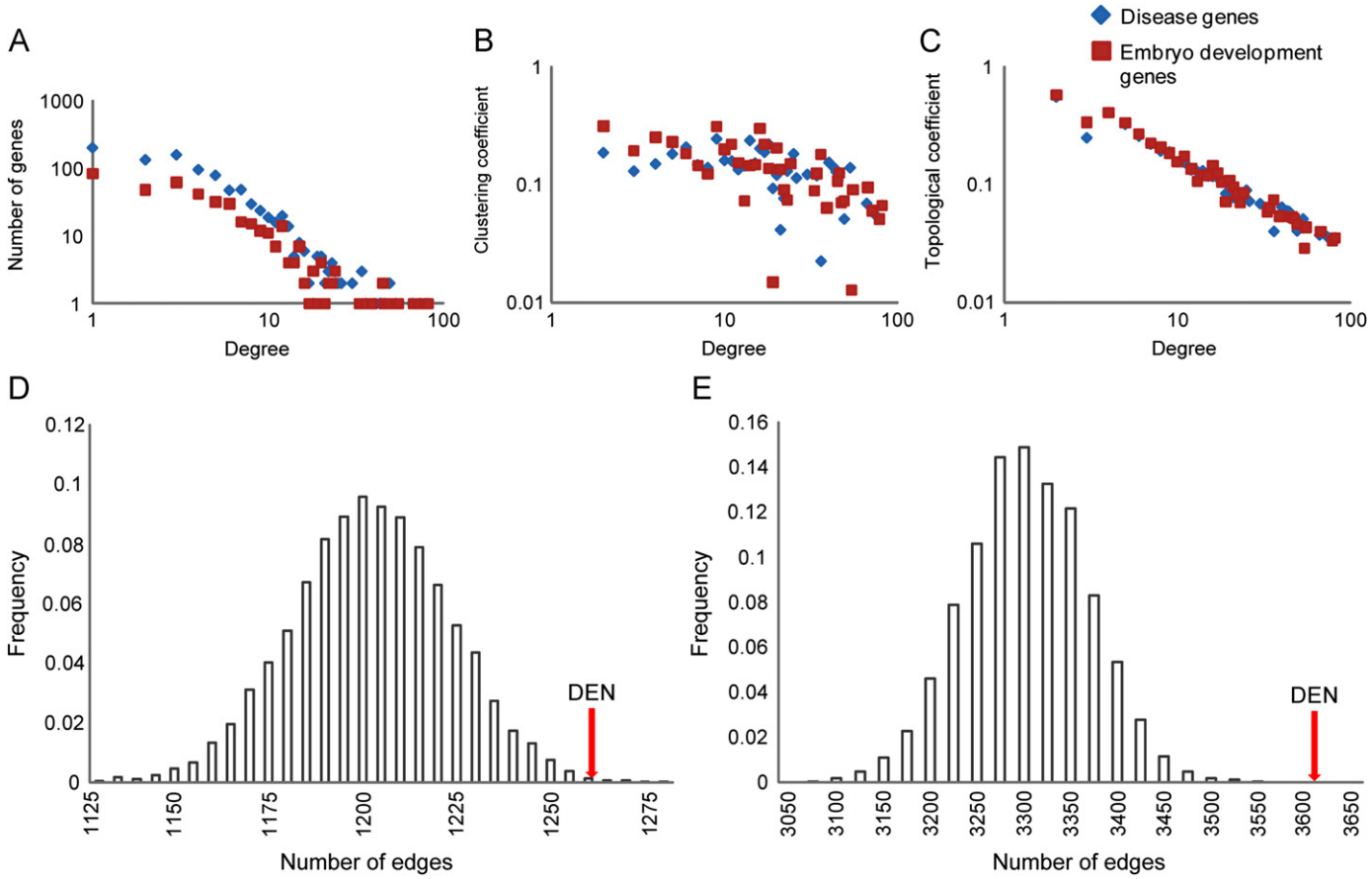
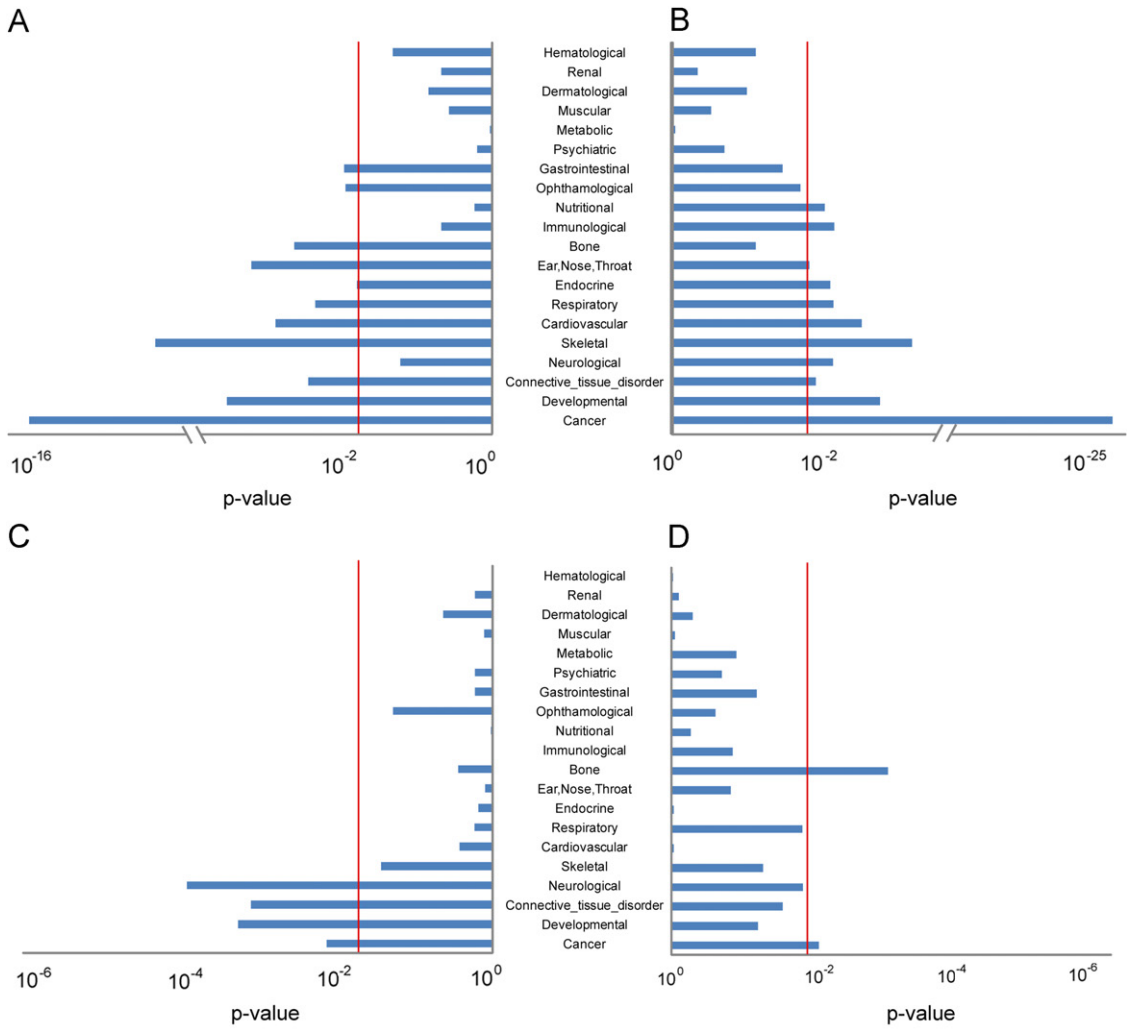


Fig. 2. The global topological properties of disease embryo development network (DEN). (A–C) Basic network features of disease embryo development network (DEN). (D) The number of nodes in DEN, which is significantly larger than that of degree-conserved random networks with p-value 3.5e−3. (E) The number of edges in DEN, which is significantly larger than that of degree-conserved random networks with p-value 3.6e−7. The process of generating the degree-conserved random network is depicted in Section 2.



E

Disease Classes	Gene Overlapping	Interact Pattern	microRNA	TF	Classification
Cancer	2.80E-16	2.13E-28	0.0323	0.0454	Embryo development High-related Diseases (EHD)
Developmental	3.54E-07	5.95E-05	0.0052	0.1623	
Connective tissue disorder	0.0084	0.0367	0.0068	0.0968	
Neurological	0.1147	0.0246	0.0018	0.0635	Embryo development Medium-related Diseases (EMD)
Skeletal	4.89E-12	8.80E-07	0.099	0.1461	
Cardiovascular	0.0037	1.54E-04	0.5027	0.9453	
Respiratory	0.0105	5.31E-04	0.6871	0.0645	
Endocrine	0.0406	0.0043	0.7401	0.9463	
Ear,Nose,Throat	4.96E-04	0.0454	0.8515	0.2858	
Bone	0.0061	0.1501	0.4906	0.001	
Immunological	0.3026	0.0035	NaN	0.2761	Embryo development Low-related Diseases (ELD)
Nutritional	0.6591	0.0218	0.9592	0.6663	
Ophthamological	0.023	0.0544	0.1277	0.396	
Gastrointestinal	0.0301	0.0817	0.6911	0.1679	
Psychiatric	0.7125	0.2958	0.6913	0.3476	
Metabolic	0.9988	0.9978	NaN	0.2537	
Muscular	0.3587	0.4152	0.8354	0.9209	
Dermatological	0.2232	0.1922	0.3595	0.6384	
Renal	0.2953	0.5398	0.6911	0.8527	
Hematological	0.0965	0.1449	NaN	0.9644	

Fig. 3. The closeness between embryo development and disease in many different biological levels. (A–D) Bar graph to show the statistical significance of closeness between embryo development and disease in gene overlapping (A), interacting partner of human PPI network (B), sharing microRNA (C) and sharing TF (D). Refer to Section 2 for details. (E) Detailed closeness in many ways. According to the number of significant closeness in different biological level, all kinds of disease are categorized into three groups of diseases. High-related embryo development disease (red shadow), which is significant in more than three levels. Medium-related embryo development disease (green shadow), which is significant in two levels. And low-related embryo development disease (blue shadow), which is significant in less than one level. "NaN" represents that the disease classed are not enriched in any of the microRNAs.

of diseases relate to embryo development process? To answer this question, we calculated the overlapping between embryo development and all disease classes from different biological levels, including gene sharing, interaction partner of PPI network, post-transcriptional modification by microRNAs and transcriptional regulation by transcription factors (TF).

The significance of overlapping between embryo development genes and disease genes is shown in Fig. 3A. Ten diseases, including cancer, developmental disorder etc., are significantly related with embryo development ($p < 0.05$, see Section 2). Then we checked the association between disease classes and embryo development by calculating the percentage of interaction partners in the human PPI network. The result indicates that eleven disease classes tend to associate with embryo development (see Fig. 3B). Furthermore, we investigated co-regulated relationship between disease classes and embryo development. As shown in Fig. 3C and D, four classes of disease and embryo development are co-regulated by microRNA; two classes of disease and embryo development are co-regulated by TF (see Section 2).

3.3. The difference among three groups of diseases

If there are close relationships between embryo development and some classes of diseases, the closeness between embryo development genes and diseases genes will be reflected in different biological levels. In our study, some kinds of disease show unequal closeness with embryo development in different biological levels. So we divided all disease classes into three groups, as shown in Fig. 3E, embryo development high-related disease (EHD), medium-related disease (EMD) and low-related disease (ELD), according to the closeness with embryo development in different biological levels. From previous investigations, cancer has shown close relationship with embryo development in many aspects (John et al., 2008; Kim et al., 2010; Lalli and Alonso, 2010; Monzo et al., 2008; Navarro and Monzo, 2010; Ullmann, 2010). And as results shown, cancer has very better significance than other kinds of disease in each level, which is consistent with previous investigations. Also, skeletal diseases are significant in gene overlapping and interact pattern level and some research support it (Merrick et al., 2009; See et al., 2008). But some of them, like metabolic disease, show few or even none relationship with embryo development. Anyway, we combined different p-values in each level to classify the disease and analyzed the difference between them.

Firstly, the comparison of the closeness centrality among them in the human PPI network shows that the EHD genes are central in human PPI network, but EMD genes and ELD genes are not. As shown in Fig. 4A, EHD genes have a significantly higher closeness centrality (0.2697) than EMD genes (0.2493) and ELD genes (0.2423) with p-value 3.1e−11 and 3.0e−21, respectively. We also compared the percentage of essential genes in three groups, and found that 60.8% of EHD genes are essential genes, which is significantly higher than 50.4% of EMD genes (p-value 7.1e−3) and 38.9% of ELD genes (p-value 1.6e−8).

Secondly, we performed a GO enrichment to analyze the biological functions of EHD genes, EMD genes and ELD genes. As shown in Table 1, EHD genes are significantly enriched in DNA binding and cell cycle phase; EMD genes are significantly enriched in negative regulation of blood pressure and cell differentiation; and ELD genes are enriched in cellular carbohydrate

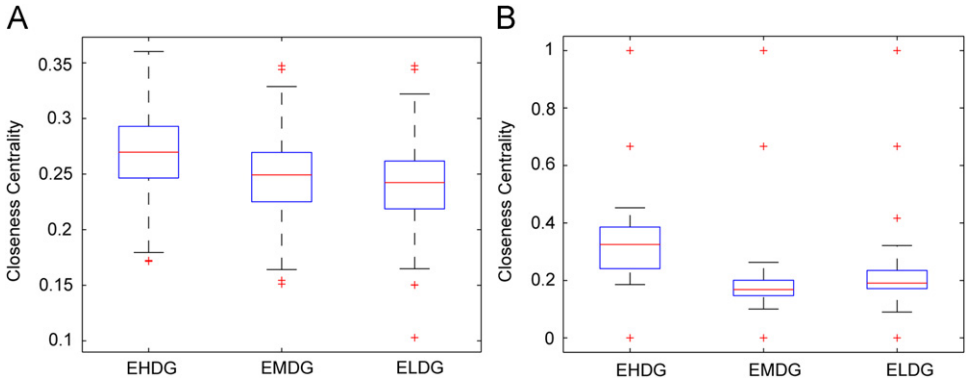


Fig. 4. The difference among EHD, EMD and ELD in network closeness centrality. (A) The comparison of network closeness centrality among three groups of disease in human PPI network. (B) The difference of network closeness centrality of embryo development genes in the network of three groups of diseases genes.

Table 1
Different GOA enrichments among EHD genes, EMD genes and ELD gene.

GO-ID	High	Medium	Low	Description
	p-value	*p*-value	*p*-value	
3676	6.9e−8	–	2.1e−6 (under)	Nucleic acid binding
5634	9.3e−11	–	1.3e−5 (under)	Nucleus
3690	2.3e−12	>0.01	>0.01	Double-stranded DNA binding
22403	7.2e−10	>0.01	>0.01	Cell cycle phase
5654	1.6e−8	>0.01	>0.01	Nucleoplasm
45776	>0.01	4.3e−11	>0.01	Negative regulation of blood pressure
7404	>0.01	4.7e−9	>0.01	Glial cell differentiation
6575	>0.01	4.7e−7	>0.01	Amino acid derivative metabolic process
55072	>0.01	>0.01	1.7e−19	Cellular carbohydrate metabolic process
9056	>0.01	>0.01	5.8e−13	Catabolic process
2250	>0.01	>0.01	1.1e−14	Adaptive immune response

EHD genes, EMD genes and ELD genes display difference in GOA enrichment. p-values followed by "under" represent underrepresentation, while others mean overrepresentation.

metabolic process, adaptive immune response, etc. Also, we compared the closeness centrality of embryo development genes in the network which constructed by three groups of genes respectively. For each of the three groups of diseases, networks are constructed in which two genes are connected if they associate in HPRD. As shown in Fig. 4B, the median of closeness centrality of embryo development genes in network of EHD is 0. 3249, and significant higher than that in network of EMD which is 0.1681 (p-value 1.2e−3, rank sum test) and ELD which is 0.1909 (p-value 3.0e−3 rank sum test). This indicates that embryo development genes tend to be the central of EHD network.

Finally, we used BLAST alignment sequence program to investigate the similarity of protein sequences between embryo development genes and three types of disease genes. With E-value less than 1.0e−10, the mean alignment sequence score of EHD genes with embryo development genes is 232, significantly higher than EMD genes (score=181, with p-value 4.2e−54) and ELD genes (score=175, with p-value 1.5e−69). And the mean alignment sequence score of EMD genes is also high than ELD genes with p-value 3.8e−4. Additionally, we recalculated the same process with E-value less than 1.0e−5, and the mean alignment sequence score of EHD genes is also significantly higher than other types of disease genes. The results indicate that the EHD genes have more similarity with embryo development genes at sequence similarity.

4. Conclusion and discussion

There is an expectation for clinicians and medical researchers to explore the relationship between embryo development and diseases, and a great deal of works have been done for it. Here we constructed the disease-embryo development network for the first time, and investigate the relationship between embryo development and disease. Our results demonstrate that there are close relationships between disease genes and embryo development genes from the perspective of network, which supplies new biological evidences for disease studies. Also, we divided all diseases into three groups according to the closeness with embryo development and compared the differences among them including closeness of PPI network, enrichment biological function, percentage of essential genes and protein sequence similarity. In our results, cancer, developmental and connective tissue diseases show strong correlation with embryo development, and recent studies coincide with our results. For example, John et al. (2008) found that embryo development gene DPPA2 is coexpressed with cancer testis antigens in non-small cell lung cancer, and some evidence show that developmental diseases are related with embryo development (Sinclair and Singh, 2007). At the same time, we found some evidences to support the correlation between embryo development and disease in gene expression level (John et al., 2008; Monk et al., 2008; Wang et al., 2008). All above these could certificate that embryo development and disease indeed have association and different kinds of disease have different relationship with embryo development. The result can imply that we can further investigate the pathogenesis of disease from perspective of similarity with embryo development, especially with high-related embryo development disease.

Previous investigations have demonstrated that different diseases tend to be co-morbid if they have a similar cellular mechanism, including component of the genome, transcriptome, proteome and metabolome in the pathogenesis (Lee et al., 2008). We analyzed the effect of embryo development genes on disease co-morbidity using the data from Park et al., 2009 studies, and found that there is a little difference in co-morbidity incidence of diseases between diseases, which share or not share embryo development genes. We used Φ to represent co-morbidity incidence of diseases (see Section 2). $\overline{\Phi} = 0.013$ for diseases that share embryo development genes, and 0.01 for diseases that do not share embryo development genes. It indicates that the diseases that share embryo development genes are apt to appearance together in the population.

Certainly, we noted that there is some potential bias in the source of data such as the preference of disease gene association and the incompleteness of embryo development genes. For instance, the co-morbidity incidences have no significant difference among three groups of diseases, whether or not they share embryo development genes. One explanation is that the metabolism may be another biological mechanism prompting disease co-morbidity (Lee et al., 2008) besides embryo development. We carefully studied the metabolism-related diseases, and found that 66.5% of diseases belong to metabolism-related diseases. In addition to the influence induced by other biological mechanisms, a number of influences, such as the environmental factor, lifestyle and treatment-related factors, may also alter the impact of the correlation between embryo development and disease. Moreover, according to a recent comprehensive review (Chou, 2011), to develop a useful model or predictor for biological systems, the following things were usually needed to consider: (i) benchmark dataset construction or selection, (ii) mathematical formulation for biological sequence samples, (iii) operating algorithm (or engine), (iv) anticipated accuracy and (v) web-server establishment. So we used the embryo development genes and disease genes to prove the relationship between embryo development and disease based on the steps that we mentioned above. Since user-friendly and publicly accessible web-servers represent the future direction for developing practically more useful models, simulated methods or predictors (Chou and Shen, 2009), we shall make efforts in our future work to provide a web-server for the method presented in this paper. Meanwhile, with the development of high-throughput molecular biotechnology, especially mature application of the next-generation sequencing, we can further study the etiologies and pathobiologies of disease using embryo development process from systems biology viewpoint.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (Grant Nos. 30871394, 61073136 and 91029717), the National High Tech Development Project of China, the 863 Program (Grant No. 2007AA02Z329) and the National Science Foundation of Heilongjiang Province (Grant Nos. ZD200816-01, ZJG0501, GB03C602-4, JC200711 and YJSCX2007-0195HLJ).

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2011.07.018.

References

Albert, R., 2005. Scale-free networks in cell biology. J. Cell Sci. 118, 4947–4957.
Almaas, E., 2007. Biological impacts and context of network theory. J. Exp. Biol. 210, 1548–1558.
Alon, U., 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8, 450–461.
Althaus, I.W., Chou, J.J., Gonzales, A.J., Deibel, M.R., Chou, K.C., Kezdy, F.J., Romero, D.L., Aristoff, P.A., Tarpley, W.G., Reusser, F., 1993a. Steady-state kinetic studies with the non-nucleoside HIV-1 reverse transcriptase inhibitor U-87201E. J. Biol. Chem. 268, 6119–6124.
Althaus, I.W., Gonzales, A.J., Chou, J.J., Romero, D.L., Deibel, M.R., Chou, K.C., Kezdy, F.J., Resnick, L., Busso, M.E., So, A.G., et al., 1993b. The quinoline U-78036 is a potent inhibitor of HIV-1 reverse transcriptase. J. Biol. Chem. 268, 14875–14880.

Althaus, I.W., Chou, J.J., Gonzales, A.J., Deibel, M.R., Chou, K.C., Kezdy, F.J., Romero, D.L., Palmer, J.R., Thomas, R.C., Aristoff, P.A., et al., 1993c. Kinetic studies with the non-nucleoside HIV-1 reverse transcriptase inhibitor U-88204E. Biochemistry 32, 6548–6554.

Argmann, C.A., Chambon, P., Auwerx, J., 2005. Mouse phenogenomics: the fast track to "systems metabolism". Cell Metab. 2, 349–360.

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., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.

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

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., Califano, A., 2005. Reverse engineering of regulatory networks in human B cells. Nat. Genet. 37, 382–390.

Braun, P., Rietman, E., Vidal, M., 2008. Networking metabolites and diseases. Proc. Natl. Acad. Sci. USA 105, 9849–9850.

Chou, K.C., 1981. Two new schematic rules for rate laws of enzyme-catalysed reactions. J. Theor. Biol. 89, 581–592.

Chou, K.C., 1989. Graphic rules in steady and non-steady state enzyme kinetics. J. Biol. Chem. 264, 12074–12079.

Chou, K.C., 1990. Applications of graph theory to enzyme kinetics and protein folding kinetics. Steady and non-steady-state systems. Biophys. Chem. 35, 1–24.

Chou, K.C., 2010. Graphic rule for drug metabolism systems. Curr. Drug Metab. 11, 369–378.

Chou, K.C., 2011. Some remarks on protein attribute prediction and pseudo amino acid composition. J. Theor. Biol. 273, 236–247.

Chou, K.C., Forsen, S., 1980. Graphical rules for enzyme-catalysed rate laws. Biochem. J. 187, 829–835.

Chou, K.C., Cai, Y.D., 2006. Predicting protein-protein interactions from sequences in a hybridization space. J. Proteome Res. 5, 316–322.

Chou, K.C., Shen, H.B., 2009. REVIEW : Recent advances in developing web-servers for predicting protein attributes. Nat. Sci. 1, 63–92.

Chou, K.C., Kezdy, F.J., Reusser, F., 1994. Kinetics of processive nucleic acid polymerases and nucleases. Anal. Biochem. 221, 217–230.

Chou, K.C., Zhang, C.T., Maggiora, G.M., 1997. Disposition of amphiphilic helices in heteropolar environments. Proteins 28, 99–108.

Cusick, M.E., Klitgord, N., Vidal, M., Hill, D.E., 2005. Interactome: gateway into systems biology. Hum. Mol. Genet. 14 (Spec No. 2), R171–R181.

Davidson, E.H., 2009. Network design principles from the sea urchin embryo. Curr. Opin. Genet. Dev. 19, 535–540.

Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., Miller, J., Schlaeger, T., Daley, G.Q., Feinberg, A.P., 2009. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. 41, 1350–1353.

Ergun, A., Lawrence, C.A., Kohanski, M.A., Brennan, T.A., Collins, J.J., 2007. A network biology approach to prostate cancer. Mol. Syst. Biol. 3, 82.

Feldman, I., Rzhetsky, A., Vitkup, D., 2008. Network properties of genes harboring inherited disease mutations. Proc. Natl. Acad. Sci. USA 105, 4323–4328.

Friedman, A., Perrimon, N., 2007. Genetic screening for signal transduction in the era of network biology. Cell 128, 225–231.

Gerstein, M., Lan, N., Jansen, R., 2002. Proteomics. Integrating interactomes. Science 295, 284–287.

Giallourakis, C., Henson, C., Reich, M., Xie, X., Mootha, V.K., 2005. Disease gene discovery through integrative genomics. Annu. Rev. Genomics Hum. Genet. 6, 381–406.

Giuffrida, D., Rogers, I.M., Nagy, A., Calogero, A.E., Brown, T.J., Casper, R.F., 2009. Human embryonic stem cells secrete soluble factors that inhibit cancer cell growth. Cell Proliferation 42, 788–798.

Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabasi, A.L., 2007. The human disease network. Proc. Natl. Acad. Sci. USA 104, 8685–8690.

Gonzalez-Diaz, H., Perez-Montoto, L.G., Duardo-Sanchez, A., Paniagua, E., Vazquez-Prieto, S., Vilas, R., Dea-Ayuela, M.A., Bolas-Fernandez, F., Munteanu, C.R., Dorado, J., Costas, J., Ubeira, F.M., 2009. Generalized lattice graphs for 2D-visualization of biological information. J. Theor. Biol. 261, 136–147.

Groenendijk, B.C., Stekelenburg-de Vos, S., Vennemann, P., Wladimiroff, J.W., Nieuwstadt, F.T., Lindken, R., Westerweel, J., Hierck, B.P., Ursem, N.T., Poelmann, R.E., 2008. The endothelin-1 pathway and the development of cardiovascular defects in the haemodynamically challenged chicken embryo. J. Vasc. Res. 45, 54–68.

Hamosh, A., Scott, A.F., Amberger, J., Valle, D., McKusick, V.A., 2000. Online Mendelian Inheritance in Man (OMIM). Hum. Mutat 15, 57–61.

Hamosh, A., Scott, A.F., Amberger, J., Bocchini, C., Valle, D., McKusick, V.A., 2002. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 30, 52–55.

Hu, L., Huang, T., Shi, X., Lu, W.C., Cai, Y.D., Chou, K.C., 2011. Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. PLoS One 6, e14556.

Ideker, T., Sharan, R., 2008. Protein networks in disease. Genome Res. 18, 644–652.

John, T., Caballero, O.L., Svobodova, S.J., Kong, A., Chua, R., Browning, J., Fortunato, S., Deb, S., Hsu, M., Gedye, C.A., Davis, I.D., Altorki, N., Simpson, A.J., Chen, Y.T., Monk, M., Cebon, J.S., 2008. ECSA/DPPA2 is an embryo-cancer antigen that is coexpressed with cancer-testis antigens in non-small cell lung cancer. Clin. Cancer Res. 14, 3291–3298.

Kann, M.G., 2007. Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform 8, 333–346.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., Orkin, S.H., 2010. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell 143, 313–324.

Lage, K., Karlberg, E.O., Storling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tumer, Z., Pociot, F., Tommerup, N., Moreau, Y., Brunak, S., 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316.

Lalli, E., Alonso, J., 2010. Targeting DAX-1 in embryonic stem cells and cancer. Expert Opin. Ther. Targets 14, 169–177.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S.A., Haggarty, S.J., Clemons, P.A., Wei, R., Carr, S.A., Lander, E.S., Golub, T.R., 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935.

Lee, D.S., Park, J., Kay, K.A., Christakis, N.A., Oltvai, Z.N., Barabasi, A.L., 2008. The implications of human metabolic network topology for disease comorbidity. Proc. Natl. Acad. Sci. USA 105, 9880–9885.

Li, Y., Agarwal, P., 2009. A pathway-based view of human diseases and disease relationships. PLoS One 4, e4346.

Loscalzo, J., Kohane, I., Barabasi, A.L., 2007. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. Mol. Syst. Biol. 3, 124.

Ma, Y.L., Zhang, P., Wang, F., Moyer, M.P., Yang, J.J., Liu, Z.H., Peng, J.Y., Chen, H.Q., Zhou, Y.K., Liu, W.J., Qin, H.L., 2010. Human embryonic stem cells and metastatic colorectal cancer cells shared the common endogenous human microRNA-26b. J. Cell Mol. Med.

Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T., 2005. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 33, D54–D58.

Merrick, D., Stadler, L.K., Larner, D., Smith, J., 2009. Muscular dystrophy begins early in embryonic development deriving from stem cell loss and disrupted skeletal muscle formation. Dis. Model Mech. 2, 374–388.

Monk, M., Hitchins, M., Hawes, S., 2008. Differential expression of the embryo/cancer gene ECSA(DPPA2), the cancer/testis gene BORIS and the pluripotency structural gene OCT4, in human preimplantation development. Mol. Hum. Reprod. 14, 347–355.

Monzo, M., Navarro, A., Bandres, E., Artells, R., Moreno, I., Gel, B., Ibeas, R., Moreno, J., Martinez, F., Diaz, T., Martinez, A., Balague, O., Garcia-Foncillas, J., 2008. Overlapping expression of microRNAs in human embryonic colon and colorectal cancer. Cell. Res. 18, 823–833.

Navarro, A., Monzo, M., 2010. MicroRNAs in human embryonic and cancer stem cells. Yonsei Med. J. 51, 622–632.

Oti, M., Huynen, M.A., Brunner, H.G., 2008. Phenome connections. Trends Genet. 24, 103–106.

Park, J., Lee, D.S., Christakis, N.A., Barabasi, A.L., 2009. The impact of cellular networks on disease comorbidity. Mol. Syst. Biol. 5, 262.

Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T.K., Chandrika, K.N., Deshpande, N., Suresh, S., Rashmi, B.P., Shanker, K., Padma, N., Niranjan, V., Harsha, H.C., Talreja, N., Vrushabendra, B.M., Ramya, M.A., Yatish, A.J., Joy, M., Shivashankar, H.N., Kavitha, M.P., Menezes, M., Choudhury, D.R., Ghosh, N., Saravana, R., Chandran, S., Mohan, S., Jonnalagadda, C.K., Prasad, C.K., Kumar-Sinha, C., Deshpande, K.S., Pandey, A., 2004. Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. 32, D497–D501.

Postovit, L.M., Margaryan, N.V., Seftor, E.A., Kirschmann, D.A., Lipavsky, A., Wheaton, W.W., Abbott, D.E., Seftor, R.E., Hendrix, M.J., 2008. Human embryonic stem cell microenvironment suppresses the tumorigenic phenotype of aggressive cancer cells. Proc. Natl. Acad. Sci. USA 105, 4329–4334.

Relaix, F., 2006. Skeletal muscle progenitor cells: from embryo to adult. Cell. Mol. Life Sci. 63, 1221–1225.

Ruiz-Vela, A., Aguilar-Gallardo, C., Simon, C., 2009. Building a framework for embryonic microenvironments and cancer stem cells. Stem. Cell. Rev. 5, 319–327.

Schadt, E.E., Lamb, J., Yang, X., Zhu, J., Edwards, S., Guhathakurta, D., Sieberts, S.K., Monks, S., Reitman, M., Zhang, C., Lum, P.Y., Leonardson, A., Thieringer, R., Metzger, J.M., Yang, L., Castle, J., Zhu, H., Kash, S.F., Drake, T.A., Sachs, A., Lusis, A.J., 2005. An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. 37, 710–717.

See, A.W., Kaiser, M.E., White, J.C., Clagett-Dame, M., 2008. A nutritional model of late embryonic vitamin A deficiency produces defects in organogenesis at a high penetrance and reveals new roles for the vitamin in skeletal development. Dev. Biol. 316, 171–190.

Sharma, S.K., Lucitti, J.L., Nordman, C., Tinney, J.P., Tobita, K., Keller, B.B., 2006. Impact of hypoxia on early chick embryo growth and cardiovascular function. Pediatr. Res. 59, 116–120.

Shen, Y.Z., Ding, Y.S., Gu, Q., Chou, K.C., 2010. Identifying the hub proteins from complicated membrane protein network systems. Med. Chem. 6, 165–173.

Sinclair, K.D., Singh, R., 2007. Modelling the developmental origins of health and disease in the early embryo. Theriogenology 67, 43–53.

Smith, D.L., Nilar, S.H., 2010. Homology modeling studies of yeast Mitogen-Activated Protein Kinases (MAPKS): structural motifs as a basis for specificity. Protein Pept. Lett. 17, 732–738.

Takebe, N., Ivy, S.P., 2010. Controversies in cancer stem cells: targeting embryonic signaling pathways. Clin. Cancer Res. 16, 3106–3112.

Tu, Z., Wang, L., Xu, M., Zhou, X., Chen, T., Sun, F., 2006. Further understanding human disease genes by comparing with housekeeping genes and other genes. BMC Genomics 7, 31.

Tutarel, O., Norozi, K., Hornung, O., Orhan, G., Wubbolt-Lehmann, P., Wessel, A., Yelbuz, T.M., 2005. Images in cardiovascular medicine. Cardiac failure in the chick embryo resembles heart failure in humans. Circulation 112, e352–e353.

Ullmann, C.D., 2010. Cancer stem cells and targeting embryonic signaling pathways. J. Thorac. Oncol. 5, S492–S494.

Wang, F., Li, X., Xie, X., Zhao, L., Chen, W., 2008. UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. FEBS Lett. 582, 1919–1927.

Wang, J., Zhang, S., Wang, Y., Chen, L., Zhang, X.S., 2009. Disease-aging network reveals significant roles of aging genes in connecting genetic diseases. PLoS Comput. Biol. 5, e1000521.

Wang, M., Yao, J.S., Huang, Z.D., Xu, Z.J., Liu, G.P., Zhao, H.Y., Wang, X.Y., Yang, J., Zhu, Y.S., Chou, K.C., 2005. A new nucleotide-composition based fingerprint of SARS-CoV with visualization analysis. Med. Chem. 1, 39–47.

Wolfram, S., 1984. Cellular automation as models of complexity. Nature 311, 419–424.

Wu, Z.C., Xiao, X., Chou, K.C., 2010. 2D-MH: a web-server for generating graphic representation of protein sequences based on the physicochemical properties of their constituent amino acids. J. Theor. Biol. 267, 29–34.

Xiao, X., Shao, S.H., Chou, K.C., 2006. A probability cellular automaton model for hepatitis B viral infections. Biochem. Biophys. Res. Commun. 342, 605–610.

Xiao, X., Wang, P., Chou, K.C., 2008. Predicting protein structural classes with pseudo amino acid composition: an approach using geometric moments of cellular automaton image. J. Theor. Biol. 254, 691–696.

Xiao, X., Wang, P., Chou, K.C., 2009. GPCR-CA: a cellular automaton image approach for predicting G-protein-coupled receptor functional classes. J. Comput. Chem. 30, 1414–1423.

Xiao, X., Shao, S., Ding, Y., Huang, Z., Chen, X., Chou, K.C., 2005a. Using cellular automata to generate image representation for biological sequences. Amino Acids 28, 29–35.

Xiao, X., Shao, S., Ding, Y., Huang, Z., Chen, X., Chou, K.C., 2005b. An application of gene comparative image for predicting the effect on replication ratio by HBV virus gene missense mutation. J. Theor. Biol. 235, 555–565.

Xie, G., Mo, Z., 2011. Three 3D graphical representations of DNA primary sequences based on the classifications of DNA bases and their applications. J. Theor. Biol. 269, 123–130.

Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L., Vidal, M., 2007. Drug-target network. Nat. Biotechnol. 25, 1119–1126.

Yu, J.F., Sun, X., Wang, J.H., 2009. TN curve: a novel 3D graphical representation of DNA sequence based on trinucleotides and its applications. J. Theor. Biol. 261, 459–468.

Zhou, G.P., Deng, M.H., 1984. An extension of Chou's graphic rules for deriving enzyme kinetic equations to systems involving parallel reaction pathways. Biochem. J. 222, 169–176.

Zhu, X., Gerstein, M., Snyder, M., 2007. Getting connected: analysis and principles of biological networks. Genes Dev. 21, 1010–1024.