SPANDIDOS PUBLICATIONS

Enriching protein-protein and functional interaction networks in human embryonic stem cells

CHANGQING ZUO[1], SHUANG LIANG[1], ZONGGUI WANG[2], HUA LI[1], WENLING ZHENG[3] and WENLI MA[1]

[1]Institute of Genetic Engineering, Southern Medical University, Guangzhou, Guangdong 510515;
[2]Department of Biochemistry and Molecular Biology, Guangdong Medical College, Dongguan, Guangdong 523808;
[3]Southern China Genomics Reseach Center, Guangzhou, Guangdong 510800, P.R. China

Received January 16, 2009; Accepted March 6, 2009

DOI: 10.3892/ijmm_00000197

Abstract. Human Embryonic Stem Cells (hESCs) have a great therapeutic potential in regenerative medicine, but the precise molecular mechanisms by which hESCs maintain or regulate their characteristics remain largely unknown. Since protein-protein interaction is vitally important in regulating hESCs, we utilized a network-based bioinformatics analysis in order to learn what and how specific proteins interact with each other. By combining protein-protein interaction data and a collection of genes over-expressed in hESCs, we constructed a protein interaction network using a breadth-first search algorithm. This scale-free network which is significantly larger than networks generated by random samplings, illustrates how these hESC-enriched proteins might interact with each other in hESCs. Of the top 5% highly connected nodes (corresponding to 21 proteins including MYC, H2AFX, RUVBL1, DDX18, CDC2, HDAC2 and HIST1H4C) presumably critical for determining the fate of hESCs, nearly half are known to be regulated by NANOG/SOX2/MYC. This underscores importance of these transcription factors in hESCs. In addition, *in silico* cis-element analysis suggests that NF-Y may be an important transcription factor regulating many of these hub proteins (high connected nodes) in hESCs. To further abstract the functional significance, directly connected proteins were matched to and grouped by gene ontology (GO) terms in molecular function category. Sixty-six interacting GO-GO terms paired through protein interactions were found over-represented in hESCs. This functional enrichment may be essential for understanding molecular characteristics in hESCs. Collectively, we analyzed hESC-enriched genes based on protein-protein interaction data, from which an hESC-enriched protein interaction network was constructed and a network of molecular functional terms was also identified. The results of this analysis, on the systems level, may shed new light to further our understanding of hESCs.

Correspondence to: Dr Wenli Ma, Institute of Genetic Engineering, Southern Medical University, Guangzhou, Guangdong 510515, P.R. China
E-mail: wenli668@gmail.com

Key words: scare-free network, bioinformatics, protein interaction, gene ontology, embryonic stem cell

Introduction

Human Embryonic Stem Cells (hESCs) are pluripotent cells isolated from the inner cell mass of the blastocyst (1). They can differentiate into all tissue types *in vivo*, and into the three primary germ layers as well as extra-embryonic tissues *in vitro*. Because of these unique characteristics, these cells are potentially invaluable in regenerative medicine (2-5). Efforts on this have identified some important genes in hESC lines, such as OCT3/4, NANOG, REX1, SOX2 and FOXD3 (6-10). However, our knowledge concerning the many characteristics of hESCs is still limited. One main reason may be that the vast majority of studies have focused on individual genes/proteins, without considering the possible role of interactions. It is well known that most proteins do not function in isolation, but rather interact with one another to form molecular networks, and larger protein complexes are, in turn, part of a more extensive biological web.

With the evolution of high-throughput technologies in the post-genomics era, studies have shifted from characterization of single protein to investigation of the entire interactome. Nowadays, biological knowledge is often represented by networks, such as regulatory and metabolic networks. Construction and analyses of these networks have revealed interesting characteristics within the framework of interactome. For example, hub proteins in networks tend to be more conserved (11,12). There are an increasing number of studies focusing on the transcriptional networks, which emphasize the roles of transcription factors, in regulating human ES cells. As a result, a number of important transcription factors responsible for self-renewal and pluripotency have been identified (13-16). In recent years, the network-based approach has gained popularity and been successfully applied for analysis of protein-protein interaction networks in many species and diseases (17-19). For instance, one study using affinity purification of NANOG under native conditions followed by mass spectrometry generated a mouse ES protein interaction network including only 37 proteins (20). However, the picture is far from complete.

Microarray technology provides us a unique opportunity to examine gene expression patterns in hESCs. However,

heterogeneity of hESCs gene expression data could exist across different laboratories or different cell lines, which could be partly circumvented by meta-analysis to give a more robust result. A 'consensus hESCs gene list' was produced (21), and some of the genes on the list are also expressed in certain other tissues. The authors explain that these proteins may not be highly specific to hESCs individually, however, they act together with other proteins to function specifically in hESCs (21). Based on these observations, we postulate that there may be an enriched protein interaction sub-network(s) among a collection of those on or off the list to maintain or to modify properties of hESCs.

In this study, two questions were raised: i) Do the hESC-enriched genes/proteins interact directly with each other more frequently than expected by chance alone? In other words, does an hESC-enriched protein interaction network exist at all? ii) Can any enriched functional interaction patterns be identified to be important for maintaining characteristics of hESCs? To address these questions, we performed network-based analysis of both enriched genes/proteins and ontological terms to propose the existence of an hESC-enriched molecular interaction network.

Materials and methods

Gene list. We collected genes over-expressed in hESCs in at least 3 independent studies, re-affirmed by a meta-analysis of human embryonic stem cells transcriptome - Amazonia (http://amazonia.montp.inserm.fr/) (21). We only collected those with known entrez gene id from the original 'consensus hESC gene list' called by the authors. The resulting list includes 1029 unique gene ids. We converted these gene ids to 1020 UniProtKB/Swiss-Prot accession numbers using UniProt ID mapping (http://www.uniprot.org/jobs) and named them hESC enriched proteins (hESPs).

Protein-protein interaction data. To obtain protein-protein interaction data, we downloaded the i2d database for human (http://ophid.utoronto.ca/ophidv2.201/) (12), which contains a new release of the Online Predicted Human Interaction Database (OPHID), and human protein-protein interactome assembled from other databases complemented by homolog interactions identified in other organisms (22). It now includes 138554 protein interaction pairs for 13560 proteins in UniProtKB/Swiss-Prot. We wrote a Perl script to remove reciprocally redundant pairs (for example: proteins A and B can form pairs A-B and B-A with one of them removed). We obtained 92545 unique protein interaction pairs (UPI). To the best of our knowledge, this is perhaps the largest protein interaction network for humans.

Tissue specificity/selectivity. Tissue-specific/selective gene expression is believed to be of physiological importance (23). To verify whether hESC-enriched genes are tissue-specific/selective or not, we compared them with 3904 tissue-selective genes surveying 97 tissue types (stem cell and other related tissue or cell types not surveyed) (24).

Network construction and analysis. Network for hESPs was constructed using a Perl script (named Max Network Program, MNP). The program took as input human protein-protein interaction data downloaded from the i2d database, and mapped interaction of hESPs within the entire protein-protein interaction network using a breadth-first search algorithm. The resulting subnet, named as Max Network Protein Interaction Pairs or MNPIP, was visually rendered by Cytoscape program (http://www.cytoscape.org/) (25). To assess the significance of MNPIP, 1000 simulations were done with equal number of input proteins randomly selected. The free statistical package R (http://www.r-project.org/) was used to compute Wilcoxon signed rank test statistics for difference between MNPIP and sub-networks randomly generated, with a $p<0.05$ considered as significant.

Hub analysis. We classified nodes in the network according to the degree of connectivity. The top 5% of the most connected proteins in this network were regarded as hubs. Promoter sequence (from 1200 bp upstream to 200 bp downstream of the transcription start site, TSS) analysis was done using Gather (http://gather.genome.duke.edu/) and TFM-Explorer (http://bioinfo.lifl.fr/TFME/).

Enriched functional interaction pairs and network. Interacting partners are likely to be functionally related. To identify functional enrichment in hESCs, we assigned GOSlim function terms to human UiprotKB/Swiss-Prot proteins in i2d database whenever possible. A total of 42 GOSlim function terms from QuickGO (www.ebi.ac.uk/ego) were used. Proteins assigned with the same GOSlim term were considered functionally similar. One protein could have multiple GOSlim assignments if it is involved in various functional aspects. Based on the underlying protein-protein interactions, pairs of interacting GOSlim terms were formed to represent interacting functional groups. This was done with the help of a Perl script (named Interaction Patterns Analysis or IPA). We calculated an enrichment score for each GOSlim-GOSlim interaction pair originated from all human interacting proteins in i2d database or from hESC-enriched proteins. We based our enrichment score on EASE, a modified Fisher Exact p-value used by DAVID (http://david.abcc.ncifcrf.gov/home.jsp) (26).

Results

Tissue specificity/selectivity. Previous study has generated 3904 tissue-selective unique protein-coding genes (from the latest Affymetrix annotation, express in 5 or less human tissues) for 97 normal human tissue types (24). When comparing 1029 hESC-enriched genes (see Materials and methods) with the tissue-selective ones, 274 genes (Fig. 1) were found intersected. This suggests that some of the hESC-enriched genes are not strictly specific to hESCs, which is not uncommon as described (21).

Protein interaction network for enriched genes. Previous studies have suggested that hESC-enriched proteins (hESPs) may act in a cooperative manner rather than in isolation. In this study, we found evidence to support this. First, out of the 1020 hESPs (see Materials and methods), 784 proteins were found in i2d interaction database, which formed 1765 binary protein interaction pairs (BPIP). With 1000 random samplings

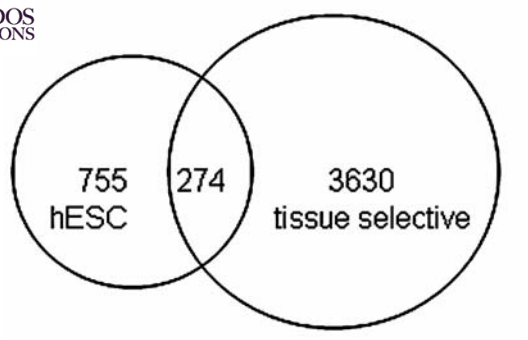

Figure 1. Venn diagrams of the hESC-enriched genes and tissue selective genes.

of the same size (i.e. 784 random proteins) from all non-redundant proteins recorded in i2d database, significantly less BPIP were found (484±50 BPIP, p<2.2e-16, estimated by one-sided Wilcoxon signed rank test with continuity correction). Next, we searched for the most dense sub-network consisting of interacting hESPs. Remarkably, a significant fraction of the interacting hESPs (403 out of 784 proteins, 51.4%) forms a large subnetwork connecting each other. However, 1000 random samplings of 784 proteins from i2d yielded significant smaller subnetworks consisting of 8.5-33.3% of input proteins (as compared to 51.4% for hESPs, p<2.2e-16, one-sided Wilcoxon signed rank test with continuity correction, Table I). Thus, the above results

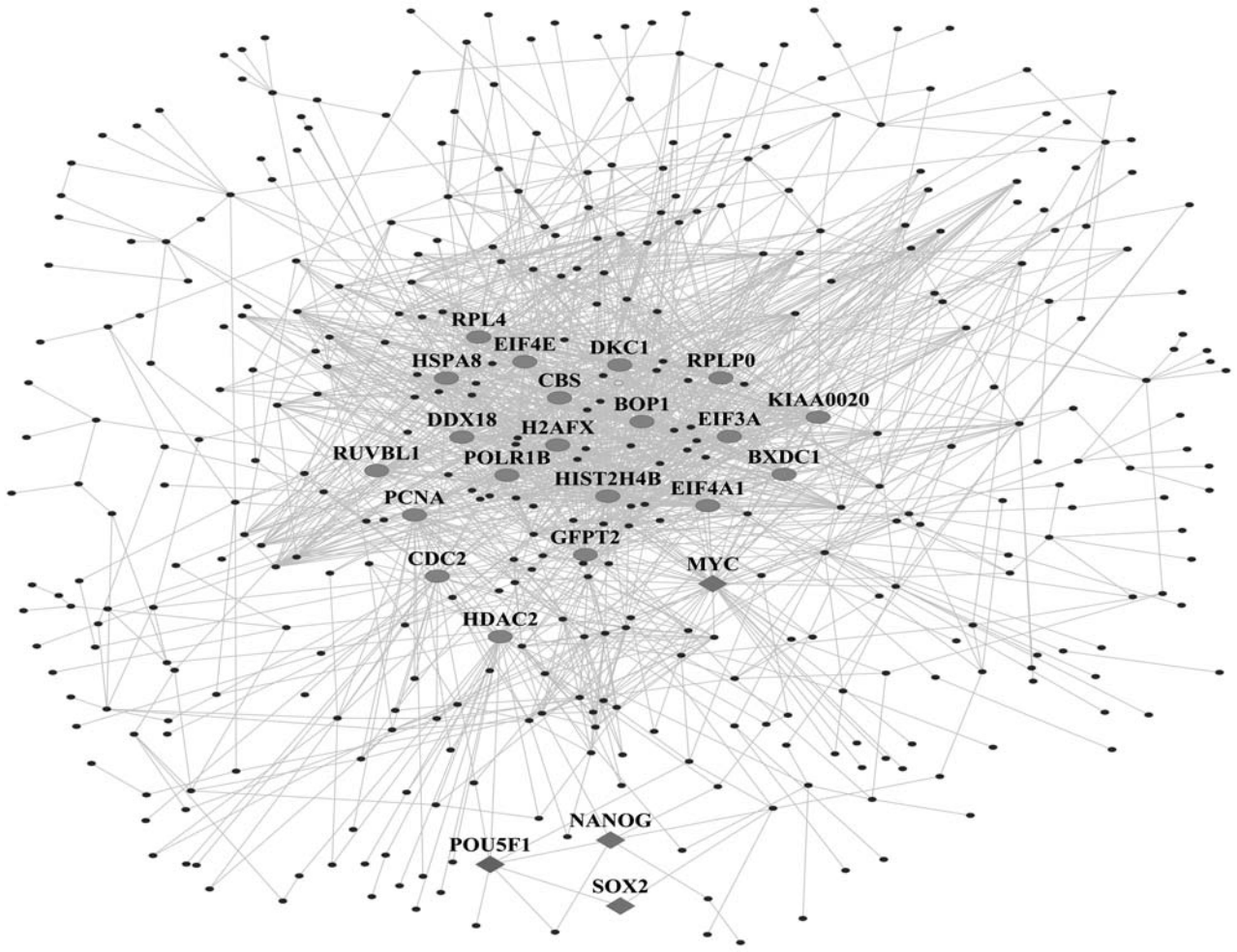

Figure 2. The hESC-enriched protein interaction network. Twenty hubs are labelled and shown as ovals and 4 known important transcription factors in the network are shown as diamonds (MYC is both a hub and a transription factor).

Table I. Significant analysis of BPIP and MNN using hESPs and randomly selected proteins.

	hESCs	Random	P-value[a]
Number of binary protein interaction pairs (BPIP)	1765	484±50	P<2.2e-16
Number of max network nodes (MNN)	403	179±32	P<2.2e-16

[a]One-sided Wilcoxon signed rank test with continuity correction.

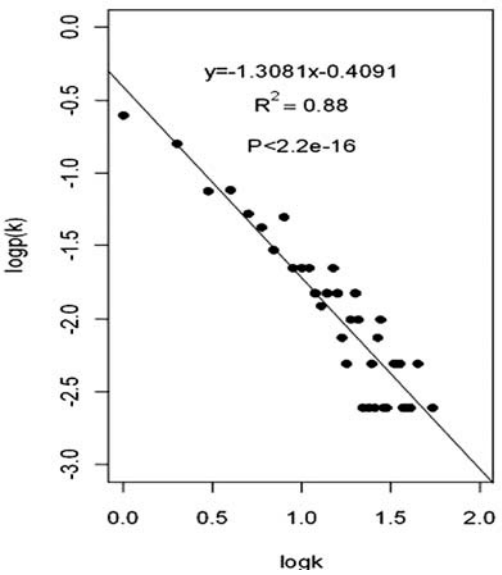

Figure 3. The log plot of P(k) against k illustrating scale-free characteristics (γ=1.3081) of hESC-enriched protein interaction network ($P(k)\sim k^{-\gamma}$). The plot was generated using the free software R.

indicate that a significant portion of the hESPs may cooperate directly or indirectly by means of interaction to form the hESC-enriched protein interaction network (Fig. 2).

Hubs analysis of hESC-enriched protein interaction network. Most biological networks follow power law distribution, $P(K) \sim K^{-\gamma}$ according to the Barabasi-Albert model (27). A network with $\gamma<3$ is a typical scale-free network which possesses a small number of highly connected nodes/hubs in contact with a large number of nodes of low connectivity. The hESC-enriched protein interaction network (γ=1.3081, Fig. 3) in which a few hubs are heavily connected with most nodes of low connectivity, is a scale-free network by definition. Since it is hard to draw a line as to what a hub is, we focused on the top 5% most connected nodes in the network and called them hubs. This accounts for a total of 21 nodes (Table II), each of which having more than 28 partners.

Some of the transcription factors identified previously, e.g. OCT4, SOX2, NANOG and c-MYC, are very important for generating induced Pluripotent Stem (iPS) cells, and therefore are believed to be important regulators for hESCs

Table II. List of 21 hubs in hESC-enriched protein interaction network.

Swiss-Prot Accession	Gene symbol	NANOG/SOX2/ MYC target[a]	Involvement in cancer[b]	Biocarta pathway[c] (part of cell growth)
P62805	HIST1H4C	B, C	-	-
P01106	MYC	-	Y	MAPKinase/p38 MAPK/WNT signaling pathway
P60842	EIF4A1	A	Y	Internal bibosome entry pathway mTOR signaling pathway
Q9NVP1	DDX18	B	Y	-
P16104	H2AFX	B, C	Y	-
Q15397	KIAA0020	-	-	-
P36578	RPL4	-	-	-
P12004	PCNA	B	Y	p53 signaling pathway
Q9H9Y6	POLR1B	-	-	-
P11142	HSPA8	A	Y	-
O60832	DKC1	A	Y	-
P06493	CDC2	B, C	Y	Cyclins and cell cycle regulation
P06730	EIF4E	A	Y	Internal ribosome entry pathway mTOR signaling pathway
Q9H7B2	BXDC1	-	-	-
Q14137	BOP1	-	Y	-
P05388	RPLP0	-	Y	-
Q14152	EIF3A	-	Y	-
Q9Y265	RUVBL1	-	Y	-
Q92769	HDAC2	C	Y	Mechanisms of transcriptional repression by DNA methylation
O94808	GFPT2	-	-	-
P35520	CBS	A	-	-

[a]A, targets of MYC; B, targets of NANOG; C, targets of SOX2. [b]Y, involvement in cancer from current knowledge. [c]http://www.biocarta.com/genes/index.asp.

List of molecular function-GO-GO term interaction pairs enriched in human embryonic stem cells.

GO term1	GO term2	GO term1 function	GO term2 function	EASE score
GO:0003677	GO:0016787	DNA binding	Hydrolase activity	3.3e-46
GO:0000166	GO:0003677	Nucleotide binding	DNA binding	7.9e-40
GO:0003677	GO:0005515	DNA binding	Protein binding	1.5e-36
GO:0005515	GO:0016787	Protein binding	Hydrolase activity	8.4e-26
GO:0003677	GO:0003824	DNA binding	Catalytic activity	2.7e-24
GO:0000166	GO:0000166	Nucleotide binding	Nucleotide binding	4.4e-23
GO:0003723	GO:0005515	RNA binding	Protein binding	4.8e-17
GO:0003824	GO:0005515	Catalytic activity	Protein binding	5.6e-17
GO:0000166	GO:0016787	Nucleotide binding	Hydrolase activity	3.2e-16
GO:0000166	GO:0005515	Nucleotide binding	Protein binding	4.9e-16
GO:0003824	GO:0016787	Catalytic activity	Hydrolase activity	1.2e-14
GO:0005515	GO:0005515	Protein binding	Protein binding	2.1e-14
GO:0003723	GO:0016787	RNA binding	Hydrolase activity	5.0e-14
GO:0003824	GO:0003824	Catalytic activity	Catalytic activity	2.6e-13
GO:0016787	GO:0016787	Hydrolase activity	Hydrolase activity	3.6e-12
GO:0003723	GO:0003723	RNA binding	RNA binding	4.9e-12
GO:0004518	GO:0005515	Nuclease activity	Protein binding	9.9e-09
GO:0003723	GO:0003824	RNA binding	Catalytic activity	2.5e-07
GO:0004518	GO:0004518	Nuclease activity	Nuclease activity	3.4e-07
GO:0005515	GO:0008135	Protein binding	Translation factor activity, nucleic acid binding	2.2e-6
GO:0005488	GO:0016787	Binding	Hydrolase activity	2.4e-6
GO:0004518	GO:0016787	Nuclease activity	Hydrolase activity	3.7e-6
GO:0003677	GO:0005215	DNA binding	Transporter activity	5.9e-6
GO:0003677	GO:0016740	DNA binding	Transferase activity	9.0e-6
GO:0003824	GO:0004518	Catalytic activity	Nuclease activity	1.0e-5
GO:0003723	GO:0009055	RNA binding	Electron carrier activity	1.6e-5
GO:0003676	GO:0003677	Nucleic acid binding	DNA binding	2.1e-5
GO:0016740	GO:0016787	Transferase activity	Hydrolase activity	4.6e-5
GO:0003723	GO:0008135	RNA binding	Translation factor activity, nucleic acid binding	6.9e-5
GO:0003677	GO:0004518	DNA binding	Nuclease activity	6.7e-5
GO:0003824	GO:0008135	Catalytic activity	Translation factor activity, nucleic acid binding	1.4e-4
GO:0000166	GO:0008135	Nucleotide binding	Translation factor activity, nucleic acid binding	1.9e-4
GO:0000166	GO:0004518	Nucleotide binding	Nuclease activity	2.3e-4
GO:0003723	GO:0004518	RNA binding	Nuclease activity	3.0e-4
GO:0003677	GO:0005488	DNA binding	Binding	4.1e-4
GO:0008135	GO:0016787	Translation factor activity, nucleic acid binding	Hydrolase activity	4.8e-4
GO:0003682	GO:0005515	Chromatin binding	Protein binding	4.8e-4
GO:0000166	GO:0003676	Nucleotide binding	Nucleic acid binding	5.0e-4
GO:0003677	GO:0003682	DNA binding	Chromatin binding	6.6e-4
GO:0003824	GO:0016740	Catalytic activity	Transferase activity	7.5e-4
GO:0003682	GO:0016787	Chromatin binding	Hydrolase activity	0.002
GO:0000166	GO:0016740	Nucleotide binding	Transferase activity	0.002
GO:0003676	GO:0016787	Nucleic acid binding	Hydrolase activity	0.002
GO:0003676	GO:0003723	Nucleic acid binding	RNA binding	0.002

Table III. Continued.

GO term1	GO term2	GO term1 function	GO term2 function	EASE score
GO:0003677	GO:0005198	DNA binding	Structural molecule activity	0.005
GO:0030234	GO:0030234	Enzyme regulator activity	Enzyme regulator activity	0.006
GO:0003676	GO:0005515	Nucleic acid binding	Protein binding	0.006
GO:0000166	GO:0003682	Nucleotide binding	Chromatin binding	0.007
GO:0003677	GO:0008135	DNA binding	Translation factor activity, nucleic acid binding	0.007
GO:0008135	GO:0009055	Translation factor activity,nucleic acid binding	Electron carrier activity	0.007
GO:0003723	GO:0005488	RNA binding	Binding	0.007
GO:0003824	GO:0005488	Catalytic activity	Binding	0.008
GO:0003676	GO:0003824	Nucleic acid binding	Catalytic activity	0.010
GO:0005488	GO:0008135	Binding	Translation factor activity, nucleic acid binding	0.011
GO:0003677	GO:0016301	DNA binding	Kinase activity	0.012
GO:0004518	GO:0016301	Nuclease activity	Kinase activity	0.014
GO:0016740	GO:0016740	Transferase activity	Transferase activity	0.015
GO:0005198	GO:0005215	Structural molecule activity	Transporter activity	0.016
GO:0005198	GO:0009055	Structural molecule activity	Electron carrier activity	0.016
GO:0003723	GO:0016740	RNA binding	Transferase activity	0.017
GO:0003824	GO:0004721	Catalytic activity	Phosphoprotein phosphatase activity	0.017
GO:0003700	GO:0016787	Transcription factor activity	Hydrolase activity	0.018
GO:0003723	GO:0004721	RNA binding	Phosphoprotein phosphatase activity	0.020
GO:0003676	GO:0008135	Nucleic acid binding	Translation factor activity, nucleic acid binding	0.023
GO:0003676	GO:0003676	Nucleic acid binding	Nucleic acid binding	0.030
GO:0008135	GO:0008135	Translation factor activity, nucleic acid binding	Translation factor activity, nucleic acid binding	0.030

(6,7,9,28,29). In the hESC-enriched protein interaction network, transcription of many hub protein-coding genes were regulated by these transcription factors (30,31) (Table II). Interestingly, nearly half of the hub proteins are known to be involved in tumorigenesis or associated with poor cancer prognosis (32-45) (Table II). Among the top 10-ranked transcription factor binding sites predicted by both Gather and TFM explorer, those for transcription factor NF-Y were found in common within the proximal promoter sequences of nine hub protein-coding genes.

Enriched functional interactions. Cellular behavior is a consequence of the complex interactions between its numerous constituents. To gain more biological insights by studying functional interactions, we first annotated each protein in the hESC-enriched protein protein interaction pairs with a subset of Gene Ontology terms. In this study, 42 GOSlim terms from the molecular function category of the Gene Ontology were used. Our goal was to identify significantly enriched functionally interacting GO-GO pairs. This resulted in 66 GO-GO interaction pairs (Table III) enriched in hESCs. Except for GO:0030234, most of them were found in a functionally interacting network (Fig. 4). The top four most connected GOSlim terms in this network are: GO:0003677, GO:0016787, GO:0003723 and GO:0003824.

Discussion

We studied the interactions of a list of genes over-expressed in hESCs. A protein-protein interaction subnetwork formed

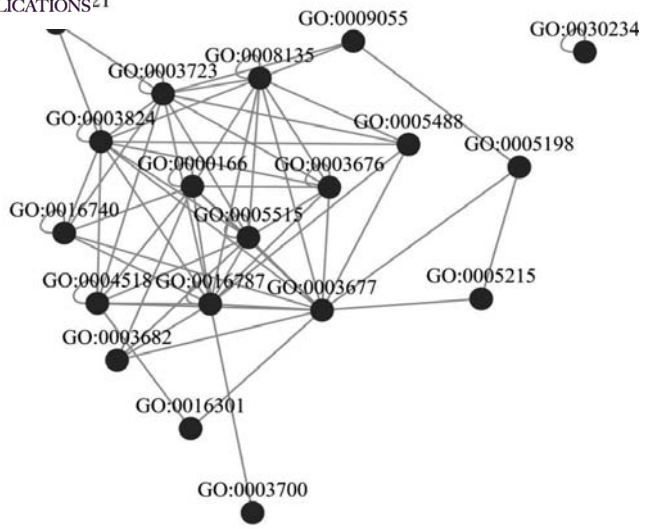

Figure 4. Enriched 'GOSlim-GOSlim interaction' network in hESC (molecular function interaction patterns).

with a significant number of hESC-enriched genes was identified by integrating gene expression and protein interaction data. It is full of hESC-enriched genes and is very likely to be responsible for maintaining characteristics of hESCs. This scale-free network has a few dominant hubs heavily connected with most nodes of low connectivity. A series of studies have shown that hub proteins in a scale-free network are likely to be essential for growth and the degree of connectivity correlates with other phenotypes in addition to essentiality (46-48). Scale-free networks, albeit more tolerant to random removal of nodes, are vulnerable to loss of highly interactive hubs (49).

MYC, one of the hubs in our hESC-enriched protein interaction network, is a key factor for inducing Pluripotent Stem (iPS) cells (29), and for regulating self-renewal and pluripotency in mouse ES cells (mESCs) (50). A recent study showed that H2AFX/H2AX, also a hub in the hESC-enriched protein interaction network, participates in a critical signaling pathway different from that of somatic cells to control stem cell proliferation (51). RUVBL1/TIP49a, an important hub identified in this study and a common hub in co-expression networks found in both hESCs and mESCs (52), plays a critical role in c-MYC and WNT signaling pathways (32,53). RUVBL1/TIP49a was also found to be evolutionarily highly conserved and essential for viability in yeast, flies and worms (54). Furthermore, WNT, mTOR and MAPK pathways are pivotal for regulating hESCs, as evidenced by the involvement of several hub proteins in these pathways (55-57). We anticipate that some other hubs or nodes identified in this study may turn out to be important factors in future investigations of hESCs.

To our surprise, several important transcription factors, namely NANOG, SOX2, OCT4, are not among the most connected hubs in this network whereas some of their targets are (Table II). These factors also share some common targets (30), which suggests that they work closely in a regulatory network. In addition, OCT4 interacts with NANOG and SOX2 in the hESC-enriched protein interaction network.

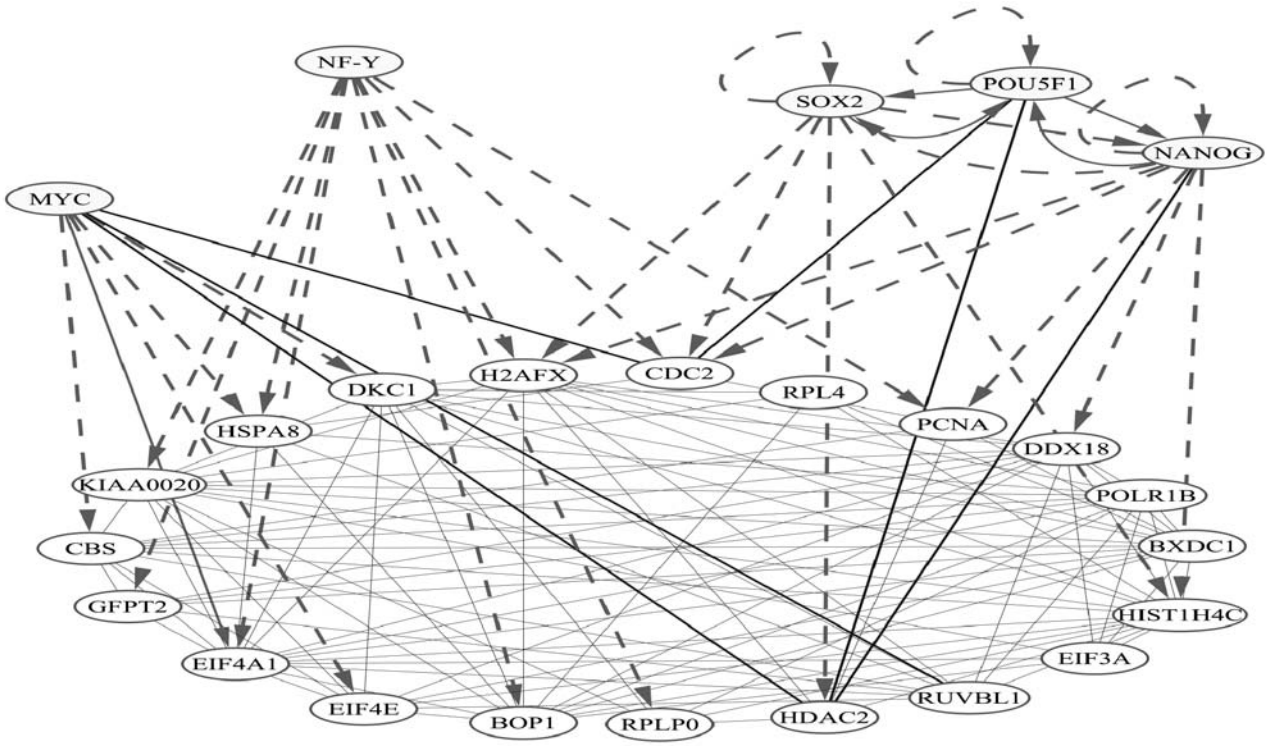

Figure 5. The relationship between hubs and important transcription factors in hESC-enriched protein interaction network: solid line, protein-protein interaction; solid line with arrows, protein-protein interactions with regulatory potential and direction (as indicated by arrows) and dash line with arrows, with regulatory potential and direction only. Important transcription factors (SOX2, POU5F1, NANOG, MYC and NF-Y), their targets (ovals) and regulatory circuits (arrows) are indicated.

Regulatory circuits formed by some of these factors and their targets, as reported previously based on experimental findings (20), may be used to fine-tune characteristics of the ES. Furthermore, the directed or self-directed regulatory loops of OCT4, NANOG and SOX2 can be used to maintain proper and stable expression level (30) through a robust synergistic and sustained network topology (through regulation of many hubs) whose stabilization is vitally important for hESC (Fig. 5). Any imbalance of a regulator with its targets or with itself in the network may produce unexpected outcomes. This also suggests that synergistic hubs target activation by their regulators is a very important mechanism by which these tran-scription factors, albeit not hubs, control self-renewal and pluripotency in hESCs.

In addition to the above important transcription factors, NF-Y was predicted by *in silico* analysis of promoter sequences to regulate close to half of the 21 hubs in the hESC-enriched protein interaction network. It has recently been shown that NF-Y binding site is rather conserved and over-represented in promoter regions of genes preferentially expressed in human and mouse pluripotent cells (58). In fact, NF-Y was down-regulated during differentiation (58). Taken all these observations together, NF-Y seems to be an important sustaining factor for the maintenance of hESCs by regulating its hub targets in the hESC-enriched protein interaction network (Fig. 5). In line with this, NF-Y also influences hematopoietic stem cell (HSC) self-renewal and differentiation (59).

Functional categories such as gene ontology terms (GO term) for genes enriched under certain condition(s) can facilitate functional interpretation and derive biologically meaningful conclusions. Previous studies mainly focused on the enrichment of GO terms for selected genes/proteins, largely ignored the interactions across various functional groups. Therefore, it makes more sense to identify biological meanings by investigating enriched functional interactions with protein interaction considered. To this end, we found 66 functional interaction pairs enriched in hESCs, which formed an interaction network. Interestingly, most of these interacting pairs, particularly those formed with GO:0003677 (DNA binding), GO:0003723 (RNA binding), are involved in transcription and translation. This is consistent with the need for self-renewal and unrestrained proliferation.

Human embryonic stem cell-derived therapy renews our hope for regenerative medicine, but we must first overcome several hurdles, one of which, perhaps the biggest, is that hESC-therapies may spur tumor formation (60,61). There are some apparent parallel traits such as self-renewal and differentiation capacity between stem cells and cancer cells, which prompt us to the hypothesis that tumors often arise from undifferentiated stem/progenitor cells, or cancer stem cells are derived from normal stem cells (62,63). Out of the 21 hubs in the hESC-enriched protein interaction network, 14 (67%) are involved in tumorigenesis or poor cancer prognosis. In addition, a recent study showed that an embryonic stem cell-like gene expression signature (part of what we used here), potentially contributing to stem cell-like phenotypes shown by many tumors, has been found in poorly differentiated aggressive human tumors (64). These observations can expand our knowledge in hESC biology and tumor formation.

In conclusion, this study has identified an enriched protein interaction network formed by 403 hESC-over-expressed gene-coding proteins. Enriched molecular functional interaction network were also found in hESCs. The existence of these interaction networks beyond randomness suggests that they are important and very likely to be responsible for the maintenance of hESCs. The hubs governing the hESC-enriched protein interaction network, such as MYC, H2AFX, RUVBL1, DDX18, CDC2, HDAC2, HIST1H4C and so on, which are possibly critical in determining the fate of hESCs, deserve more attention in future investigations. It is worth noting that some hESC-associated proteins for self-renewal and pluripotency, for instance, KLF4 (14) and possibly other factors, are missing in the hESC-enriched protein interaction network. Despite this, our findings represent a step in the right direction, on the system levels, to gain more significant biological information in stem cell research. When more data become available, it is possible to refine the network and make it more informative for studying hESCs. It is also hoped that continuous research on hESCs will speed up its therapeutic applications.

Acknowledgements

This study was supported by a grant for Key lab of Biochip Research of Guangdong Province, China (NO. 2004B60144).

References

1. Thomson JA, Itskovitz-Eldor J, Shapiro SS, *et al*: Embryonic stem cell lines derived from human blastocysts. Science 282: 1145-1147, 1998.
2. Newman MB and Bakay RA: Therapeutic potentials of human embryonic stem cells in Parkinson's disease. Neurotherapeutics 5: 237-251, 2008.
3. Docherty K, Bernardo AS and Vallier L: Embryonic stem cell therapy for diabetes mellitus. Semin Cell Dev Biol 18: 827-838, 2007.
4. Siu CW, Moore JC and Li RA: Human embryonic stem cell-derived cardiomyocytes for heart therapies. Cardiovasc Hematol Disord Drug Targets 7: 145-152, 2007.
5. Kang HK, Roh S, Lee G, Hong SD, Kang H and Min BM: Osteogenic potential of embryonic stem cells in tooth sockets. Int J Mol Med 21: 539-544, 2008.
6. Nichols J, Zevnik B, Anastassiadis K, *et al*: Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95: 379-391, 1998.
7. Chambers I, Colby D, Robertson M, *et al*: Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113: 643-655, 2003.
8. Rogers MB, Hosler BA and Gudas LJ: Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. Development 113: 815-824, 1991.
9. Graham V, Khudyakov J, Ellis P and Pevny L: SOX2 functions to maintain neural progenitor identity. Neuron 39: 749-765, 2003.
10. Hanna LA, Foreman RK, Tarasenko IA, Kessler DS and Labosky PA: Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. Genes Dev 16: 2650-2661, 2002.
11. Wuchty S, Barabasi AL and Ferdig MT: Stable evolutionary signal in a yeast protein interaction network. BMC Evol Biol 6: 8, 2006.
12. Brown KR and Jurisica I: Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biol 8: R95, 2007.
13. Kim J, Chu J, Shen X, Wang J and Orkin SH: An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132: 1049-1061, 2008.
14. Jiang J, Chan YS, Loh YH, *et al*: A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol 10: 353-360, 2008.

SPANDIDOS PUBLICATIONS
CH, Wu Q, Chew JL, *et al*: The Oct4 and Nanog iption network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38: 431-440, 2006.
16. Zhou Q, Chipperfield H, Melton DA and Wong WH: A gene regulatory network in mouse embryonic stem cells. Proc Natl Acad Sci USA 104: 16438-16443, 2007.
17. Chuang HY, Lee E, Liu YT, Lee D and Ideker T: Network-based classification of breast cancer metastasis. Mol Syst Biol 3: 140, 2007.
18. Hahn MW and Kern AD: Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 22: 803-806, 2005.
19. Ideker T and Sharan R: Protein networks in disease. Genome Res 18: 644-652, 2008.
20. Wang J, Rao S, Chu J, *et al*: A protein interaction network for pluripotency of embryonic stem cells. Nature 444: 364-368, 2006.
21. Assou S, Le Carrour T, Tondeur S, *et al*: A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells 25: 961-973, 2007.
22. Brown KR and Jurisica I: Online predicted human interaction database. Bioinformatics 21: 2076-2082, 2005.
23. Yu X, Lin J, Zack DJ and Qian J: Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. Nucleic Acids Res 34: 4925-4936, 2006.
24. Liang S, Li Y, Be X, Howes S and Liu W: Detecting and profiling tissue-selective genes. Physiol Genomics 26: 158-162, 2006.
25. Shannon P, Markiel A, Ozier O, *et al*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504, 2003.
26. Hosack DA, Dennis G Jr, Sherman BT, Lane HC and Lempicki RA: Identifying biological themes within lists of genes with EASE. Genome Biol 4: R70, 2003.
27. Barabasi AL and Oltvai ZN: Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101-113, 2004.
28. Yu J, Vodyanik MA, Smuga-Otto K, *et al*: Induced pluripotent stem cell lines derived from human somatic cells. Science 318: 1917-1920, 2007.
29. Takahashi K and Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663-676, 2006.
30. Boyer LA, Lee TI, Cole MF, *et al*: Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122: 947-956, 2005.
31. Fernandez PC, Frank SR, Wang L, *et al*: Genomic targets of the human c-Myc protein. Genes Dev 17: 1115-1129, 2003.
32. Feng Y, Lee N and Fearon ER: TIP49 regulates beta-catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. Cancer Res 63: 8726-8734, 2003.
33. Prochownik EV: c-Myc: linking transformation and genomic instability. Curr Mol Med 8: 446-458, 2008.
34. Ji P, Diederichs S, Wang W, *et al*: MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22: 8031-8041, 2003.
35. Dubaele S and Chene P: Cellular studies of MrDb (DDX18). Oncol Res 16: 549-556, 2007.
36. Bassing CH, Suh H, Ferguson DO, *et al*: Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. Cell 114: 359-370, 2003.
37. Malkas LH, Herbert BS, Abdel-Aziz W, *et al*: A cancer-associated PCNA expressed in breast cancer has implications as a potential biomarker. Proc Natl Acad Sci USA 103: 19472-19477, 2006.
38. Powers MV, Clarke PA and Workman P: Dual targeting of HSC70 and HSP72 inhibits HSP90 function and induces tumor-specific apoptosis. Cancer Cell 14: 250-262, 2008.
39. Montanaro L, Brigotti M, Clohessy J, *et al*: Dyskerin expression influences the level of ribosomal RNA pseudo-uridylation and telomerase RNA component in human breast cancer. J Pathol 210: 10-18, 2006.
40. Chen H, Huang Q, Dong J, Zhai DZ, Wang AD and Lan Q: Overexpression of CDC2/CyclinB1 in gliomas, and CDC2 depletion inhibits proliferation of human glioma cells in vitro and in vivo. BMC Cancer 8: 29, 2008.
41. Graff JR, Konicek BW, Carter JH and Marcusson EG: Targeting the eukaryotic translation initiation factor 4E for cancer therapy. Cancer Res 68: 631-634, 2008.
42. Killian A, Sarafan-Vasseur N, Sesboue R, *et al*: Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis. Genes Chromosomes Cancer 45: 874-881, 2006.
43. Chang TW, Chen CC, Chen KY, Su JH, Chang JH and Chang MC: Ribosomal phosphoprotein P0 interacts with GCIP and over-expression of P0 is associated with cellular proliferation in breast and liver carcinoma cells. Oncogene 27: 332-338, 2008.
44. Zhang L, Pan X and Hershey JW: Individual overexpression of five subunits of human translation initiation factor eIF3 promotes malignant transformation of immortal fibroblast cells. J Biol Chem 282: 5790-5800, 2007.
45. Weichert W, Roske A, Gekeler V, *et al*: Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. Br J Cancer 98: 604-610, 2008.
46. Jeong H, Mason SP, Barabasi AL and Oltvai ZN: Lethality and centrality in protein networks. Nature 411: 41-42, 2001.
47. Said MR, Begley TJ, Oppenheim AV, Lauffenburger DA and Samson LD: Global network analysis of phenotypic effects: protein networks and toxicity modulation in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 101: 18006-18011, 2004.
48. Shachar R, Ungar L, Kupiec M, Ruppin E and Sharan R: A systems-level approach to mapping the telomere length maintenance gene circuitry. Mol Syst Biol 4: 172, 2008.
49. Albert R: Scale-free networks in cell biology. J Cell Sci 118: 4947-4957, 2005.
50. Cartwright P, McLean C, Sheppard A, Rivett D, Jones K and Dalton S: LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development 132: 885-896, 2005.
51. Andang M, Hjerling-Leffler J, Moliner A, *et al*: Histone H2AX-dependent GABA(A) receptor regulation of stem cell proliferation. Nature 451: 460-464, 2008.
52. Sun Y, Li H, Liu Y, Mattson MP, Rao MS and Zhan M: Evolutionarily conserved transcriptional co-expression guiding embryonic stem cell differentiation. PLoS ONE 3: E3406, 2008.
53. Wood MA, McMahon SB and Cole MD: An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. Mol Cell 5: 321-330, 2000.
54. Qiu XB, Lin YL, Thome KC, *et al*: An eukaryotic RuvB-like protein (RUVBL1) essential for growth. J Biol Chem 273: 27786-27793, 1998.
55. Armstrong L, Hughes O, Yung S, *et al*: The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. Hum Mol Genet 15: 1894-1913, 2006.
56. Murakami M, Ichisaka T, Maeda M, *et al*: mTOR is essential for growth and proliferation in early mouse embryos and embryonic stem cells. Mol Cell Biol 24: 6710-6718, 2004.
57. Miyabayashi T, Teo JL, Yamamoto M, McMillan M, Nguyen C and Kahn M: Wnt/beta-catenin/CBP signaling maintains long-term murine embryonic stem cell pluripotency. Proc Natl Acad Sci USA 104: 5668-5673, 2007.
58. Grskovic M, Chaivorapol C, Gaspar-Maia A, Li H and Ramalho-Santos M: Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet 3: E145, 2007.
59. Zhu J, Zhang Y, Joe GJ, Pompetti R and Emerson SG: NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. Proc Natl Acad Sci USA 102: 11728-11733, 2005.
60. Hentze H, Graichen R and Colman A: Cell therapy and the safety of embryonic stem cell-derived grafts. Trends Biotechnol 25: 24-32, 2007.
61. Vogel G: Cell biology. Ready or not? Human ES cells head toward the clinic. Science 308: 1534-1538, 2005.
62. Lobo NA, Shimono Y, Qian D and Clarke MF: The biology of cancer stem cells. Annu Rev Cell Dev Biol 23: 675-699, 2007.
63. Reya T, Morrison SJ, Clarke MF and Weissman IL: Stem cells, cancer, and cancer stem cells. Nature 414: 105-111, 2001.
64. Ben-Porath I, Thomson MW, Carey VJ, *et al*: An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 40: 499-507, 2008.