Network-based association of hypoxia-responsive genes with cardiovascular diseases

Rui-Sheng Wang, William M Oldham and Joseph Loscalzo
Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA
E-mail: jloscalzo@partners.org

Received 19 May 2014, revised 21 July 2014
Accepted for publication 11 August 2014
Published 24 October 2014

Abstract
Molecular oxygen is indispensable for cellular viability and function. Hypoxia is a stress condition in which oxygen demand exceeds supply. Low cellular oxygen content induces a number of molecular changes to activate regulatory pathways responsible for increasing the oxygen supply and optimizing cellular metabolism under limited oxygen conditions. Hypoxia plays critical roles in the pathobiology of many diseases, such as cancer, heart failure, myocardial ischemia, stroke, and chronic lung diseases. Although the complicated associations between hypoxia and cardiovascular (and cerebrovascular) diseases (CVD) have been recognized for some time, there are few studies that investigate their biological link from a systems biology perspective. In this study, we integrate hypoxia genes, CVD genes, and the human protein interactome in order to explore the relationship between hypoxia and cardiovascular diseases at a systems level. We show that hypoxia genes are much closer to CVD genes in the human protein interactome than that expected by chance. We also find that hypoxia genes play significant bridging roles in connecting different cardiovascular diseases. We construct a hypoxia-CVD bipartite network and find several interesting hypoxia-CVD modules with significant gene ontology similarity. Finally, we show that hypoxia genes tend to have more CVD interactors in the human interactome than in random networks of matching topology. Based on these observations, we can predict novel genes that may be associated with CVD. This network-based association study gives us a broad view of the relationships between hypoxia and...
cardiovascular diseases and provides new insights into the role of hypoxia in cardiovascular biology.

Online supplementary data available from stacks.iop.org/NJP/16/105014/mmedia

Keywords: hypoxia, cardiovascular disease, human interactome

Introduction

Molecular oxygen is essential for living beings. All multicellular organisms depend on a constant supply of oxygen and maintain oxygen homeostasis for survival. Oxygen is the primary electron acceptor in many cellular reactions and is harnessed by mitochondria to generate ATP via oxidative phosphorylation. An excess or deficiency of oxygen may result in the death of cells, tissues, or organisms. Lower oxygen levels result in hypoxic stress (hypoxia) causing cells to activate regulatory pathways responsible for increasing the oxygen supply and optimizing metabolism under limited oxygen tension. A critical response to hypoxia is mediated through changes in gene expression regulated by the oxygen sensing pathway involving the hypoxia-inducible factors (HIFs) to promote adaptation to low oxygen content [1]. One way in which HIFs protect cells from hypoxic stress is by inducing a shift from oxidative to glycolytic metabolism converting glucose to pyruvate that can be further converted to lactate as a glycolytic end product. In hypoxic cells, the expression of several hundred mRNAs is changed through regulation by HIF, which binds to hypoxia responsive elements in their gene promoter regions [2]. There is also a growing number of HIF-independent pathways that have been found to promote optimal oxygen utilization and hypoxia tolerance [3].

Hypoxia is commonly associated with the pathobiology of many diseases, e.g., cardiovascular diseases, such as myocardial ischemia, myocardial infarction, and stroke, as well as chronic lung diseases, cancer, and inflammation [1]. These diseases are either caused by or a consequence of hypoxia. Cardiovascular disease, the leading cause of death worldwide, is a class of diseases that affects the cardiovascular system (heart, blood vessels, or both). Oxygen is not only required for normal oxidative metabolism but is also a critical participant in the generation of many small-molecule signaling intermediates, such as nitric oxide (NO) and reactive oxygen species, and in many other cellular redox processes. These oxygen-associated reactions can be either beneficial or contribute to cardiac dysfunction and death [4]. In addition, oxygen is a major determinant of myocardial gene expression. Under hypoxic conditions, gene expression patterns in the heart are significantly altered [4, 5]. Previous studies have also shown that the oxygen sensitive master transcription factors, the HIFs, play a protective role in the pathophysiology of myocardial ischemia and pressure-overload heart failure, and contribute to the pathogenesis of pulmonary arterial hypertension [2, 6]. Therefore, consideration of the relationships between hypoxia and cardiovascular diseases is essential in understanding the pathobiology of these common disorders.

Although the relationship between hypoxia and cardiovascular diseases has been recognized for many years, few studies have investigated this relationship from a systems-level perspective. The growth of high-throughput data provides new avenues by which to study complex human diseases, including cardiovascular diseases, using systems biology approaches
[7–13]. Dewey and colleagues assembled myocardial transcript data and used gene coexpression network analysis to derive functional modules and regulatory mediators in developing myocardium that were not present in normal adult tissue [14]. Jensen and coworkers integrated protein interaction data and genome-wide association study (GWAS) data in an analysis of coronary artery disease (CAD) patients and found a novel gene module that was overrepresented in CAD-associated genes [15]. Huan and colleagues used a systems biology framework that integrated gene coexpression networks, tissue-specific Bayesian networks, and protein–protein interaction networks to identify key regulatory driver genes for coronary heart disease [16]. In addition, metabolomics data have been used to monitor the level of hundreds to thousands of metabolites in normal or diseased tissues, allowing the identification of metabolic biomarkers of cardiovascular diseases [17]. These examples clearly demonstrate that network modeling facilitates systems-level analysis of cardiovascular diseases. In addition, diseases with similar pathophenotypes tend to share genes or cluster in the human interactome [18]. Given the fact that hypoxia is involved in the pathobiology of cardiovascular diseases, we hypothesize that hypoxia-induced signaling pathways may be intimately linked to the molecular mechanisms of CVDs. In light of the successes of systems biology approaches in studying cardiovascular diseases and other diseases that are closely linked to hypoxic stress, we chose to explore the proximity of the molecular determinants of CVD and hypoxia responses at a biological network level.

In this study, we investigate the relationships between hypoxia genes and CVD by integrating hypoxia genes and CVD genes through the comprehensive human interactome. For hypoxia genes, we focus on those that are differentially expressed in response to hypoxic stress in endothelial cells, smooth muscle cells, and fibroblasts. The CVD genes are taken from public databases in which curated literature information about disease genes is deposited. From this systems-level study, we show that hypoxia genes are much closer to CVD genes in the human protein interactome than that expected by chance. We also find that hypoxia genes play significant bridging roles in connecting different cardiovascular diseases. By mapping genes common to hypoxia and CVD to knowledge-based signaling pathways, we identify several unique signaling pathways common to hypoxia and CVD. We construct a hypoxia-CVD bipartite network and find several interesting hypoxia-CVD modules with significant gene ontology (GO) similarity. Finally, we show that hypoxia genes tend to have more CVD interactors in the human interactome than in random networks of matching topology. Based on these observations, we attempt to predict new genes that may be associated with CVD.

**Materials and methods**

1. **Collecting hypoxia-responsive genes**

We collected hypoxia-responsive genes from two genome-wide gene expression microarray studies. One study is from our own laboratory and comprises a microarray dataset from three cell types exposed to hypoxia (William M Oldham, unpublished data): pulmonary arterial endothelial cells (PAEC), pulmonary arterial smooth muscle cells (PASMC), and lung fibroblasts (LF). We used a bioconductor package in R, *Limma* [19], to identify hypoxia-responsive genes. After applying 0.05 as the Benjamini–Hochberg-adjusted *P*-value threshold and mapping the probe sets to genes, we obtained 376, 484, and 25 hypoxia-responsive genes in PAEC, LF, and PASMC, respectively. The other microarray dataset is taken from [20] in which...
primary human PAECs were subjected to 4 different culture conditions: first, exposed to 20% O₂; second exposed to 1% O₂; and third and fourth, exposed under non-hypoxic conditions to 50 pfu of adenovirus encoding the constitutively expressed form of HIF-1 (AdCA5) or *Escherichia coli* beta-galactosidase (AdLacZ), respectively. Those probes for which a significant difference (Student’s t-test, *P*<0.05, and 1.5-fold change) was observed with similar variation in response to both hypoxia and AdCA5 were viewed as hypoxia-responsive genes [20]. These constraints led to the identification of 245 genes with increased expression and 325 genes with decreased expression in response to hypoxic stress. Combining the data from the two studies gives a total of 1210 hypoxia-responsive genes (supplemental file 1).

2. Compiling cardiovascular disease genes

The cardiovascular disease genes were compiled from two databases: the Disease Ontology (DO) database (http://disease-ontology.org/) [21] and the Human Genome Epidemiology (HuGE) Navigator [22]. DO has been developed as a standardized ontology for human disease. It not only integrates disease and medical vocabularies through extensive cross-mapping of DO terms to other disease databases, including OMIM, but also provides gene-disease association relationships through FunDO [23]. We selected a few principal (common) cardiovascular diseases (DO terms) with a significant number of associated genes according to FunDO, such as atherosclerosis, pulmonary hypertension, cerebrovascular disorders, myocardial ischemia, stroke, brain ischemia, intermediate coronary syndrome and thrombophlebitis. HuGE Navigator is an integrated, searchable knowledge base of genetic associations and HuGE [22]. Phenopedia (HuGEpedia) in this database provides a disease-centered view of genetic association studies summarized in the online HuGE encyclopedia. We followed the disease classification of DO and selected several diseases identical or close to the DO terms we used, such as atherosclerosis, brain ischemia, cerebrovascular accident, cerebrovascular disorders, coronary artery disease, pulmonary hypertension, myocardial ischemia, hypoxia-ischemia, thrombophlebitis and thrombosis. To obtain reliable disease genes, we only retained those genes with more than one publication from different research labs in support of an association.

Combining the two resources provides a total of 938 CVD genes associated with six diseases (supplemental file 1): atherosclerosis, coronary artery disease (CAD) (including intermediate coronary syndrome), pulmonary hypertension (PH), cerebrovascular disease (CeVD) (including stroke, cerebrovascular disorder, and cerebrovascular accident), myocardial ischemia (MIS) (including brain ischemia and hypoxia-ischemia), and thrombosis (including thrombophlebitis). The distribution of CVD genes over all diseases studied is shown in figure 1. As shown in figure 1(A), each disease has some genes specific to this disease, and different cardiovascular diseases have a large number of overlapping genes, which is not surprising as they are often mechanistically linked (such as atherosclerosis and CAD). Figure 1(B) plots the number of genes against the number of associated diseases, with many genes involved in more than one disease.

3. Construction of a human protein interactome

Biological processes involve different types of molecular interactions. To build a comprehensive interactome, we pooled human molecular interaction data from different sources, including protein–protein interactions, protein complexes, protein–DNA interactions, kinase–substrate interactions, metabolic interactions, and signaling pathways. Protein–protein
interactions from several high-throughput yeast-two-hybrid studies \[24–27\] were combined with the CCSB Human Interactome (HI-2012, \url{http://interactome.dfci.harvard.edu/H_sapiens/}) and the binary subsets of interactions reported in the IntAct, MINT, HPRD, and BioGrid databases \[28–31\]. A protein complex is a group of two or more associated polypeptide chains linked by non-covalent protein–protein interactions. CORUM is a database collection of manually curated human protein complexes derived from a variety of experimental tools \[32\]. CORUM, literature curated protein interactions (LCI) from the CCSB, and experimentally determined human protein complexes \[33\] were also included in the set of protein–protein interactions. Protein–DNA regulatory interactions were taken from the TRANSFAC database, which lists experimentally derived regulatory interactions \[34\]. Kinase–substrate interactions were obtained from the PhosphositePlus database \[35\], which provides comprehensive information on post-translational protein modifications, including phosphorylation, ubiquitination, acetylation, and methylation. Metabolic enzyme-coupled interactions (two enzymes sharing adjacent reactions) were derived from the KEGG and BiGG databases \[36–38\]. In addition, protein interactions from 3D structural prediction \[39\] and signaling interactions \[40\] were also combined in the network.

The consolidated human interactome (July 2013 version) described above has 14 174 proteins with 170 303 interactions, after removing duplicate interactions and self-loops. The edges in the consolidated human interactome represent functional associations among genes at different levels rather than just physical protein-protein binding. Among the hypoxia genes and CVD genes compiled in the analysis described above, 850 hypoxia-responsive genes and 864 CVD genes can be mapped in this interactome.

4. Null models for assessing emergent properties

We adopted two null models to assess the emergent properties of hypoxia genes and CVD genes in the context of the human interactome. The first null model (Null Model I) builds the null distribution of hypoxia genes and CVD genes at random. To put it another way, Null Model I keeps the human protein interactome network unchanged and randomly selects the
same numbers of genes from the network as hypoxia genes and CVD genes. Then the
topological properties of random gene sets are compared with those of real hypoxia genes and
CVD genes in the original human interactome network. We also used a second null model (Null
Model II) when applicable. Null Model II keeps hypoxia and CVD gene sets unchanged, and
randomly shuffles the human interactome network while keeping its size and degree
distribution. The topological properties of hypoxia and CVD genes are compared in the
human interactome versus in random networks of matching topology. We generated 1 000
random gene sets (Null Model I) or 1 000 random networks (Null Model II) to build the null
distributions as controls when we assessed the significance of emergent properties of the
experimental distributions.

5. Network analysis and implementation

Most of the network analyses in this study were performed using Python, with the assistance of
NetworkX (http://networkx.github.io/)—a Python library software package for the creation,
manipulation, and study of the structure, dynamics, and functions of complex networks [41]. It
contains many built-in graph and network analysis algorithms, such as shortest path algorithms,
connected component algorithms, subgraph induction, random graph generators, etc. We can
readily use them for examining the topological relationships between hypoxia genes and CVD
genes at a network level.

Closeness is one of the centrality measures of a node in a network, and is defined as the
reciprocal of the sum of the shortest path distances from this node to all other nodes [42].
Closeness centrality (CC) measures how far this node is from all other nodes, i.e., how long it
takes to spread information from a given node to all other nodes (assuming identical
transmission rates). In our case, we are interested in examining how close hypoxia genes are to
CVD genes, compared to all proteins, in the human interactome. CC of a hypoxia gene \( h_i \)
to CVD genes is based on the average length of the shortest paths between \( h_i \) and all CVD genes it
can reach in the human interactome:

\[
C_C(h_i) = \frac{|\text{CVD}|}{\sum_{u \in \text{CVD}} d(h_i, u)},
\]

where \( d(h_i, u) \) is the shortest-path distance between \( h_i \) and \( u \), and \(|\text{CVD}|\) is the number of CVD
genes that \( h_i \) can reach in the interactome.

Betweenness is another centrality measure of a node in a network. It is defined as the sum
of the fraction of all node pairs’ shortest paths that pass through this node [43]. Betweenness
centrality (BC) quantifies the control of a node on the communication between other nodes. In
our case, we aim to evaluate the role of hypoxia genes in the transition of two cardiovascular
diseases. BC of a hypoxia gene \( h_i \) in connecting two cardiovascular diseases (\( D_1 \) and \( D_2 \)) is the
sum of the fraction of all gene pairs’ shortest paths between \( D_1 \) and \( D_2 \) that pass through \( h_i \):

\[
C_B(h_i) = \sum_{s \in D_1, t \in D_2} \frac{\sigma(s, t|h_i)}{\sigma(s, t)},
\]

where \( \sigma(s, t) \) is the total number of shortest paths from gene \( s \) associated with disease \( D_1 \) to gene
\( t \) associated with disease \( D_2 \), and \( \sigma(s, t|h_i) \) is the number of those shortest paths from \( s \) to \( t \) that
pass through hypoxia gene \( h_i \). This BC measure scales with the number of gene pairs as implied
by the summation indices; alternatively, it can be normalized by dividing by the number of gene pairs:

$$\hat{C}_B(h_i) = \frac{1}{|D_{1,i}| \cdot |D_{2,i}|} \sum_{s \in D_{1,i}, t \in D_{2}} \frac{\sigma(s, \{h_i\})}{\sigma(s, t)},$$

where $|D_{1,i}|$ is the number of genes associated with disease $D_1$ that $h_i$ can reach in the interactome, and $|D_{2,i}|$ is the number of genes associated with disease $D_2$ that $h_i$ can reach in the interactome.

To find modules densely connecting hypoxia genes and CVD genes, we constructed a hypoxia-CVD bipartite network such that the connectivity within hypoxia genes or within CVD genes does not affect module detection. The hypoxia-CVD bipartite network only contains interactions between hypoxia genes and CVD genes. We used the Louvain method to maximize the modularity function defined for bipartite networks and, thereby, identify hypoxia-CVD gene modules [44–46]:

$$Q = \frac{1}{2m} \sum_{i,j} \left( A_{i,j} - \frac{k_i d_j}{m} \right) \delta(g_i, g_j),$$

where $m$ is the total number of edges in the bipartite network; the $A_{i,j}$ are the adjacency matrix elements representing the weight of the edge between hypoxia gene $h_i$ and CVD gene $c_j$; $k_i$ and $d_j$ are the degrees of $h_i$ and node $c_j$, respectively; $g_i$ and $g_j$ are the module indices of $h_i$ and $c_j$, respectively; and $\delta$ is the delta function and is equal to 1 if $g_i = g_j$ and equal to 0 otherwise.

The hypergeometric distribution of a random variable $X$ has the following probability mass function:

$$P(X = k | N, K, n) = \binom{K}{k} \binom{N - K}{n - k} \binom{N}{n},$$

where $N$ is the population size, $K$ is the number of success states in the population, $n$ is the number of samplings without replacement, and $k$ is the number of successes in the sampled set. The cumulative probability mass function

$$P(X \geq k | N, K, n) = 1 - \sum_{i=0}^{k-1} P(X = i | N, K, n)$$

represents the probability that the random variable $X$ takes a value equal to or larger than $k$. We use this hypergeometric distribution-based statistical test to assess which signaling pathways are significantly enriched with genes that are common to hypoxia and CVD. In this case, $N$ is the total number of proteins in the human interactome, $K$ is the number of proteins overlapping with a signaling pathway, $n$ is the number of genes common to hypoxia and CVD, and $k$ is the number of genes common to hypoxia and CVD that overlap with the signaling pathway. The $P$-value indicates the probability of a random set of proteins with the same size as the gene set common to hypoxia and CVD having $k$ proteins overlapping with a signaling pathway. We also use this test to determine which hypoxia genes have a significant number of CVD interactors. In this case, $N$ is the total number of proteins in the human interactome, $K$ is the number of CVD genes in the human interactome, $n$ is the total number of interactors that a hypoxia gene has, and
$k$ is the number of CVD interactors that the hypoxia gene has. The $P$-value indicates the probability of a random set of proteins with the same size as the degree of the hypoxia gene containing $k$ CVD genes. For both cases, the $P$-values are adjusted for multiple comparisons by using $P.\ adjust$ in R (https://stat.ethz.ch/R-manual/Rdevel/library/stats/html/p.adjust.html) and the Bonferroni procedure to assess the false discovery rate.

**Results and discussion**

1. **Integrating hypoxia genes, CVD genes, and human interactome**

Hypoxia is a critical determinant of the pathobiology of many common cardiovascular diseases. To assess the functional association between hypoxia and cardiovascular diseases at a systems-level, we mapped hypoxia genes and CVD genes into the human interactome and constructed a hypoxia-CVD network that incorporates 850 hypoxia genes and 864 CVD genes which overlap with the proteins in the consolidated interactome. This subnetwork has 1290 proteins and 4860 interactions. We grouped the proteins according to their disease assignments, yielding the interaction patterns between different cardiovascular diseases as shown in figure 2(A). We note that in addition to the large number of overlapping disease genes, these CVDs also have many protein interactions between them, indicating possible transitions from one disease to another. As shown in figure 2(B), the hypoxia-CVD network has 4860 interactions, which is significantly greater than the number of interactions (1982 ± 176) in a subnetwork of random gene sets (Null Model I, $P$-value < $1.0 \times 10^{-16}$). The largest connected component (LCC) in the hypoxia-CVD network has 1269 nodes, which is also significantly larger than the size of the LCC (834 ± 35) in a subnetwork of random gene sets (Null Model I, $P$-value < $1.0 \times 10^{-16}$), as demonstrated in figure 2(C). These observations indicate that hypoxia and CVD-related genes tend to be densely connected in the human interactome.

The results obtained using degree-preserving random networks as the control (Null Model II) are consistent with what we observed using Null Model I (supplemental figure S1(A)-(B)). The hypoxia-CVD networks consist of three types of interactions: the interactions among hypoxia genes, the interactions among CVD genes, and the interactions between hypoxia genes and CVD genes. To assess the contribution of hypoxia genes and CVD genes to the high connectivity of the hypoxia-CVD network, we also examined the subnetwork induced by hypoxia genes only (hypoxia network) and the subnetwork induced by CVD genes only (CVD network). As shown in figure S1(C), the hypoxia network has 889 interactions, which are significantly more than the interactions (577 ± 77) in a subnetwork induced by a random gene set (Null Model I, $P$-value = $2.4 \times 10^{-5}$). The size of the LCC in the hypoxia network is significantly larger than that expected by chance (figure S1(D)). The CVD network has 2648 interactions, which are significantly more than the interactions (598 ± 76) in a subnetwork induced by a random gene set (Null Model I, $P$-value < $1.0 \times 10^{-16}$, figure S1(E)). The size of the LCC in the CVD network is significantly larger than random expectation (figure S1(F)). These results confirm that the high connectedness of the hypoxia-CVD network is not due to the combination of hypoxia genes and CVD genes, nor due to the primary contribution by one of the two groups.

Significantly more interactions in the hypoxia-CVD network may be due to the dense connections among disease genes caused by research biases. To reduce the potential bias caused by literature-based associations, we removed the literature-curated protein complexes from the
CORUM database and the literature-curated interactions from CCSB, which equates with removing \( \sim 30\% \) of the interactions from the human interactome. The results based on this reduced (but unbiased) human interactome are shown in supplemental figure S2, which are consistent with the results obtained using the whole human interactome, confirming the significance of the density and connectivity in the hypoxia-CVD network.

**Figure 2.** The hypoxia-CVD network and its topological properties. (A) Interaction patterns between hypoxia and different CVDs. The node size is proportional to the number of genes associated with the disease, and the edge width is proportional to the number of interactions between two diseases. (B) The number of edges in the hypoxia-CVD network is significantly larger than that expected by chance (Null Model I, \( P \)-value < \( 1.0 \times 10^{-16} \)). (C) The size of the largest connected component (LCC) in the hypoxia-CVD network is significantly larger than that expected by chance (Null Model I, \( P \)-value < \( 1.0 \times 10^{-16} \)).
2. Closeness between hypoxia and CVD genes

We next evaluated the topological closeness relationships between hypoxia and CVD genes in the human interactome. Among 850 hypoxia genes and 864 CVD genes, 95 are both hypoxia and CVD genes, i.e., 95 hypoxia-CVD gene pairs have distance 0. Considering the total number of proteins in the human interactome (14,174), the number of hypoxia and CVD genes that overlap is significantly larger than that expected by chance (hypergeometric test, \( P \)-value < \( 4.2 \times 10^{-9} \)), as shown in figure 3(A). Hypoxia and CVD are closely related not only by gene overlap analysis, but also at the network level. To test this hypothesis, we calculated how many hypoxia-CVD gene pairs have interactions, i.e., gene pairs with distance 1. There are 2314 interactions between hypoxia genes and CVD genes, which are significantly more than the interactions (1244 ± 115) between two random sets with the same number of genes as hypoxia and CVD gene sets (Null Model I, \( P \)-value < \( 1.0 \times 10^{-16} \)), as shown in figure 3(B). This analysis shows that 1030 hypoxia or CVD genes (64%) are connected by these interactions. Results obtained by using degree-preserving random networks as the control (Null Model II) supports the same conclusions (supplemental figure S3). Common neighbors have previously been used to predict whether two proteins have an (indirect) functional association. We, therefore, next checked the number of hypoxia-CVD gene pairs that have common neighbors, i.e., gene pairs with distance 2, and provide the results in figure 3(C). We note that there are significantly more hypoxia-CVD gene pairs with common neighbors than gene pairs from two random sets (\( P \)-value < \( 1.0 \times 10^{-16} \)). There are 1570 hypoxia/CVD genes (97%) involved in these hypoxia-CVD gene pairs, indicating that almost all hypoxia genes and CVD genes can be connected by only one neighbor. Figure 3(D) collectively shows that the number of hypoxia-CVD gene pairs with distance equal to or less than 3 is significantly larger than that for gene pairs from random sets (\( P \)-value < \( 1.0 \times 10^{-16} \)). These observations indicate that hypoxia genes are very closely connected to CVD genes in the human interactome.

We also calculated the average shortest path length between hypoxia genes and CVD genes in the human interactome. As shown in figure 3(E), the average shortest path length between hypoxia genes and CVD genes is 4.34, significantly smaller than the average shortest path length between two random sets (4.54 ± 0.02, \( P \)-value = \( 7.6 \times 10^{-24} \)). CC measures how near a node is to the rest of the network, i.e., how long it takes to spread information from a given node to all other nodes (assuming identical transmission rates). To determine whether or not hypoxia genes are central to CVD genes, we defined a modified CC parameter to measure the CC of hypoxia genes to CVD genes (see materials and methods). Figure 3(F) shows that hypoxia genes have an average CC of 0.303 to CVD genes, significantly higher than a random gene set’s CC (0.286 ± 0.002) to CVD genes (\( P \)-value < \( 1.0 \times 10^{-16} \)). These observations provide further evidence that hypoxia genes have close topological relationships to CVD genes compared to random expectation, which, in turn, reflects the close functional crosstalk or pathobiological relationships between hypoxia and CVDs.

To demonstrate the robustness of our conclusions, we performed two additional analyses to address the proximity of hypoxia genes and CVD genes at the systems-level. In the first analysis, we reduced the number of hypoxia genes and CVD genes by using tissue-specific gene expression data [47]. Specifically, the top 10% of genes that are highly expressed in cardiac tissues were compared with our lists of hypoxia genes and CVD genes, which leads to 169 hypoxia genes and 158 CVD genes that are cardiac-specific. We then checked whether or not these two small tissue-specific gene sets have significant more interactions than that expected by...
Figure 3. Closeness relationships between hypoxia genes and CVD genes. (A) The overlap between hypoxia genes and CVD genes. (B) Hypoxia genes and CVD genes have significantly more interactions than that expected by chance ($P$-value < $1.0 \times 10^{-16}$). (C) There are significantly more hypoxia-CVD gene pairs with common neighbors than expected by chance ($P$-value < $1.0 \times 10^{-16}$). (D) There are significantly more hypoxia-CVD gene pairs with distance equal to or less than 3 than that expected by chance ($P$-value < $1.0 \times 10^{-16}$). (E) The average shortest path length between hypoxia genes and CVD genes is significantly smaller than that between two random gene sets ($P$-value < $1.0 \times 10^{-16}$). (F) Hypoxia genes have significantly greater closeness centrality to CVD genes than that expected by chance ($P$-value < $1.0 \times 10^{-16}$).
chance. The results are summarized in figures S3(B)–(D). We can see that the two small tissue-specific gene sets have stronger overlap (in terms of the $P$-value), and, moreover, still have significantly more interactions than random expectation. We also constructed the cardiac-specific hypoxia-CVD network as shown in figure S4. In the second analysis, we separated different types of networks and evaluated whether there are still significantly more interactions between hypoxia genes and CVD genes in these different networks. To avoid the bias resulting from literature-based interactions (LCI), we removed the interactions from this source. The interactions from protein complexes [32, 33, 39] constitute a co-complex interaction network. All other interactions including protein interactions generated by yeast-2-hybrid, signaling interactions, metabolic interactions, and kinase-substrate interactions form a binary interaction network. The two types of network may have different degree distributions, therefore, we use Null model II to generate the control. The results are summarized in figures S3(E) and (F), which show that there are still significantly more interactions between hypoxia genes and CVD genes in the binary interaction network and the co-complex interaction network, indicating that the close relationships of hypoxia genes and CVD genes are not simply a consequence of the bias of literature-curated interactions.

3. Common pathways of hypoxia and CVDs

The six CVDs we considered here are highly related, and one disease may be the precursor or consequence of another (such as thrombosis and myocardial ischemia). It would, therefore, be interesting to determine if network analysis can be used to show that hypoxia plays a central role in accounting for the transition of two linked diseases. To test this hypothesis, we defined a modified BC measure, which reflects the control of a node on the communication between other nodes (materials and methods). We used this modified BC measure to quantify the ‘bridgeness’ of a hypoxia gene in connecting two diseases. To remove the potential biases caused by overlapping disease genes, we only consider disease-specific genes. The BC values of hypoxia genes in connecting different disease pairs are summarized in figure 4. The BC of a node is highly correlated with its degree [48]. Thus, to evaluate the significance of the bridging role of hypoxia genes in connecting two diseases, we generated 1000 random networks each preserving the degree-distribution of the network as the control (Null Model II). From figure 4 we note that hypoxia genes play a significant role in bridging most of the disease pairs. The unnormalized BC measure scales with the number of gene pairs (materials and methods), which may not facilitate a comparison of different disease pairs. The results based on a normalized BC measure are given in supplemental figure S4, which confirms the results obtained based on the unnormalized BC measure. We also plotted heat maps for the normalized BC values and their associated $P$-values in supplemental figure S5, which indicates that hypoxia is more involved in connecting cerebrovascular-related and PH-related disease pairs than other disease pairs.

In addition to the bridgeness of hypoxia genes in connecting different CVDs, we observed that there are 95 hypoxia genes that are also CVD genes. We, therefore, next examined which signaling pathways are common to both hypoxic stress signaling and CVDs. To this end, we mapped the hypoxia-CVD overlapping genes to knowledge-based signaling pathways from two public databases and one licensed database. ConsensusPathDB integrates 4576 signaling pathways from different source databases [49]. The Molecular Signatures Database (MSigDB) (www.broadinstitute.org/gsea/msigdb) is a collection of annotated gene sets for gene set enrichment analysis [50]. The version we used is msigdb.v4.0, which contains 10 295 annotated
gene sets or signaling pathways. For the signaling pathways in ConsensusPathDB and MSigDB, we used the hypergeometric distribution-based statistical test (materials and methods) to determine which signaling pathways are significantly enriched with genes common to hypoxia and CVD. MetaCore (http://thomsonreuters.com/metacore/) is an integrated commercial software suite from Thomson Reuters for functional and pathway analysis. Through the pathway mapping analysis, we obtained a list of signaling pathways that are significantly

**Figure 4.** Bridgeness of hypoxia genes in connecting different CVDs. (A) Atherosclerosis versus cerebrovascular diseases; (B) atherosclerosis versus CAD; (C) atherosclerosis versus myocardial ischemia; (D) atherosclerosis versus pulmonary hypertension; (E) cerebrovascular diseases versus thrombosis; (F) cerebrovascular diseases versus CAD; (G) cerebrovascular diseases versus myocardial ischemia; (H) cerebrovascular diseases versus pulmonary hypertension; (I) cerebrovascular diseases versus thrombosis; (J) CAD versus myocardial ischemia; (K) CAD versus pulmonary hypertension; (L) CAD versus thrombosis; (M) myocardial ischemia versus pulmonary hypertension; (N) myocardial ischemia versus thrombosis; and (O) pulmonary hypertension versus thrombosis. ‘Interactome’ represents the human interactome, and ‘RN’ denotes the random networks.
enriched with genes common to hypoxia and CVD. These significant pathways are viewed as common pathways of the hypoxic response and CVDs. The results are provided in supplemental file 2. From each database, we find a number of common signaling pathways of hypoxia and CVD. We randomly selected 10 from each database and provided them in Table 1, which shows that some of these signaling pathways are biologically plausible, e.g.,

| Pathway maps                                      | Sources          | Size | In data | P-Values     |
|--------------------------------------------------|------------------|------|---------|--------------|
| Adipogenesis                                     | ConsensusPathDB  | 130  | 12      | 7.2 × 10⁻⁹   |
| HIF-1-alpha transcription factor network         | ConsensusPathDB  | 66   | 10      | 2.0 × 10⁻⁸   |
| Antimetabolite pathway–folate cycle              | ConsensusPathDB  | 13   | 4       | 3.9 × 10⁻⁵   |
| Metabolism of vitamins and cofactors             | ConsensusPathDB  | 75   | 7       | 7.5 × 10⁻⁵   |
| Platelet aggregation inhibitor pathway           | ConsensusPathDB  | 49   | 6       | 1.2 × 10⁻⁴   |
| Reversal of insulin resistance by leptin         | ConsensusPathDB  | 9    | 3       | 5.4 × 10⁻⁴   |
| Angiogenesis                                     | ConsensusPathDB  | 23   | 4       | 1.2 × 10⁻³   |
| Collagen formation                               | ConsensusPathDB  | 88   | 6       | 2.8 × 10⁻³   |
| HDL-mediated lipid transport                     | ConsensusPathDB  | 15   | 3       | 7.5 × 10⁻³   |
| Hypoxia-inducible factor in the cardiovascular system | ConsensusPathDB  | 16   | 3       | 0.01         |
| Vart_KSHV_infection_angiogenic_markers_up       | MSigDB           | 165  | 17      | 1.4 × 10⁻⁷   |
| Burton_adipogenesis_peak_at_2_h                  | MSigDB           | 51   | 7       | 1.3 × 10⁻⁵   |
| Smirnov_circulating_endotheliocytes_in_cancer_up | MSigDB           | 158  | 10      | 4.2 × 10⁻⁵   |
| Reactome_metabolism_of_vitamins_and_cofactors    | MSigDB           | 51   | 6       | 1.8 × 10⁻⁴   |
| Baker_hematopoiesis_STAT3_targets                | MSigDB           | 16   | 4       | 3.6 × 10⁻⁴   |
| Rozanov_MMP14_targets_subset                     | MSigDB           | 35   | 5       | 4.2 × 10⁻⁴   |
| Weinmann_adaptation_to_hypoxia_dn               | MSigDB           | 41   | 5       | 2.1 × 10⁻³   |
| Regulation_of_angiogenesis                       | MSigDB           | 26   | 4       | 6.0 × 10⁻³   |
| Vasculature_development                          | MSigDB           | 55   | 5       | 0.013        |
| Thum_Mir21_targets_heart_disease_up              | MSigDB           | 17   | 3       | 0.025        |
| Development_leptin signaling via JAK/STAT and MAPK cascades | MetaCore         | 26   | 6       | 2.1 × 10⁻⁵   |
| Development_regulation_of_epithelial-to-mesenchymal transition (EMT) | MetaCore         | 64   | 7       | 3.2 × 10⁻⁴   |
| Regulation of lipid metabolism_G-alpha(q) regulation of lipid metabolism | MetaCore         | 66   | 6       | 1.1 × 10⁻³   |
| Role of tissue factor-induced thrombin signaling in carcinogenesis | MetaCore         | 66   | 6       | 1.4 × 10⁻³   |
| Hypoxia-induced EMT in cancer and fibrosis       | MetaCore         | 9    | 3       | 2.1 × 10⁻³   |
| Resolution of inflammation in healing myocardial infarction | MetaCore         | 25   | 4       | 2.3 × 10⁻³   |
| Cell adhesion_ECM remodeling                     | MetaCore         | 52   | 5       | 2.9 × 10⁻³   |
| Adiponectin in pathogenesis of type 2 diabetes   | MetaCore         | 29   | 3       | 0.022        |
| Development_inhibition_of_angiogenesis_by_PEDF  | MetaCore         | 31   | 3       | 0.026        |
| Muscle contraction_regulation_of_eNOS_activity_in endothelial cells | MetaCore         | 65   | 4       | 0.027        |
angiogenesis, regulation of eNOS activity in endothelial cells, and hypoxia-induced EMT in cancer and fibrosis. These significant signaling pathways provide important biological insights into the crosstalk of hypoxic signaling and the pathobiology of CVDs.

4. Hypoxia-CVD functional modules

Pathway mapping can provide insights into potential crosstalk in signaling pathways involved in both hypoxia and CVDs; however, this analysis is limited to our (reductionist) knowledge of signaling pathways. Importantly in this regard, the current signaling pathway databases are far from complete. There is increasing evidence that, at the both molecular level and physiological level, the function of a biological interaction network is closely related to its topological structure [51, 52]. Therefore, densely connected subgraphs in protein interaction networks have been widely used to predict protein complexes and novel functional modules [53]. These network modules, when dysfunctional owing to biological variants that change the strength of interactions, define disease modules, a general concept that serves as the basis for the new discipline of network medicine [18]. To detect potential functional associations between hypoxia and CVDs, we constructed a bipartite network of hypoxia genes and CVD genes (materials and methods). The bipartite network has 1030 proteins and 2314 interactions. As it does not consider the connectivity within hypoxia genes or within CVD genes, the connectivity of a functional module in the bipartite network represents the association strength of hypoxia-CVD in this module. We used the Louvain method to maximize the modularity function defined for bipartite networks and thereby identify hypoxia-CVD gene modules (see materials and methods). A total of 20 hypoxia-CVD functional modules comprising more than five proteins were detected in the hypoxia-CVD bipartite network. We then restored interactions among hypoxia genes or CVD genes within the identified modules, as shown by three examples in figure 5. In figure 5(C), Module 16 illustrates the central linking role of the transforming growth factor-beta signaling pathway (bone morphogenetic receptor protein 2 (BMPR2) and transforming growth factor beta-1 (TGFβ1)) and of the tumor necrosis factor signaling pathway (tumor necrosis factor receptor-associated protein 1 (TRAP1)), neither of which had been recognized previously as common mechanistic links from hypoxia to CVDs. A full list of all such functional modules is provided in supplemental file 3 and supplemental figure S6.

Proteins in a common protein complex tend to perform similar functions. To validate the functionality of the detected densely connected modules, we assessed whether or not the proteins in these modules are functionally similar. A number of measures have been developed to evaluate numerically the functional similarity of genes or proteins based on their annotations. We used the measure GS2 (GO-based similarity of gene sets) developed in [54] to quantify the functional similarity of two proteins; this measure is based on the GO annotations of genes or gene products. In addition to having accuracy comparable to other semantic methods [55, 56], this measure is faster and can be used for analysis of large-scale gene sets. The daily snapshot of the GO tree and human gene annotations were downloaded from the GO web site (www.geneontology.org) on 5 December 2013. We evaluated the functional similarity of each module in two ways. First, we calculated the node-based functional similarity, i.e., the functional similarity of all node pairs in a module, and then took the average value. As a control for significance, we randomly selected a gene set with the same number of proteins as the module and calculated its functional similarity for comparison. Second, we calculated the edge-based functional similarity, i.e., the functional similarity of interacting node pairs in a module. As a
control for significance, we randomly selected an interaction set with the same number of interactions as the module and calculated its functional similarity for comparison. Figure 6 shows both the node-based similarity and the edge-based similarity of each module and comparative random sets. We note that most of the identified modules have higher functional similarity than random sets, suggesting that these modules are mechanistically rational and represent functional crosstalk between hypoxia and CVD pathways.

5. Predicting new CVD genes

In the previous sections, we examined the relationships between hypoxia genes and CVD genes at a macroscale level—global topological analyses, pathway mapping, and functional module

![Network Diagram](image-url)
detection. In this section, we analyze individual hypoxia genes and check the biological link between hypoxia and CVD genes at a microscale level. Specifically, we wish to predict whether or not a hypoxia gene is highly likely to be involved in CVDs. To this end, we first checked whether or not hypoxia genes tend to interact with CVD genes overall. We generated 1000 random networks with the same degree distribution as the human interactome, and calculated the number of CVD interactors with each hypoxia gene in these random networks. Figure 7 shows the average number of CVD interactors with hypoxia genes in the human interactome and random networks, respectively. We can see that hypoxia genes falling in different degree bins all have a significantly higher number of CVD interactors in the human interactome than in random networks. This observation supports the view that hypoxia genes tend to interact with CVD genes. We next evaluated the enrichment of CVD interactors for each hypoxia gene using the hypergeometric test followed by the Bonferroni correction of P-values to control the false discovery rate caused by multiple hypothesis testing (see materials and methods). Table 2 lists the prediction of hypoxia genes associated with CVDs with Bonferroni-corrected P-values. Many of these predictions are mechanistically rational, and a few of them have been found to be associated with cardiovascular risk, such as IDE, VLDLR, HGF, and FLT1 [57–60]. Our prediction results emphasize the hypoxic stress responsive role of these genes in the pathobiology of cardiovascular diseases.

Figure 6. Average GO functional similarity of each module. The dashed lines represent the GO similarity of random sets, i.e., the average functional similarity of the random sets for each module; (A) node-based GO similarity; (B) edge-based GO similarity.
Figure 7. The average number of CVD interactors of hypoxia genes in the human interactome (blue circles) and in random networks with preserved degree distribution (red diamonds), respectively. $K_{\text{max}} = 414$.

Table 2. Prediction of hypoxia genes associated with cardiovascular diseases with Bonferroni-corrected $P$-values.

| Gene symbols | Descriptions                                                                 | # CVD Interactors | $P$-Values  |
|--------------|------------------------------------------------------------------------------|-------------------|-------------|
| PRKCA        | Protein kinase C alpha type                                                   | 66                | $8.8 \times 10^{-9}$ |
| CD276        | Costimulatory molecule                                                       | 18                | $1.5 \times 10^{-8}$ |
| OSMR         | Oncostatin-M specific receptor beta subunit                                   | 12                | $2.4 \times 10^{-8}$ |
| IDE          | Insulin-degrading enzyme                                                      | 6                 | $1.6 \times 10^{-6}$ |
| VLDLR        | Very low density lipoprotein receptor                                         | 7                 | $3.8 \times 10^{-6}$ |
| TGFB2        | Glioblastoma-derived T-cell suppressor factor                                 | 16                | $1.3 \times 10^{-5}$ |
| VEGFC        | Vascular endothelial growth factor-related protein                           | 8                 | $8.9 \times 10^{-5}$ |
| GDF6         | Growth/differentiation factor 16                                              | 11                | $4.0 \times 10^{-4}$ |
| ITGA11       | Integrin, alpha 11                                                           | 9                 | $7.5 \times 10^{-4}$ |
| HLA-E        | HLA class I histocompatibility antigen, E alpha chain                         | 9                 | 0.002        |
| HGF          | Hepatocyte growth factor (hepapoitin A; scatter factor)                      | 8                 | 0.003        |
| ITGA10       | Integrin, alpha 10                                                           | 6                 | 0.004        |
| BGN          | Dermatan sulfate proteoglycan I                                              | 7                 | 0.004        |
| APAF1        | Apoptotic peptidase activating factor                                        | 8                 | 0.005        |
| CADM1        | Immunoglobulin superfamily, member 4D variant 2                              | 9                 | 0.007        |
| DOK5         | Downstream of tyrosine kinase 5                                              | 6                 | 0.007        |
| ADRA1B       | Adrenergic, alpha-1B-, receptor                                              | 10                | 0.007        |
| MYD88        | Myeloid differentiation primary response protein                             | 10                | 0.007        |
| F2RL1        | Coagulation factor II (thrombin) receptor-like 1                             | 5                 | 0.009        |
| CEBPD        | CCAAT/Enhancer binding protein (C/EBP), delta                                | 7                 | 0.009        |
| FLT1         | Fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor) | 12                | 0.018        |
| GDF10        | Growth differentiation factor 10                                             | 9                 | 0.020        |
| EPHB2        | Developmentally-regulated eph-related tyrosine kinase                        | 15                | 0.024        |
| MGLL         | Lysophospholipase homologue                                                  | 6                 | 0.028        |
| PREP         | Prolyl endopeptidase                                                         | 4                 | 0.044        |
Conclusions

The biological link between hypoxia and CVDs has been recognized for many years, but few studies have investigated this relationship from a systems-level perspective. In this study, we carried out a thorough investigation of the closeness relationships of hypoxia and cardiovascular diseases in the human interactome by integrating hypoxia genes, CVD genes, and human molecular interaction network data. The analyses from macroscale to microscale all clearly indicate the presence of close relationships between hypoxia genes and CVD genes. We find that hypoxia genes are much closer to CVD genes in the human protein interactome than that expected by chance. More importantly, we observe that hypoxia genes play significant bridging roles in connecting different CVDs. A few signaling pathways have been identified as common to both hypoxic stress responses and CVDs. These signaling pathways and the newly-identified hypoxia-CVD functional modules provide new insights for the role of hypoxia in cardiovascular biology and disease.

Acknowledgements

This work was supported by NIH grants HL61795, HL108630 (MAPGen Consortium), and HL048743 (to JL). We wish to thank Professor Barabasi’s laboratory at Northeastern University for providing us with the consolidated human interactome data they compiled.

Reference

[1] Semenza G L 2011 Oxygen sensing, homeostasis, and disease N. Engl. J. Med. 365 537–47
[2] Semenza G L 2014 Hypoxia-inducible factor 1 and cardiovascular disease Annu. Rev. Physiol. 76 39–56
[3] Simon M C et al 2008 Hypoxia-induced signaling in the cardiovascular system Annu. Rev. Physiol. 70 51–71
[4] Giordano F J 2005 Oxygen, oxidative stress, hypoxia, and heart failure J. Clin. Invest. 115 500–8
[5] Huang Y et al 2004 Cardiac myocyte-specific HIF-1alpha deletion alters vascularization, energy availability, calcium flux, and contractility in the normoxic heart FASEB J. 18 1138–40
[6] Semenza G L 2014 Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology Annu. Rev. Pathol. 9 47–71
[7] MacLellan W R, Wang Y and Lusis A J 2012 Systems-based approaches to cardiovascular disease Nat. Rev. Cardiol. 9 172–84
[8] Lusis A J and Weiss J N 2010 Cardiovascular networks: systems-based approaches to cardiovascular disease Circulation 121 157–70
[9] Ramsey S A, Gold E S and Aderem A 2010 A systems biology approach to understanding atherosclerosis EMBO Mol. Med. 2 79–89
[10] Vidal M, Cusick M E and Barabasi A L 2011 Interactome networks and human disease Cell 144 986–98
[11] Meng Q et al 2013 Systems biology approaches and applications in obesity, diabetes, and cardiovascular diseases Curr. Cardiovasc. Risk Rep. 7 73–83
[12] Wang J et al 2009 Disease-aging network reveals significant roles of aging genes in connecting genetic diseases PLoS Comput. Biol. 5 e1000521
[13] Sarajlic A et al 2013 Network topology reveals key cardiovascular disease genes PLoS One 8 e71537
[14] Dewey F E et al 2011 Gene coexpression network topology of cardiac development, hypertrophy, and failure Circ. Cardiovasc. Genet. 4 26–35
[15] Jensen M K et al 2011 Protein interaction-based genome-wide analysis of incident coronary heart disease Circ. Cardiovasc. Genet. 4 549–56
[16] Huan T et al 2013 A systems biology framework identifies molecular underpinnings of coronary heart disease Arterioscler. Thromb. Vasc. Biol. 33 1427–34
[17] Mayr M 2011 Recent highlights of metabolomics in cardiovascular research Circ. Cardiovasc. Genet. 4 463–4
[18] Barabasi A L, Gulbahce N and Loscalzo J 2011 Network medicine: a network-based approach to human disease Nat. Rev. Genet. 12 56–68
[19] Smyth G K 2005 Limma: linear models for microarray data Bioinformatics and Computational Biology Solutions Using R and Bioconductor ed R Gentleman et al (New York: Springer) pp 397–420
[20] Manalo D J et al 2005 Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1 Blood 105 659–69
[21] Schriml L M et al 2012 Disease ontology: a backbone for disease semantic integration Nucleic Acids Res. 40 D940–6 (Database issue)
[22] Yu W et al 2008 A navigator for human genome epidemiology Nat. Genet. 40 124–5
[23] Du P et al 2009 From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations Bioinformatics 25 i63–8
[24] Stelzl U et al 2005 A human protein–protein interaction network: a resource for annotating the proteome Cell 122 957–68
[25] Yu H et al 2011 Next-generation sequencing to generate interactome datasets Nat. Methods 8 478–80
[26] Venkatesan K et al 2009 An empirical framework for binary interactome mapping Nat. Methods. 6 83–90
[27] Rual J F et al 2005 Towards a proteome-scale map of the human protein–protein interaction network Nature 437 1173–8
[28] Aranda B et al 2010 The IntAct molecular interaction database in 2010 Nucleic Acids Res. 38 (Database issue) D525–31
[29] Ceol A et al 2009 MINT, the molecular interaction database: 2009 update Nucleic Acids Res. 2010 38 (Database issue) D532–9
[30] Keshava Prasad T S et al 2009 Human protein reference database: 2009 update Nucleic Acids Res. 37 D767–72 (Database issue)
[31] Stark C et al 2011 The BioGRID Interaction Database: 2011 update Nucleic Acids Res. 39 (Database issue) D698–704
[32] Ruepp A et al 2010 CORUM: the comprehensive resource of mammalian protein complexes—2009 Nucleic Acids Res. 38 (Database issue) D497–501
[33] Havugimana P C et al 2012 A census of human soluble protein complexes Cell 150 1068–81
[34] Matys V et al 2003 TRANSFAC: transcriptional regulation, from patterns to profiles Nucleic Acids Res. 31 374–8
[35] Hornbeck P V et al 2012 PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse Nucleic Acids Res. 40 (Database issue) D261–70
[36] Lee D S et al 2008 The implications of human metabolic network topology for disease comorbidity Proc. Natl Acad. Sci. USA 105 9880–5
[37] Kanehisa M et al 2004 The KEGG resource for deciphering the genome Nucleic Acids Res. 32 (Database issue) D277–80
[38] Schellenberger J et al 2010 BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions BMC Bioinformatics 11 213
[39] Zhang Q C et al 2012 Structure-based prediction of protein–protein interactions on a genome-wide scale Nature 490 556–60
[40] Vinayagam A et al 2011 A directed protein interaction network for investigating intracellular signal transduction Sci. Signal 4 rs8
[41] Hagberg A A, Schult D A and Swart P J 2008 Exploring network structure, dynamics, and function using NetworkX Proc. of the 7th Python in Science Conference (SciPy2008) ed G Varoquaux, T Vaught and J Millman (Pasadena, CA, USA) pp 11–15

[42] Freeman L C 1978 Centrality in social networks: conceptual clarification Social Netw. 1 215–39

[43] Freeman L C 1977 A set of measures of centrality based on betweenness Sociometry 40 35–41

[44] Barber M J 2007 Modularity and community detection in bipartite networks Phys. Rev. E 76 066102

[45] Guimera R and Nunes Amaral L A 2005 Functional cartography of complex metabolic networks Nature 433 895–900

[46] Blondel V D et al 2008 Fast unfolding of communities in large networks J. Stat. Mech. 10 P10008

[47] Su A I et al 2004 A gene atlas of the mouse and human protein-encoding transcriptomes Proc. Natl Acad. Sci. USA 101 6062–7

[48] Valente T W et al 2008 How correlated are network centrality measures? Connect (Tor) 28 16–26

[49] Kamburov A et al 2013 The ConsensusPathDB interaction database: 2013 update Nucleic Acids Res. 41 (Database issue) D793–800

[50] Subramanian A et al 2005 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles Proc. Natl Acad. Sci. USA 102 15545–50

[51] Bashan A et al 2012 Network physiology reveals relations between network topology and physiological function Nat. Commun. 3 702

[52] Jeong H et al 2001 Lethality and centrality in protein networks Nature 411 41–2

[53] Spirin V and Miny L A 2003 Protein complexes and functional modules in molecular networks Proc. Natl Acad. Sci. USA 100 12123–8

[54] Ruths T, Ruths D and Nakhleh L 2009 GS2: an efficiently computable measure of GO-based similarity of gene sets Bioinformatics 25 1178–84

[55] Sevilla J L et al 2005 Correlation between gene expression and GO semantic similarity IEEE/ACM Trans. Comput. Biol. Bioinform. 2 330–8

[56] Wang J Z et al 2007 A new method to measure the semantic similarity of GO terms Bioinformatics 23 1274–81

[57] Crawford D C et al 2008 A common VLDLR polymorphism interacts with APOE genotype in the prediction of carotid artery disease risk J. Lipid Res. 49 588–96

[58] Florez J C et al 2006 High-density haplotype structure and association testing of the insulin-degrading enzyme (IDE) gene with type 2 diabetes in 4206 people Diabetes 55 128–35

[59] Lonn J et al 2011 Hepatocyte growth factor in patients with coronary artery disease and its relation to periodontal condition. Results Immunol. 2 7–12

[60] Madonna R and De Caterina R 2009 VEGF receptor switching in heart development and disease Cardiovasc. Res. 84 4–6