BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 12 2014, pages 1721–1729
doi:10.1093/bioinformatics/btu089

Systems biology

Advance Access publication February 18, 2014

Dynamic networks reveal key players in aging

Fazle E. Faisal and Tijana Milenković*

Department of Computer Science and Engineering, ECK Institute for Global Health and Interdisciplinary Center for Network Science and Applications, University of Notre Dame, Notre Dame, IN 46556, USA

Associate Editor: Igor Jurisica

ABSTRACT

Motivation: Because susceptibility to diseases increases with age, studying aging gains importance. Analyses of gene expression or sequence data, which have been indispensable for investigating aging, have been limited to studying genes and their protein products in isolation, ignoring their connectivities. However, proteins function by interacting with other proteins, and this is exactly what biological networks (BNs) model. Thus, analyzing the proteins' BN topologies could contribute to the understanding of aging. Current methods for analyzing systems-level BNs deal with their static representations, even though cells are dynamic. For this reason, and because different data types can give complementary biological insights, we integrate current static BNs with aging-related gene expression data to construct dynamic age-specific BNs. Then, we apply sensitive measures of topology to the dynamic BNs to study cellular changes with age.

Results: While global BN topologies do not significantly change with age, local topologies of a number of genes do. We predict such genes to be aging-related. We demonstrate credibility of our predictions by (i) observing significant overlap between our predicted aging-related genes and 'ground truth' aging-related genes; (ii) observing significant overlap between functions and diseases that are enriched in our aging-related predictions and those that are enriched in 'ground truth' aging-related data; (iii) providing evidence that diseases which are enriched in our aging-related predictions are linked to human aging; and (iv) validating our high-scoring novel predictions in the literature.

Availability and implementation: Software executables are available upon request.

Contact: tmilenko@nd.edu

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

Received on March 15, 2013; revised on December 6, 2013; accepted on February 10, 2014

1 INTRODUCTION

1.1 Motivation and background

Because the US population is, on average, growing older as a result of ~78 million baby boomers who began turning 65 in 2011, and because susceptibility to diseases increases with age, studying human aging gains importance. Analysis of gene expression data has been indispensable for investigating aging (Fortney *et al.*, 2010; Wieser *et al.*, 2011). However, it has mostly been limited to studying differential expression of individual genes, without considering their connectivities (Fortney *et al.*, 2010). However, it is the proteins (gene products) that carry out cellular processes, and they do so by interacting with other proteins instead of acting alone. And this is exactly what BNs and *protein–protein interaction (PPI) networks* in particular model; in PPI networks, nodes are proteins and edges correspond to physical interactions between the proteins. Thus, analyzing topologies of proteins in PPI networks could contribute to the understanding of the processes of aging. Although, as a proof of concept, this study focuses on PPI networks, it is applicable to other types of BNs. High-throughput screens for PPI detection have yielded systems-level (though incomplete) PPI networks for many organisms, which are publicly available (Breitkreutz *et al.*, 2008).

The majority of current methods for analyzing systems-level PPI networks deal with their *static* representations, due to limitations of biotechnologies for PPI collection, even though cells are dynamic (Przytycka and Kim, 2010). For this reason, and because different data types can give complementary biological insights (Memišević *et al.*, 2010b; Przytycka and Kim, 2010), we *integrate* current static PPI network data (Breitkreutz *et al.*, 2008; Peri *et al.*, 2004) with age-specific gene expression data (Berchtold *et al.*, 2008) to computationally construct *dynamic age-specific PPI networks*, to study cellular changes with age from such networks.

Furthermore, topological positions of aging-related genes in the *static* networks have been studied (de Magalhães, 2009; Ferrarini *et al.*, 2005; Kriete *et al.*, 2011; Promislow, 2004; Reja *et al.*, 2009), but mostly with *crude* measures of topology that cannot cope with the complexity of PPI networks (Pržulj, 2011). For example, node degrees have been used to argue the central role or aging-related proteins in the yeast network compared with proteins that are not associated with aging, or to study the role of chaperones (heat shock proteins) in aging (Promislow, 2004; Sőti and Csermely, 2007). In addition to aging, many approaches have aimed to link node degrees with, for example, essentiality (Jeong *et al.*, 2001), disease (Sharan and Ideker, 2008; Vanunu *et al.*, 2010), cancer (Aragues *et al.*, 2008; Jonsson and Bates, 2006) or pathogenicity (Dyer *et al.*, 2008). However, it is possible that the high-degree proteins have been more studied simply because of their known relevance to human health (Pržulj, 2011; Ratmann *et al.*, 2009). Hence, more constraining measures of topology might be needed that go beyond capturing only the direct network neighborhood of a node (Milenković and Pržulj, 2008; Milenković *et al.*, 2010). While such measures exist and have been used to link proteins' network positions with their involvement in some biological processes (Milenković *et al.*, 2011; Sharan *et al.*, 2007), to our knowledge,

*To whom correspondence should be addressed.

they have not been linked to proteins' involvement in aging even in static, and especially in dynamic PPI networks. Here, we apply *a series* of measures of topology, including some highly sensitive measures (Milenković and Pržulj, 2008; Milenković *et al.*, 2011), to the *dynamic* PPI networks to identify key players in aging.

1.2 Our study

We aim to study human aging via integration of aging-related gene expression data with static PPI network data (Fig. 1). We obtain dynamic age-specific PPI networks by selecting in the static network (i) all proteins that correspond to actively expressed genes at different ages and (ii) all PPIs involving these 'active' proteins. Hence, each age-specific network is the network that is 'active' at a given age. We hypothesize that the dynamic and integrative network analysis provides a valuable model of cellular functioning that can reveal aging-related information and that can reveal more of the information than static analysis of individual data types.

Given the dynamic network data, we first aim to answer whether the overall network topologies change with age (Fig. 1). Because this is *not* the case and because the gene expression data alone may reveal only a small portion of all genes as aging-related (Berchtold *et al.*, 2008; Lu *et al.*, 2004), it could be that *local* topologies around only a *subset* of proteins in the network do change with age. Hence, we study positions of proteins in each age-specific PPI network with respect to measures of local topology, called *node centralities*, with the goal of identifying proteins whose centralities significantly change with age. We find 537 such proteins (5.7% of all proteins in the static network), which is quantitatively consistent with the result of the gene expression study by Lu *et al.* (2004). We predict these proteins to be aging-related and validate them as follows:

(1) The predictions are statistically significant, i.e. non-random.

(2) The overlap of our predictions and 'ground truth' aging-related genes is significant. Nonetheless, many of our predictions are *novel*, i.e. absent from the 'ground truth' data. This confirms that dynamic network analysis of integrated data types can reveal additional biological knowledge compared with static analysis of individual data.

(3) There is an overlap between functions and diseases that are enriched in our predictions and those that are enriched in the 'ground truth' aging-related data. Also, diseases that are enriched in our predictions are linked to human aging.

(4) We *manually* search in the literature for 10% of our highest-scoring predictions that are not present in the 'ground truth' aging-related data, and we successfully validate all of them.

2 MATERIALS AND METHODS

2.1 Data

2.1.1 Aging-related gene expression data As a proof of concept of our approach, we use a microarray human brain gene expression data set consisting of 173 samples obtained from 55 individuals, spanning 37 different ages between 20 and 99 years (Berchtold *et al.*, 2008). We define a gene as being *expressed* (or active) in a given sample by following the methodology of Lu *et al.*'s work (2004). Because the expression data consist of multiple samples per age, we integrate all samples belonging to a given age and define a gene as being expressed at that age according to the majority vote rule (Lee *et al.*, 2000). For details, see Supplementary Section S1.1.

In addition, to test the robustness of our approach, we use an alternative RNA-seq human brain gene expression data consisting of 15 samples obtained from 15 individuals, spanning 15 age-groups between 0 and 98 years (Mazin *et al.*, 2013) (Supplementary Section S1.1). As we show in Section 3.3, using this alternative RNA-seq data instead of the above microarray data makes no significant effect on our aging-related predictions.

2.1.2 Static PPI network data We obtain human static PPIs from HPRD (Peri *et al.*, 2004) and BioGRID (Breitkreutz *et al.*, 2008). HPRD data consist of 9617 unique proteins (with respect to their gene IDs) and

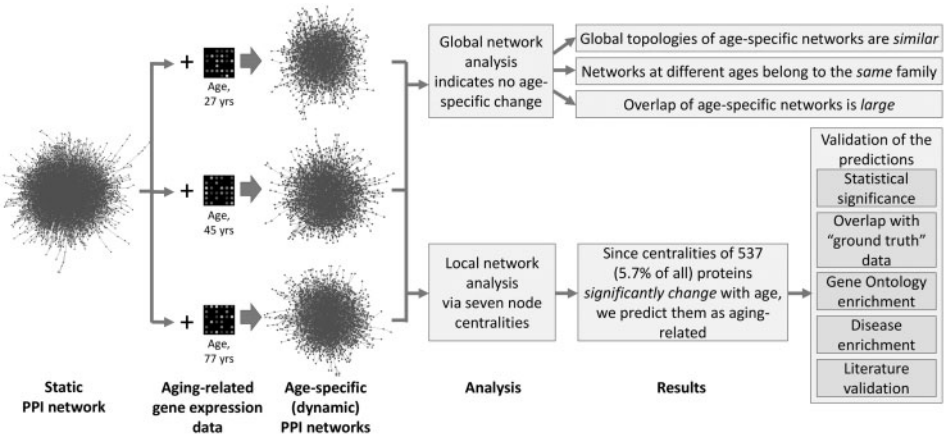

Fig. 1. Summary of our study. We integrate a static PPI network with aging-related gene expression data to obtain age-specific networks. We analyze changes in global and local network topologies with age. While global network analysis indicates no age-specific change, local topologies (as captured by seven node centrality measures) of some proteins do significantly change with age. We predict such proteins to be aging-related and validate our predictions in several ways

37 039 unique PPIs between the proteins. BioGRID data consist of 10 078 unique proteins and 50 954 unique PPIs between the proteins.

2.1.3 Integrating static PPI network with gene expression data to form age-specific PPI networks We form dynamic age-specific networks as follows. To form a network specific to a given age, we select in the static network those proteins that are expressed at that age (Section 2.1.1) and all PPIs that exist between the expressed proteins (see Supplementary Section S1.2 for a formal description). Because gene expression data are collected for 37 ages, 37 age-specific networks are formed from the given static network. Because we study two static networks (HPRD and BioGRID), we obtain two sets of dynamic networks. We run subsequent analyses on each of the network sets. Because we find that results are similar across the two sets, for simplicity, here we report results only for the HPRD network. Results for the BioGRID network are reported in Supplementary Table S2 and Supplementary Figs S2 and S5.

2.1.4 'Ground truth' aging-related data We denote the set of 9426 genes that are present in both the static PPI network (Section 2.1.2) and brain gene expression data (Section 2.1.1) as *StatNetExpression*.

By studying the brain gene expression data from Section 2.1.1, Berchtold *et al.* (2008) identified 8277 genes whose expression significantly changed with age. Of these, 4520 are present in StatNetExpression. Henceforth, we denote this 'ground truth' aging-related set of 4520 genes predicted from *brain gene expression data alone* as *BrainExpression2008Age*.

By studying a different brain gene expression data, Lu *et al.* (2004) predicted 442 genes as aging-related, as their expression significantly correlated with age. Of these, 343 genes are present in StatNetExpression. Henceforth, we denote this 'ground truth' aging-related set of 343 genes predicted from *brain gene expression data alone* as *BrainExpression2004Age*.

Clearly, BrainExpression2008Age and BrainExpression2004Age are similar in the sense that their aging-related genes have been inferred from brain gene expression data. [And as such, among all 'ground truth' data sets (see below), these two sets are expected to be the most similar to our aging-related predictions, because our predictions are also partly based on brain human gene expression data.] However, it is important to note that BrainExpression2008Age and BrainExpression2004Age were predicted from two independent data sets, and compared with BrainExpression2004Age, BrainExpression2008Age is a result of a newer microarray study; it covers more samples and more individuals, and it covers more samples per individual. And this is exactly why we base our study (Section 2.1.1) on the expression data by Berchtold *et al.* (2008) rather than on the data by Lu *et al.* (2004).

By studying the brain gene expression data set related to different stages of Alzheimer's disease (AD), Simpson *et al.* (2011) identified 2911 genes that have significantly different expression levels at different stages of AD. Of these, 1541 are present in StatNetExpression. Henceforth, we denote this 'ground truth' AD-related set of 1541 genes predicted from *brain gene expression data* as *ADExpressionAge*.

In July 2012, GenAge contained 261 human genes that have been linked to aging as sequence-based orthologs of aging-related genes in model species (de Magalhães *et al.*, 2009). Of these, 242 are present in StatNetExpression. Henceforth, we denote this 'ground truth' aging-related set of 242 genes predicted from *sequence data* by *SequenceAge*.

2.1.5 Complements of the 'ground truth' aging-related data We define a set of genes as the *complement* of a 'ground truth' aging-related data set if the genes are present in StatNetExpression but not in the 'ground truth' data set. We denote the complements of BrainExpression2008Age, BrainExpression2004Age, ADExpressionAge and SequenceAge as *BrainExpression2008Complement*, *BrainExpression2004Complement*, *ADExpressionComplement* and *SequenceComplement*, respectively.

All above data sets are defined with respect to HPRD PPI data. For BioGRID data, see Supplementary Section S1.3.

2.2 Do global network topologies change with age?

Given the dynamic age-specific PPI networks, we test whether the overall (global) topologies of the networks change with age. We do so by comparing the different networks with respect to several commonly used global network properties (Section 2.2.1), by evaluating the fit of each of the age-specific networks to a series of well-known graph families, i.e. network models (Section 2.2.2) (Kuchaiev *et al.*, 2011; Milenković *et al.*, 2008), and by measuring the overlap of the age-specific networks (Section 2.2.3).

2.2.1 Comparing global properties of age-specific networks We analyze three properties: the average clustering coefficient, average diameter and graphlet frequency distribution (Memišević *et al.*, 2010a). The properties are defined in Supplementary Section S1.4.

2.2.2 Evaluating the fit of age-specific networks to different graph families or network models We compare the fit of the dynamic PPI networks to different graph families, i.e. network models (Milenković *et al.*, 2009), to test whether the best fitting model changes with age. Various network models have been proposed. We use (i) Erdös–Rényi random graphs (ER), (ii) generalized Erdös–Rényi random graphs with same degree distribution as the data (ERDD), (iii) geometric random graphs (GEO), (iv) geometric gene duplication and mutation model (GEOGD), (v) scale-free networks (SF) and (vi) scale-free gene duplication and mutation model (SFGD) (Kuchaiev *et al.*, 2011; Milenković *et al.*, 2008). To evaluate the fit of the data network to a given model, we compare the topology of the data network with the topology of a random network instance drawn from the model with respect to a highly constraining measure of network topological similarity called *graphlet degree distribution agreement* (Pržulj, 2007). For details, see Supplementary Section S1.5.

2.2.3 Computing the overlap between age-specific networks We measure the overlap between each pair of age-specific networks as the percentage of nodes (or edges) in the smaller of the two networks that are common to the two networks. For details, see Supplementary Section S1.6.

2.3 Do local topologies of proteins change with age?

We study topological positions of proteins in each age-specific network with respect to seven node centrality measures (Section 2.3.1). We predict to be aging-related those proteins whose centralities significantly change with age (Section 2.3.2). We validate our predictions in several ways (Section 2.3.3).

2.3.1 Local measures of topology or node centralities Various centrality measures have been used to link topological importance of a node in the network to its functional importance. Below, we define each of the seven measures that we use and provide biological justification for their use.

Degree centrality (DEGC) measures the degree of a node in the network, i.e. the number of the node's neighbors. The higher the degree of a node, the more central the node according to DEGC. Because current PPI networks have 'power-law' degree distributions, with many low-degree nodes and few high-degree nodes, and because the removal of the high-degree nodes would impact the network structure (by disconnecting it), DEGC of a gene has been related to the gene's essentiality as well as its involvement in disease (Barabási and Oltvai, 2004; Sharan and Ideker, 2008).

Clustering coefficient centrality (CLUSC) measures, for a given node, how many pairs of neighbors of the node are connected by an edge, out of all pairs of the node's neighbors. Intuitively, the more interconnected the neighborhood of the node, the more central the node is according to

CLUSC. In a PPI network, a node with high-clustering coefficient, together with the node's neighbors, forms a highly interconnected network region, which is likely to correspond to a functional module (Barabási and Oltvai, 2004).

K-core of a network is a maximal subset of nodes in the network such that each node is connected to at least k others in the subset. *K-coreness centrality* (KC) of a node is k if the node is in k-core. Nodes with high KC in the human network have been found to correspond to the 'core disease ome' that is significantly enriched in disease genes and drug targets (Janjić and Pržulj, 2012).

Graphlet degree centrality (GDC) measures how many graphlets a node participates in, for all 2–5-node graphlets (Milenković *et al.*, 2011). Intuitively, the more graphlets a node touches, the more central the node is according to GDC. Because it captures the *extended* network neighborhood of a node, GDC is a highly sensitive measure of network topology. Thus, in a PPI network, proteins with high GDCs represent potential candidates for therapeutic intervention, as targeting such proteins with drugs would have more significant impact on the network structure than targeting proteins that reside in sparse and non-complex network regions (Milenković *et al.*, 2011). Indeed, GDC has been found to capture well disease and pathogen-interacting proteins and drug targets (Milenković *et al.*, 2011).

Betweenness centrality (BETWC) measures the involvement of a node in the shortest paths in the network. Intuitively, nodes that occur in many shortest paths have high centrality according to BETWC. BETWC of node v, $C_{\text{betwc}}(v)$, is: $C_{\text{betwc}}(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where V is the set of nodes in the network, σ_{st} is the number of shortest paths between nodes s and t and $\sigma_{st}(v)$ is the number of shortest paths between s and t that go through v. In a PPI network, BETWC of a protein indicates the 'likelihood' of the protein to participate in pathways connecting all other proteins (Koschützki and Schreiber, 2008). Removal of a protein that is on critical pathways between many other proteins could cause loss of communication between the proteins. Also, targeting such a node with a drug could cause the drug effects to spread fast to all the nodes (Milenković *et al.*, 2011). This property has been used to identify gene–disease associations by encoding each gene in the network based on the distribution of shortest path lengths to all genes associated with disease (Radivojac *et al.*, 2008).

Closeness centrality (CLOSEC) measures the 'closeness' of a node to all other nodes in the network. Intuitively, nodes with small shortest path distances to all other nodes have high centrality according to CLOSEC. CLOSEC of node v, $C_{\text{closec}}(v)$, is: $C_{\text{closec}}(v) = \frac{1}{\sum_{u \in V} \sigma(u,v)}$, where $\sigma(u, v)$ is the shortest path distance between nodes u and v. In a PPI network, CLOSEC of a protein indicates the 'likelihood' of the protein to reach or be reachable from all other proteins (Scardoni *et al.*, 2009). And it has been a widely accepted assumption that proteins that are closer to each other are more likely to perform the same function (Sharan *et al.*, 2007).

Eccentricity centrality (ECC) is related to CLOSEC, except that it measures the 'closeness' of a node *only* to the *farthest* node in the network (Wuchty and Stadler, 2003). Intuitively, nodes with small shortest path distances to the furthest node in the network have high centrality according to ECC. ECC of node v, $C_{\text{ecc}}(v)$, is: $C_{\text{ecc}}(v) = \frac{1}{\max_{u \in V}\{\sigma(u,v)\}}$.

2.3.2 Prediction of aging-related genes For each measure, we compute centrality values for a node in each of the 37 age-specific networks. Then, we calculate Pearson or Spearman correlation between the 37 ages and the node's 37 centrality values (Supplementary Section S1.7). We do this for all genes that are active in >20% of the ages in the gene expression data (Lu *et al.*, 2004). If such a gene is unexpressed at a given age, we assign it a centrality value of zero at that age. We note that ignoring the centrality value of an unexpressed gene as opposed to treating it as a zero value makes no significant difference on our results. Also, because results are consistent for both correlation measures, here we report them only for Pearson correlation. Results for Spearman correlation are shown in Supplementary Fig. S4.

We quantify the statistical significance of a given correlation value observed from the data by measuring the probability (i.e. *P*-value) of observing by chance a better value (i.e. the same or higher value when the original value is positive or the same or lower value when the original value is negative). We do this by randomly reshuffling the 37 node centrality values at the 37 ages and by computing the resulting 'random correlation'. We repeat this 999 999 times to get 999 999 random correlations, which in addition to the data correlation totals to 1 000 000 correlation values (Phipson and Smyth, 2010). We compute the *P*-value as the percentage of the 1 000 000 values in which the random correlation (including the data correlation) is better than the data correlation. We perform multiple test correction of each *P*-value to control the false-discovery rate in our predictions (Benjamini and Hochberg, 1995). We predict a gene to be aging-related if its q-value (the adjusted *P*-value after multiple test correction) is <0.01.

Because we study multiple node centralities, each of which can predict the given gene to be aging-related, we score our predictions so that the more centrality measures support a prediction and the higher the significance of the change of its centrality values with age, the higher the score and the more credible the prediction. For details, see Supplementary Section S1.7.

2.3.3 Validation of predicted aging-related genes

Statistical significance of our predictions. To test whether our approach of combining static network data with aging-related expression data into the dynamic network data actually gives meaningful predictions, we study whether the number of aging-related genes that we predict from the actual data is statistically significantly larger than the number of aging-related genes that we predict from 'randomized data'. By 'randomized data', we mean that we randomize the expression data before integrating it with the static network data (Supplementary Section S1.8). Then, we integrate the randomized expression data with the static PPI network, construct randomized age-specific networks just as in Section 2.1.3 and predict aging-related genes from the randomized networks just as in Section 2.3.2. We repeat the above procedure multiple times, to assign a *P*-value to the number of predictions that we make from the actual data (Supplementary Section S1.8).

Overlap between genes of different data sets. We measure the statistical significance of the overlap between genes of two data sets by using the hypergeometric test, which computes probability *P* (i.e. *P*-value) of observing the same or larger overlap by chance as follows. Let E be the set of genes from StatNetExpression. Let A be the subset of genes from E that are in any one of the data sets. Let G be the subset of genes from E that are in the other data set. Let O be the overlap between A and G. Then: $p = 1 - \sum_{i=0}^{|O|-1} \frac{\binom{|E|}{i}\binom{|E|-|A|}{|G|-i}}{\binom{|E|}{|G|}}$ We use the *P*-value threshold of 0.05.

Gene Ontology enrichment. We study the enrichment of a data set in biological process Gene Ontology **(GO)** terms (Ashburner *et al.*, 2000). We use (i) all 5450 GO terms that annotate (independent on the evidence code) at least two genes from StatNetExpression and (ii) 2392 GO terms that annotate (with respect to an experimental evidence code only) at least two genes from StatNetExpression. For a GO term g, we compute the statistical significance of its enrichment via the above hypergeometric test formula, where now E is the set of genes from StatNetExpression that are annotated by any GO term, A is the gene set in which we are measuring GO term enrichment, G is the subset of genes from E that are annotated by GO term g and O is the set of genes in the overlap between A and G. We use the *P*-value threshold of 0.05.

GO term overlap. We measure the statistical significance of the overlap of GO terms enriched in one data set and GO terms enriched in another data set via the above hypergeometric test formula, where now E is the set of GO terms that annotate at least two genes from StatNetExpression, A

is the set of GO terms enriched in any one of the two data sets, G is the set of GO terms enriched in the other data set and O is the set of GO terms that are in the overlap between A and G. We use the P-value threshold of 0.05.

Disease Ontology enrichment. We study the enrichment of a data set in all 527 Disease Ontology (DO) terms that annotate at least two genes from StatNetExpression (Du *et al.*, 2009) in the same way as when we study GO term enrichments.

DO term overlap. We study the overlap of DO terms from different data sets in the same way as when we study GO term overlaps.

Literature validation. We *automatically* search for a gene in PubMed (http://www.pubmed.gov) and consider the gene to be validated in the context of aging if its name is mentioned (according to NCBI's E-utilities – http://www.ncbi.nlm.nih.gov/books/NBK25500) with 'age', 'aging' or 'ageing' in the title or abstract of at least one article. Also, we *manually* search for a gene by reading relevant PubMed articles *more closely*.

3 RESULTS AND DISCUSSIONS

We study *global* topologies of the age-specific networks in Section 3.1. We study *local* topologies of proteins in each network and predict aging-related genes in Sections 3.2.1 and 3.2.2. We validate our predictions in Section 3.2.3.

3.1 Global network topologies do not change with age

3.1.1 Global properties of age-specific networks are similar Average clustering coefficients, average diameters and graphlet frequency distributions (Section 2.2.1) of the age-specific networks do not significantly change with age (Supplementary Figs. S1 and S2).

3.1.2 Networks at different ages belong to the same graph family We compare the fit of the age-specific networks to six network models (Section 2.2.2). The best-fitting model does not change with age (Supplementary Figs. S1 and S2). Note that our primary goal is not to identify the best-fitting model for dynamic PPI networks. Nonetheless, consistent with results for static PPI networks (Kuchaiev *et al.*, 2011; Pržulj *et al.*, 2010; Ratmann *et al.*, 2009), it is gene duplication models that fit the age-specific networks the best.

3.1.3 Overlap of age-specific networks is large The age-specific networks share *on average* 92% of the nodes and 89% of the edges, depending on age, whereas every pair of the networks shares *at least* 82% of the nodes and 74% of the edges (Supplementary Fig. S3). Hence, the network overlaps are quite large.

3.2 Local topologies of proteins do change with age

3.2.1 Prediction of aging-related genes Depending on a study, gene expression data alone can reveal as few as 442 (Lu *et al.*, 2004) or as many as 8277 (Berchtold *et al.*, 2008) of the human genes as aging-related. We integrate the gene expression data (Berchtold *et al.*, 2008) with the human PPI network, forming dynamic age-specific networks, to investigate whether data integration can contribute to our understanding of aging. While global network analysis has failed to uncover any aging-related information, it could be that the dynamic network data encode aging-related information only locally and around only a subset of nodes. Therefore, we use centrality measures (Section 2.3.1) to quantify local positions of nodes in the age-specific networks and find nodes whose centralities correlate well with age; as such proteins could be key players in aging.

We predict a gene to be aging-related if its centrality values are statistically significantly correlated with age (Section 2.3.2) for at least one centrality measure. This results in 537 (5.7%) predictions out of all 9426 genes. Fig. 2a shows the number of aging-related predictions for each centrality individually and all centralities combined. Some measures, and GDC in particular, are superior to others in terms of the number of their aging-related predictions. Importantly, GDC is a sensitive measure of network topology, as it captures the *extended* network neighborhood around a node. As such, its superiority is not necesarily surprising (Milenković *et al.*, 2011).

A gene's centrality can be positively correlated with age (the gene becomes more network-central with age) or it can be negatively correlated with age (the gene becomes less network-central with age). Interestingly, even though it has already been argued that aging is associated with failure of 'hubs', highly interconnected and thus network-central proteins (Soltow *et al.*, 2010), the majority of our predictions are positively correlated with age (Fig. 2).

3.2.2 Relationships and potential redundancies of different node centralities We predict a gene to be aging-related if its centrality values correlate well with age with respect to *at least one* centrality. So, we study whether any genes are predicted by more than one or even all of the centralities. We find that almost half (47%) of the 537 aging-related predictions are supported by multiple centralities, while the remaining predictions are supported by a single centrality (Fig. 2b). As expected, the number of predictions decreases as the number of centralities supporting the predictions increases.

We study redundancies of the different centralities by computing, for each pair of centralities, the correlation between their centrality values over all nodes in a network, and by averaging correlations over the 37 age-specific networks. We observe high correlations between some measures, such as BETWC, DEGC, KC and GDC or CLOSEC and ECC (Fig. 3a and Supplementary Section S2.1). Thus, some centralities appear to be redundant to others.

When we study pairwise overlaps of aging-related predictions produced by the different centralities, we still observe high overlaps between some (though not all) centralities (Fig. 3b). For example, each prediction made by ECC is supported by at least one other centrality measure. The overlaps are encouraging, as they increase the credibility of predictions supported by multiple centrality measures.

Yet, given that the majority (~53%) of the predictions are identified by a single centrality (Fig. 2b), it seems that not all centralities are redundant to each other. And when we focus on our aging-related genes predicted by exactly one centrality, in most cases, these predictions are not even marginally significant with respect to other centralities (Fig. 3c and Supplementary Fig. S6), indicating again that the different centralities are in general not redundant to each other. Thus, we keep all 537 predictions, independent on the number of centralities supporting them. Henceforth, we denote this set of 537 aging-related genes

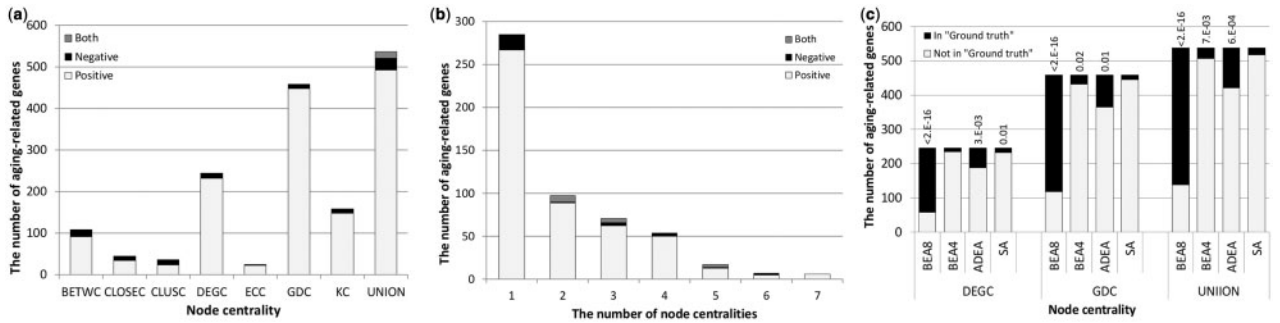

Fig. 2. The number of our predicted aging-related genes. Panel (a) shows the number of predictions for each of seven node centralities individually (BETWC, CLOSEC, CLUSC, DEGC, ECC, GDC and KC) or by at least one of them (UNION). Panel (b) shows the number of genes predicted by exactly k node centralities ($k = 1, 2, \ldots, 7$). In the panels, white and black bars show the number of genes that are positively and negatively correlated with age, respectively. Gray bars denote the number of genes for which one centrality measure identifies the given gene as positively correlated with age, while another measure identifies the same gene as negatively correlated with age. (Panel (c) shows the overlaps of our predictions with aging-related genes from the four 'ground truth' data sets [BrainExpression2008Age (BEA8), BrainExpression2004Age (BEA4), ADExpressionAge (ADEA) and SequenceAge (SA)], for each of the two best-performing centralities from Panel (a) (DEGC and GDC) individually, as well as for all seven centralities combined (UNION). The same results for the remaining five centralities individually are shown in the Supplement. In the panel, white and black bars show the number of predicted genes that are absent from and present in the 'ground truth' data, respectively. The P-values for statistically significant overlaps are noted at the top of the bars. The results in the figure are consistent when we use Spearman correlation to predict aging-related genes instead of Pearson correlation (Section 2.3.2 and Supplementary Fig. S4). Also, the results are consistent when we use BioGRID data as the static PPI network instead of HPRD data (Section 2.1.3 and Supplementary Fig. S5)

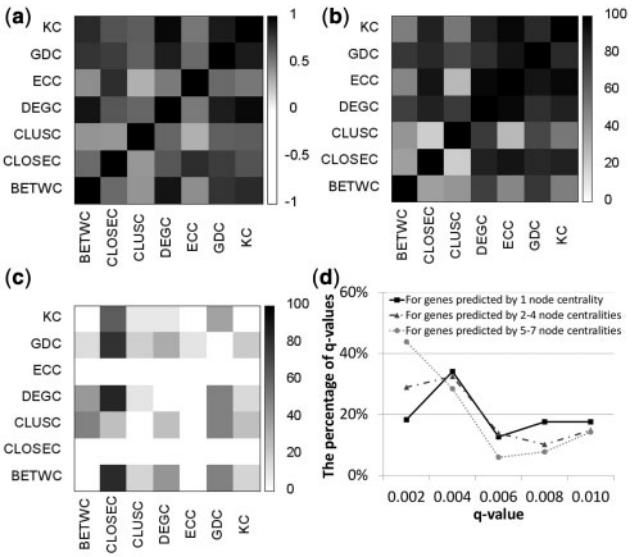

Fig. 3. Relationships between different centralities (BETWC, CLOSEC, CLUSC, DEGC, ECC, GDC and KC): (**a**) Spearman correlation between each pair of centralities averaged over the 37 age-specific networks; (**b**) pairwise overlap of aging-related genes predicted by the different centralities; (**c**) percentage of genes predicted as aging-related (q-value ≤ 0.01; q-value is the adjusted P-value after multiple test correction) by exactly one centrality (listed in a row), which have a 'marginal' q-value between 0.01 and 0.05 with respect to one of the six remaining centralities (listed in a column) and (**d**) distributions of q-values of our predicted aging-related genes. Panel (c) can be interpreted as follows. For example, many predictions made by BETWC (the last row) are marginally significant with respect to CLOSEC, DEGC and GDC (intensive color), whereas very few of the predictions made by ECC (the third row) are marginally significant with respect to any other centrality (light color). Hence, the predictions made by BETWC only may be more credible than the predictions made by ECC only, as the former are marginally supported by additional centralities, whereas the latter are not

predicted via dynamic network analysis as *DyNetAge*. We denote the complement of this set, i.e. the set of genes that are present in StatNetExpression but not in DyNetAge, as *DyNetComplement*. Clearly, DyNetComplement is the set of genes whose network centralities do not significantly correlate with age, and as such, we do not predict them to be aging-related.

Next, we study the effect of the number of centralities supporting a prediction on the prediction's quality. It is not necessarily the case that predictions supported by many centralities are more enriched in 'ground truth' aging-related genes (see below) compared with predictions supported by only few centralities (Supplementary Fig. S7). However, it *is* the case that the more centralities support a prediction, the more significant the correlation of its centrality values with age (Fig. 3d). Hence, to account for the number of centralities supporting a prediction, we rank the prediction so that the more centralities support it and the more significant the change of its centrality values with age, the more credible the prediction (Section 2.3.2). (Supplementary Tables S1 and S2 rank the predictions.)

3.2.3 Validation of predicted aging-related genes

Dynamic network analysis gives meaningful and statistically significant aging-related predictions. The number of aging-related genes in DyNetAge is statistically significantly larger than the number of aging-related genes that we can predict from the 'randomized' data (P-value of 5.3×10^{-9}; Section 2.3.3).

Overlap of our predictions with 'ground truth' aging-related genes is significant. Human aging is hard to study experimentally due to long life span and ethical constraints. Instead, human aging-related 'ground truth' knowledge has been *predicted* more or less *computationally*, by studying expression data or by transferring aging-related knowledge from model species via sequence comparison. (We intentionally use quotes, as we are *not* dealing with experimental ground truth data.) Similarly, here, we aim to predict new 'ground truth' data from an additional data

type – PPIs. (But, by no means do we claim to identify *all* aging-related genes.)

Because 'ground truth' data are predicted computationally, they could be noisy. Also, different 'ground truth' sets could be biased toward different data types from which the predictions have been made, be it expression, sequence or PPI data. Because different data types could be capturing different functional slices of the cell (Memišević *et al*., 2010b; Przytycka and Kim, 2010), it might not be alarming if overlaps between the different 'ground truth' sets are not large. However, because all the sets aim to capture the same biological phenomenon (aging), some overlap should exist. Ideally, one would hope for a statistically significant overlap. However, the existence of any (even non-significant) overlap would be encouraging, because (i) the overlap could be low due to the noisiness of each 'ground truth' set and (ii) statistically non-significant results may be biologically important, whereas statistically significant results may not be (Ho *et al*., 2010; Milenković *et al*., 2010).

We measure the overlap of DyNetAge with four 'ground truth' data sets: BrainExpression2008Age, BrainExpression2004Age, ADExpressionAge and SequenceAge (Section 2.1.4). Of the four, BrainExpression2008Age and BrainExpression2004Age are the most likely to be similar to DyNetAge (as all three are based on *brain-related* data, brain-related *gene expression* data and brain *aging-related* gene expression data; Supplementary Section S2.2), followed by ADExpressionAge (as our predictions as well as these data are both based on *brain-related* data and brain-related *gene expression* data), followed by SequenceAge (as our predictions and these data both capture aging-related information but from different data types) (Section 2.1.4). Therefore, a high overlap of DyNetAge with BrainExpression2008Age or BrainExpression2004Age would validate our method. A high overlap with ADExpressionAge would suggest that our method could capture not only brain aging-related genes but also brain aging-related *disease* genes. A high overlap with SequenceAge would suggest that our method could capture genes identified from a different data type, namely, sequence data.

And this is exactly what we observe in general (Supplementary Fig. S8 and Supplementary Table S4). DyNetAge's overlap is statistically significant for BrainExpression2008Age, BrainExpression2004Age and ADExpressionAge, and marginally significant for SequenceAge (Table 1). The (marginally) significant overlap between DyNetAge and all four 'ground truth' data sets is encouraging. Importantly, even though most of the overlaps are significant, 26%, 94%, 78% and 96% of our DyNetAge predictions are not in BrainExpression2008Age, BrainExpression2004Age, ADExpressionAge and SequenceAge, respectively, and 20% of our predictions are not in *any* of the four data sets (Supplementary Table S5). This confirms that data integration can reveal *additional* knowledge compared with studying individual data types.

Recall that the complement of each 'ground truth' data set (including DyNetAge) is the set of genes not predicted as aging-related by the given study (Section 2.1.5). Hence, it would be encouraging to see (i) low (non-significant) overlaps between DyNetAge and complements of the 'ground truth' data sets, (ii) low (non-significant) overlaps between DyNetComplement and 'ground truth' sets and (iii) high (significant) overlaps between DyNetComplement and complements of 'ground truth' sets. This is what we typically observe in all three cases (Supplementary Table S4).

Table 1. Overlap and its statistical significance (*P*-value) between genes, GO terms and DO terms in DyNetAge and those in the 'ground truth' aging-related sets [BrainExpression2008Age (BE8A), BrainExpression2004Age (BE4A), ADExpressionAge (ADEA) and SequenceAge (SA)]

		BE8A	BE4A	ADEA	SA
Genes	Overlap	74%	9%	22%	8%
	P-value	<2.20E-16	7.10E-03	6.00E-04	0.10
GO terms	Overlap	24%	13%	6%	36%
	P-value	<2.20E-16	3.35E-06	0.15	4.44E-16
DO terms	Overlap	8%	8%	0%	33%
	P-value	N/A	N/A	N/A	0.43

GO enrichment. Two-hundred thirty-three GO terms are significantly enriched in DyNetAge (*P*-values between 0.049 and 5×10^{-6}), whereas only 9 GO terms are enriched in DyNetComplement (*P*-values between 0.043 and 6.5×10^{-4}) (Section 2.3.3). Of *aging-related* GO terms, it is encouraging that DyNetAge is marginally significantly enriched in *cell aging* (*P*-value of 0.052) as well as (though not significantly) in *aging*, *multicellular organismal aging* and *cellular senescence*. Importantly, DyNetAge's performance in this context is typically comparable with that of the 'ground truth' data, especially when using only gene-GO term associations obtained by *experimental* evidence codes (Supplementary Section S2.3).

GO overlap. Our predictions are further validated by (i) high GO term overlaps between DyNetAge and the 'ground truth' data sets, (ii) low GO term overlaps between DyNetAge and complements of the 'ground truth' sets, (iii) low GO term overlaps between DyNetComplement and the 'ground truth' sets and (iv) high GO term overlaps between DyNetComplement and complements of the 'ground truth' sets (Table 1 and Supplementary Table S6). For example, when considering gene-GO term associations of any evidence code, GO terms from DyNetAge significantly overlap with GO terms from three of the four 'ground truth' sets (BrainExpression2008Age, BrainExpression2004Age and SequenceAge), and some (though non-significant) overlap with GO terms from ADExpressionAge is also encouraging (Table 1). Results when considering gene-GO term associations of *experimental* evidence codes are shown in Supplementary Table S7. Importantly, GO term overlap is significant between DyNetAge and BrainExpressionAge2008, whereas it is non-significant between SequenceAge and BrainExpressionAge2008, indicating the superiority of DyNetAge over SequenceAge in this context (Supplementary Section S2.4).

DO enrichment. Twelve diseases are significantly enriched in DyNetAge (*P*-values between 0.048 and 1.5×10^{-3}), whereas no disease is enriched in DyNetComplement (Section 2.3.3). The 12 diseases in DyNetAge are *pemphigoid bullous*, *lupus erythematosus*, *infection by Cryptococcus neoformans*, *capillaries disease*, *Wiskott–Aldrich syndrome*, *Kuhnt–Junius degeneration*, *carcinoma*, *Rett syndrome*, *skin cancer*, *infection*, *myeloproliferative disease* and *IgA glomerulonephritis*. Kuhnt–Junius degeneration,

lupus erythematosus, myeloproliferative disease and skin cancer are not only enriched in SequenceAge as well but these disease have also been explicitly linked to aging in the literature [PubMed IDs (PMIDs): 11587915, 11318593, 24138668 and 17942417, respectively]. Pemphigoid bullous is not only enriched in BrainExpression2008Age as well but this disease also shows increased risk in old patients (PMID: 10426901). Rett syndrome is not only enriched in BrainExpression2004Age as well but this disease also shows aging-related changes in physiological processes of patients (PMID: 11247011). Supplementary Table S8 maps the PMIDs to full-paper references. Importantly, even five of the six diseases from DyNetAge that are missed by *all* 'ground truth' sets can *all* be linked to aging in the literature as well (Supplementary Section S2.5).

DO overlap. We further validate our predictions by demonstrating (i) high DO term overlaps between DyNetAge and the 'ground truth' data sets, (ii) low DO term overlaps between DyNetAge and complements of the 'ground truth' sets, (iii) low DO term overlaps between DyNetComplement and the 'ground truth' sets and (iv) high DO term overlaps between DyNetComplement and complements of the 'ground truth' sets (Supplementary Table S9). It is encouraging that DO terms from DyNetAge overlap with DO terms from BrainExpressionAge2008, BrainExpressionAge2004 and SequenceAge (though the overlaps are non-significant) (Table 1). Importantly, diseases from DyNetAge do not significantly overlap with diseases from complements of the 'ground truth' sets, whereas diseases from SequenceAge *do* significantly overlap with diseases from both BrainExpression2008Age and ADExpressionAge. This indicates superiority of DyNetAge over SequenceAge in this context. For details, see Supplementary Section S2.6.

Literature validation. Automatic literature validation (Section 2.3.3) is prone to errors: we 'validate' in this manner equal portion of *each* 'ground truth' aging-related set *and* its complement. Therefore, we aim to validate our predictions manually (Section 2.3.3). Because manual validation is laborious, we focus on our 10% highest-scoring predictions (Section 2.3.2). Of these, we study all eight predictions that are absent from any 'ground truth' set, namely, CD22, CTRL, HMG20B, TRIM21, UBA52, USP45, OXSR1 and POLE2. Importantly, we successfully validate *all* these genes as follows. An aging-associated decline has been found in B cell precursors of bone marrow, whereas CD22 is one of the markers of B cell precursors (PMID: 19967915). An aging-dependent degradation has been found in CTRL that affects chymotrypsin-like activity in the retina (PMID: 17258201). HMG20B is regulated by the activity of BRCA2 tumor suppressor, whereas cancers and tumors are known to be aging-associated diseases (PMID: 21399666). TRIM21 is involved in a pathway that acts as a potential driver of lung tumorigenesis, whereas cancers and tumors are known to be aging-associated diseases (PMID: 21399661). UBA52 has been found to be significantly downregulated with respect to aging-associated degenerative diseases of the eye (PMID: 23601964). Aging-related changes have been found in the ubiquitin–proteasome system, whereas USP45 belongs to the ubiquitin-specific protease gene family (PMID: 12944592). It has been found that oxidative stress responsive transcription factor pathway is activated during aging in cardiac muscle (PMID: 9011632), whereas OXSR1 performs oxidative stress responsive function. Finally, mutations in POLE2 have been found to be linked to human colorectal cancer, whereas cancers and tumors are known to be aging-associated diseases (PMID: 20065316).

As a negative control for literature search, we also aim to validate genes from DyNetComplement, in hope that their validation rate will be lower than the validation rate of the above eight highest-scoring predictions. For a fair evaluation, we arbitrarily select eight lowest-scoring genes from DyNetComplement, and we perform manual literature validation on these genes in the same manner as for the above eight highest-scoring predictions. As we had hoped, we find only one of them to be aging-related, which further increases confidence in our aging-related predictions from DyNetAge. For details, see Supplementary Section S2.7.

3.3 Robustness to the choice of gene expression data

To validate the robustness of our approach, we also base our predictions on an alternative gene expression data obtained by RNA-seq instead of microarray technology (Section 2.1.1), in the hope that the resulting predictions (which we denote by DyNetAgeRNASeq) will significantly overlap with the 'ground truth' data (including DyNetAge). Indeed, 172 predictions from DyNetAgeRNASeq (Supplementary Table S3) significantly overlap with BrainExpressionAge2008 (P-value of 1.7×10^{-3}), BrainExpressionAge2004 (P-value of 0.01) and ADExpressionAge (P-value of 3.9×10^{-3}). Also, they overlap (though not significantly) with SequenceAge and DyNetAge. Existence of any (even non-significant) overlap is still encouraging (Section 3.2.3). Thus, our approach results in meaningful predictions independent on the choice of the data.

4 CONCLUSION

Together, our results confirm that dynamic PPI network analysis via integration of static PPI network data with aging-related gene expression data can reveal meaningful key players in aging.

Funding: This work was supported by the National Science Foundation (NSF) grants CCF-1319469 and EAGER CCF-1243295.

Conflict of Interest: none declared.

REFERENCES

Aragues,R. *et al.* (2008) Predicting cancer involvement of genes from heterogeneous data. *BMC Bioinformatics*, **9**, 172.

Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.

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

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.

Berchtold,N.C. *et al.* (2008) Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl Acad. Sci. USA*, **105**, 15605–10.

Breitkreutz,B.J. *et al.* (2008) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.

de Magalhães,J. (2009a) Aging research in the post-genome era: new technologies for an old problem. In: Foyer,C. *et al.* (ed.) *Redox Metabolism and Longevity Relationships in Animals and Plants*. Taylor and Francis, New York, pp. 99–115.

de Magalhães,J. *et al.* (2009b) The Human Ageing Genomic Resources: online databases and tools for biogerontologists. *Aging Cell*, **8**, 65–72.

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–i68.
Dyer,M. *et al.* (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.*, **4**, e32.
Ferrarini,L. *et al.* (2005) A more efficient search strategy for aging genes based on connectivity. *Bioinformatics*, **21**, 338–348.
Fortney,K. *et al.* (2010) Inferring the functions of longevity genes with modular subnetwork biomarkers of *Caenorhabditis elegans* aging. *Genome Biol.*, **11**, R13.
Ho,H. *et al.* (2010) Protein interaction network uncovers melanogenesis regulatory network components within functional genomics datasets. *BMC Syst. Biol.*, **4**, 84.
Janjić,V. and Pržulj,N. (2012) The core diseasome. *Mol. Biosyst.*, **8**, 2614–2625.
Jeong,H. *et al.* (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.
Jonsson,P.F. and Bates,P.A. (2006) Lobal topological features of cancer proteins in the human interactome. *Bioinformatics*, **22**, 2291–2297.
Koschützki,D. and Schreiber,F. (2008) Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul. Syst. Biol.*, **2**, 193–201.
Kriete,A. *et al.* (2011) Computational systems biology of aging. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **3**, 414–428.
Kuchaiev,O. *et al.* (2011) GraphCrunch 2: software tool for network modeling, alignment and clustering. *BMC Bioinformatics*, **12**, 24.
Lee,M.L. *et al.* (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cdna hybridizations. *Proc. Natl Acad. Sci. USA*, **97**, 9834–9839.
Lu,T. *et al.* (2004) Gene regulation and DNA damage in the ageing human brain. *Nature*, **429**, 883–891.
Mazin,P. *et al.* (2013) Widespread splicing changes in human brain development and aging. *Mol. Syst. Biol.*, **9**, 633.
Memišević,V. *et al.* (2010a) An integrative approach to modeling biological networks. *J. Integr. Bioinform.*, **7**, 120.
Memišević,V. *et al.* (2010b) Complementarity of network and sequence information in homologous proteins. *J. Integ. Bioinform.*, **7**, 135.
Milenković,T. and Pržulj,N. (2008) Uncovering biological network function via graphlet degree signatures. *Cancer Inform.*, **6**, 257–273.
Milenković,T. *et al.* (2008) GraphCrunch: a tool for large network analyses. *BMC Bioinformatics*, **9**, 70.
Milenković,T. *et al.* (2009) Optimized null model for protein structure networks. *PLoS One*, **4**, e5967.
Milenković,T. *et al.* (2010) Systems-level cancer gene identification from protein interaction network topology applied to melanogenesis-related interaction networks. *J. R. Soc. Interface*, **7**, 423–437.
Milenković,T. *et al.* (2011) Dominating biological networks. *PLoS One*, **6**, e23016.
Peri,S. *et al.* (2004) Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.*, **32**, D497–D501.
Phipson,B. and Smyth,G.K. (2010) Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.*, **9**. Article39.
Promislow,D. (2004) Protein networks, pleiotropy and the evolution of senescence. *Proc. R. Soc. B Biol. Sci.*, **271**, 1225–1234.
Pržulj,N. (2007) Biological network comparison using graphlet degree distribution. *Bioinformatics*, **23**, e177–e183.
Pržulj,N. (2011) Protein-protein interactions: making sense of networks via graph-theoretic modeling. *Bioessays*, **33**, 115–123.
Pržulj,N. *et al.* (2010) Geometric evolutionary dynamics of protein interaction networks. *Pac. Symp. Biocomput.*, 178–189.
Przytycka,T. and Kim,Y. (2010) Network integration meets network dynamics. *BMC Biol.*, **8**, 48.
Radivojac,P. *et al.* (2008) An integrated approach to inferring gene-disease associations in humans. *Proteins*, **72**, 1030–1037.
Ratmann,O. *et al.* (2009) From evidence to inference: probing the evolution of protein interaction networks. *HFSP J.*, **3**, 290–306.
Reja,R. *et al.* (2009) MitoInteractome: mitochondrial protein interactome database, and its application in'aging network' analysis. *BMC Genomics*, **10** (**Suppl. 3**), S20.
Scardoni,G., Petterlini,M. and Laudanna,C. (2009) Analyzing biological network parameters with centiscape. *Bioinformatics*, **25**, 2857–2859.
Sharan,R. and Ideker,T. (2008) Protein networks in disease. *Genome Res.*, **18**, 644–652.
Sharan,R. *et al.* (2007) Network-based prediction of protein function. *Mol. Syst. Biol.*, **3**, 1–13.
Sőti,C. and Csermely,P. (2007) Aging cellular networks: chaperones as major participants. *Exp. Gerontol.*, **42**, 113–119.
Simpson,J.E. *et al.* (2011) Microarray analysis of the astrocyte transcriptome in the aging brain: relationship to Alzheimer's pathology and apoe genotype. *Neurobiol. Aging*, **32**, 1795–1807.
Soltow,Q. *et al.* (2010) A network perspective on metabolism and aging. *Integr. Comp. Biol.*, **50**, 844–854.
Vanunu,O. *et al.* (2010) Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.*, **6**, e1000641.
Wieser,D. *et al.* (2011) Computational biology for ageing. *Phil. Trans. R. Soc. B Biol. Sci.*, **366**, 51–63.
Wuchty,S. and Stadler,P.F. (2003) Centers of complex networks. *J. Theor. Biol.*, **223**, 45–53.