

RESEARCH

Open Access



Screening lifespan-extending drugs in *Caenorhabditis elegans* via label propagation on drug-protein networks

Hui Liu^{1,2}, Mengmeng Guo¹, Ting Xue¹, Jihong Guan³, Libo Luo^{2*} and Ziheng Zhuang^{1,2*}

From The 27th International Conference on Genome Informatics
Shanghai, China.3-5 October 2016

Abstract

Background: One of the most challenging tasks in the exploration of anti-aging is to discover drugs that can promote longevity and delay the incidence of age-associated diseases of human. Up to date, a number of drugs, including some antioxidants, metabolites and synthetic compounds, have been found to effectively delay the aging of nematodes and insects.

Results: We proposed a label propagation algorithm on drug-protein network to infer drugs that can extend the lifespan of *C. elegans*. We collected a set of drugs of which functions on lifespan extension of *C. elegans* have been reliably determined, and then built a large-scale drug-protein network by collecting a set of high-confidence drug-protein interactions. A label propagation algorithm was run on the drug-protein bipartite network to predict new drugs with lifespan-extending effect on *C. elegans*. We calibrated the performance of the proposed method by conducting performance comparison with two classical models, kNN and SVM. We also showed that the screened drugs significantly mediate in the aging-related pathways, and have higher chemical similarities to the effective drugs than ineffective drugs in promoting longevity of *C. elegans*. Moreover, we carried out wet-lab experiments to verify a screened drug, 2-Bromo-4'-nitroacetophenone, and found that it can effectively extend the lifespan of *C. elegans*. These results showed that our method is effective in screening lifespan-extending drugs in *C. elegans*.

Conclusions: In this paper, we proposed a semi-supervised algorithm to predict drugs with lifespan-extending effects on *C. elegans*. *In silico* empirical evaluations and *in vivo* experiments in *C. elegans* have demonstrated that our method can effectively narrow down the scope of candidate drugs needed to be verified by wet lab experiments.

Keywords: Longevity, Label propagation, Drug targets, *Caenorhabditis elegans*

Background

Aging is a physiological process in accompany with continuous accumulation of damages to cells and organs, which gradually lead to loss of normal organ functions and rise vulnerability to disease [1]. With the increasing average age of people, the risk of age-related diseases also concomitantly increases. Therefore, there is a rising interest in exploring drugs to promote healthy longevity, which

basically aim at preventing or delaying the onset of age-associated illnesses, such as cardiovascular disease, type 2 diabetes, neurodegenerative disease and cancer [2, 3].

It has been shown that restriction of energy intake can effectively extend lifespan of diverse species from yeast and worms to mammals [4, 5]. Dietary restriction can also protect against age-related risk factors of diabetes, cardiovascular disease, and cancer in human [6]. Previous studies have verified that the beneficial effects of dietary restriction in mammals are primarily obtained by increasing insulin sensitivity and decreasing blood glucose in mammals [7, 8]. Also, reduced activity of

*Correspondence: @fudan.edu.cn; llb213001@163.com; cczuzh@163.com

¹Changzhou University, Jiangsu 213164, China

²Changzhou NO. 7 People's Hospital, Changzhou, Jiangsu 213011, China

Full list of author information is available at the end of the article

nutrient-sensing pathways, including IIS (insulin/insulin-like growth factor), AMPK and mTOR signaling pathways, can slow down the aging of yeast, worms, flies and mice [9–12], suggesting that inhibition of these signaling pathways may induce a physiological state similar to that resulting from food shortage. Moreover, the lifespan-control mechanisms are remarkably conserved across diverse species [13].

Up to date, a number of drugs, including some antioxidants, metabolites and synthetic compounds, have been found to effectively delay the aging of nematodes and insects [14–16]. Some natural products and extract from plants can also extend the lifespan of invertebrates [17–20]. In particular, some drugs approved for human disease can promote the longevity of worm and mice, including Rapamycin [14], Aspirin [21], Metformin [22] and Resveratrol [23]. These findings imply there exist some common physiological mechanism between aging and the diseases treated by these drugs [24].

Some public data resources have been released for exploring aging [25–27]. For instance, GenAge, the benchmark database of genes related to ageing, has collected the genes associated with changes in the ageing phenotype or longevity in human and four model organisms [26]. All gene entries in GenAge are compiled from experimentally validated results published in peer-reviewed scientific literature. A gene is considered for inclusion if genetic manipulations (including knockout, mutations, overexpression or RNA interference) result in noticeable changes in the ageing phenotype and/or lifespan. NetAge is another aging-related web resource, which provides access to gene, protein and miRNA interaction networks that are involved in complex processes of aging and age-related diseases [27]. Gene Aging Nexus (GAN) have collected together numbers of ageing-related microarray gene expression data from human, rat, mouse, fruitfly, worm and yeast studies [28]. These public databases provide a valuable resource for us to develop computational methods for screening lifespan-extending drugs.

It is worth noting that the model organism *Caenorhabditis elegans* (*C. elegans*) has advantageous features for aging exploration, including its short lifespan, stereotypical development and small size [29]. These features make it a popular model species to conduct whole-organism assessment of anti-aging effect and mechanism of action of drugs [30, 31]. Thanks to the evolutionarily conserved mechanism of lifespan control from worm to mice and human, potential anti-aging drugs can be tested on worms and then transferred to mammals. On the other hand, drug efficacy is primarily exerted through inhibiting (or activating) the functions of target proteins by drug molecules, which specifically bind to the protein functional domains so that the corresponding biological functions are desirably blocked (or catalyzed) [32].

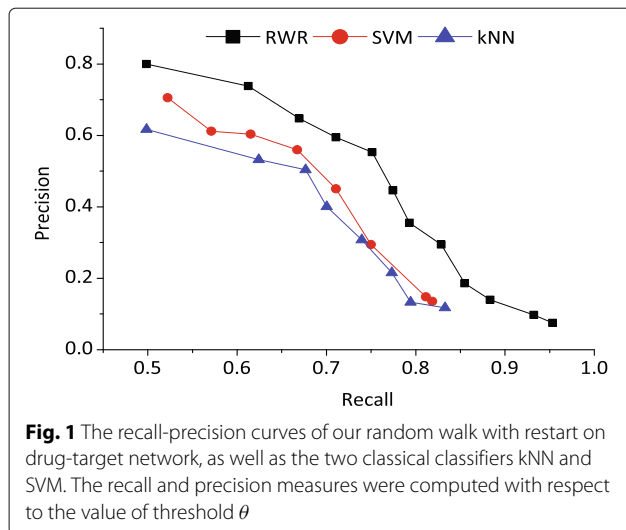
As a result, drug-protein interactions have been intensively studied and exploited in the drug research and development [33–35].

In this paper, we proposed a semi-supervised learning algorithm to predict drugs with lifespan-extending effects on *C.elegans*. We built a set of drugs and genes of which (effective or ineffective) influence on lifespan of *C. elegans* has been experimentally determined. The drug set included 1,309 drugs, which were collected from a large-scale bioassay screening for anti-aging drugs [36], and manually curated anti-aging drugs from literature [31]. The gene set included 681 genes, which were collected from the aging-related benchmark database GenAge [26]. By extracting a set of high-confidence drug-protein interactions from STITCH [37], we built a drug-protein bipartite network and then run a label propagation algorithm to predict new effective drugs. To calibrate the performance of the proposed method, we conducted 5-fold cross-validations on the gold-standard set of drugs. The empirical experiments showed that our method achieves higher performance than two classical models, kNN and SVM. Moreover, our screened drugs significantly mediate in the aging-related pathways, and have higher chemical similarities to the effective drugs than ineffective drugs in promoting longevity in *C. elegans*. Finally, we carried out wet-lab experiments to test the effectiveness of one screened drugs, 2-Bromo-4'-nitroacetophenone (PubChem CID000066840), and found that it can significantly promote the longevity of *C.elegans*. Both the *in silico* and *in vivo* experimental results demonstrated the performance of our computational approach for screening anti-aging drugs.

Results

Performance evaluation by 5-fold cross-validations

We conducted 5-fold cross-validations on the gold-standard set of drugs (see Data resources for details about the dataset) to evaluate the performance of our method. The drugs in the gold standard set were randomly split into 5 subsets with roughly equal size, and then each subset was taken in turn as a test set and the remaining four subsets were taken as input to run our method. The prediction accuracy was evaluated on the test set, and the averages over the 5-fold test subsets are regarded as final performance measures. Based on the predicted scores by our method and two classical classifiers, kNN and SVM, we computed the precision and recall measures for a given threshold, i.e. the drugs with scores greater than the threshold were classified as positive (effective) ones, and negative (ineffective) ones otherwise. We built the recall-precision curves for the three methods by gradually increasing the threshold, as shown in Fig. 1. It can be found that our method significantly outperformed the two classical classifiers kNN and SVM.

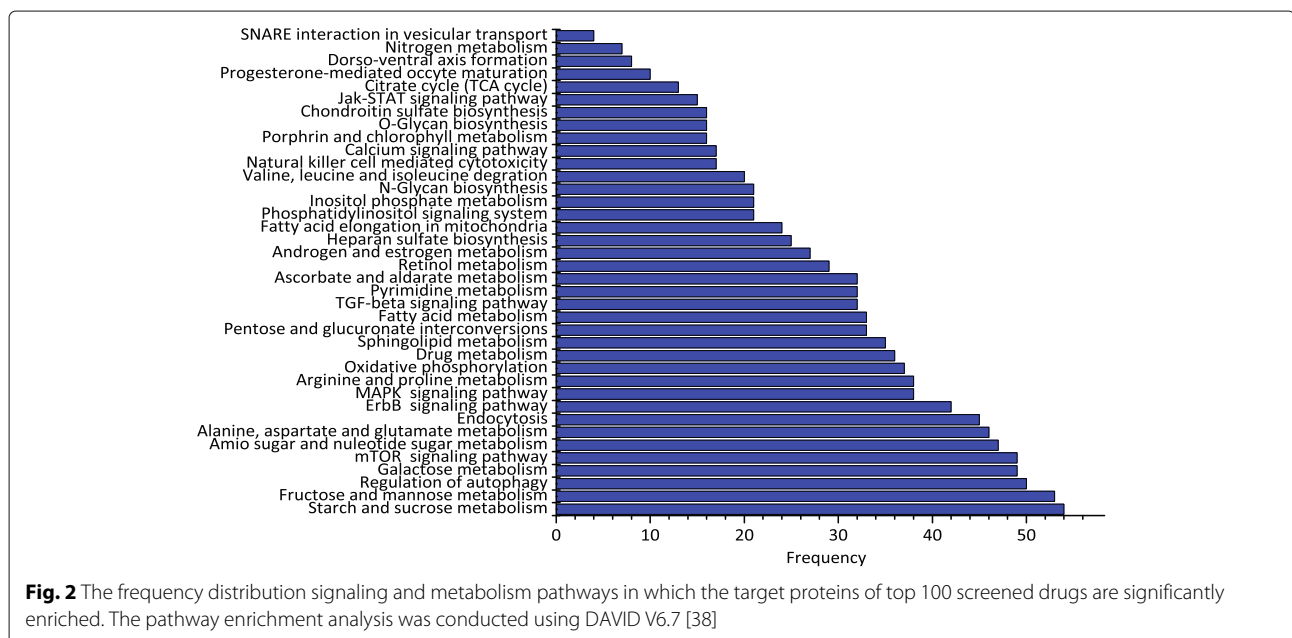


Highly scored drugs mediate in aging-related pathways

We carried out pathway analysis based on the target proteins of 100 top-rank drugs predicted by our method (for whole list of the predicted drugs see Additional file 1). The target proteins of each drug are obtained from STITCH database, and the pathway analysis were conducted by using DAVID V6.7 [38]. We found that screened drugs primarily mediate in sugar catabolism, energy metabolism and other processes related to cellular detoxification, as shown in Fig. 2. The results are consistent with several previous studies that have shown that these pathways mediate in the aging processes [39–41]. For example, McElwee et al. analyzed the differentially expressed genes during the aging processes

of mice, fly and worm, and identified a group of evolutionarily conserved biological processes related to aging, including sugar and energy catabolism [40]. Moreover, de Magalhaes et al. performed a meta-analysis of publicly available age-related expression microarray datasets on healthy and non-treated adult specimens, and then identified consistently under- or over-expressed genes related to aging processes. The enrichment analysis based on this set of aging-related genes demonstrated that various pathways mediate in aging processes across species, including inflammation and immune response, energy metabolism, etc [41]. The pathway enrichment analysis results of our screened drugs are notably consistent with the conclusions of previous studies, which indeed support the top-rank drugs are potential anti-aging agents.

In particular, we found that the drug with the highest F-ratio score, ZINC218147572 (PubChem CID100005691), has 309 target proteins deposited in STITCH. The pathway analysis by DAVID showed that this drug significantly mediates in TOR signaling pathway (p -value=1.8e-12), as shown in Fig. 3. It has been confirmed that inhibition of TOR pathway can slow down aging and extend healthy lifespan in diverse organisms, including yeast, worms, flies and mice [1, 42]. Interestingly, we conducted pathway enrichment analysis by using the 681 genes in GenAge, as shown in Fig. 4. It can be seen that there are remarkable overlaps between the significant pathways in which the two different set of genes are enriched. (for screenshots of the pathway analysis results see Additional file 2). We strongly suggest that ZINC218147572 is a promising drug for lifespan extension. We would like to conduct wet-lab



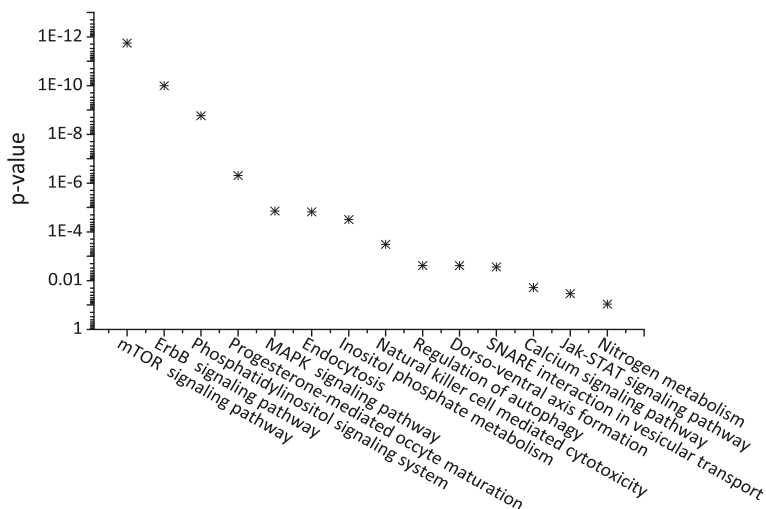


Fig. 3 The signaling and metabolism pathways analysis based on the 309 target proteins of the drug with highest *F*-ratio score, ZINC218147572 (PubChem CID100005691). The pathway enrichment analysis was conducted using DAVID V6.7 [38]

experiments to test its effectiveness on lifespan extension of *C.elegans*, once we get the drug in future (we failed to buy or produce this drug during the preparation of this paper).

Screened drugs have higher chemical similarity to effective drugs than ineffective drugs

We also checked whether the screened drugs have higher chemical similarity to the known effective drugs than ineffective drugs in promoting lifespan or not. For this purpose, we selected 195 drugs of which the *F*-ratio are greater than 2. Based on the chemical fingerprints obtained from PubChem [43], we compute the chemical similarities between the 195 screened drugs and the set of known effective/ineffective drugs. The similarity measure is defined as the cosine angle of the chemical

attribute vectors of two drugs, as describe in Eq. (6). For convenience of presentation, we computed the mean similarity to the effective/ineffective drugs for each screened drug. As shown in Fig. 5, it can be found that the screened drugs have significantly higher mean similarities to the effective drugs than the ineffective drugs. Furthermore, we conducted a pair-sample *t*-test against the null hypothesis that the mean similarity to effective drugs is not greater than that to ineffective drugs (degree of freedom, $df=194$). The result of pair-sample *t*-test indicated that the null hypothesis should be rejected and accept alternative hypothesis ($p\text{-value} \leq 2.33 \times 10^{-24}$).

Wet-lab validation of effectiveness of one screened drug

To further validate the performance of our method, we have conducted wet-lab experiments to check

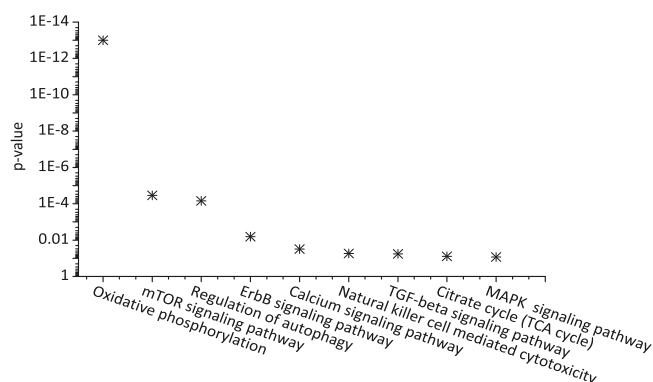


Fig. 4 The signaling and metabolism pathways analysis based on the 681 aging-related genes in GenAge database. The pathway enrichment analysis was conducted using DAVID V6.7 [38]

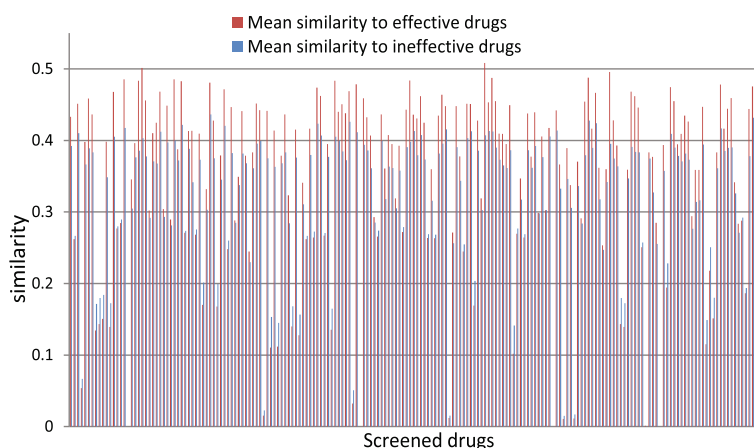


Fig. 5 The mean chemical similarity of the 195 screened drugs (F -ratio ≥ 2) to the effective and ineffective drugs included in the training set. The chemical similarities were calculated according to Eq. (6), on the basis of chemical attributes retrieved from PubChem [43]

the effectiveness of one screened drug, 2-Bromo-4'-nitroacetophenone (PubChem CID000066840), on the lifespan extension of *C. elegans*. Nematodes were treated with 2-Bromo-4'-nitroacetophenone at three different concentrations, 1 $\mu\text{g/ml}$, 5 $\mu\text{g/ml}$ and 10 $\mu\text{g/ml}$, during their lifespan in order to study the effect of 2-Bromo-4'-nitroacetophenone on lifespan of *C. elegans*. As shown in Fig. 6a and b, the drug showed concentration-dependent effects on the lifespan extension of *C. elegans*. The nematodes treated with the drug at the concentration of 1 $\mu\text{g/ml}$ and 5 $\mu\text{g/ml}$ showed significantly extended lifespan compared with untreated animals (one-way ANOVA test, p -value <0.01). The results of the wet-lab experiments provided strong support for the performance of our method.

Data resources

Drugs

The Library of Pharmacologically Active Compounds (LOPAC¹²⁸⁰) contains 1,280 types of compounds that can

be categorized into 55 pharmacological classes according to their mammalian targets. A recent high-throughput biochemical assay has been conducted to screen compounds in LOPAC¹²⁸⁰ to identify drugs that can increase *C. elegans* lifespan [36]. As a result, 60 compounds in LOPAC¹²⁸⁰ have been identified as effective drugs in promoting longevity of *C. elegans*, and the remaining drugs are ineffective drugs. Besides, Lucanic et al. have manually curated 29 types of synthetic compounds and natural products that can promote longevity of *C. elegans* from literature in their review paper [31]. Based on the known effective or ineffective drugs, we built a set of drugs including 1,312 compounds and natural products in total. In particular, 89 effective drugs (60 compounds in LOPAC¹²⁸⁰ plus 29 manually curated synthetic compounds and natural products) were labeled as positive samples, and the remaining 1,220 drugs in LOPAC¹²⁸⁰ were labeled as negative samples (detailed information of the drugs is presented in Additional file 3).

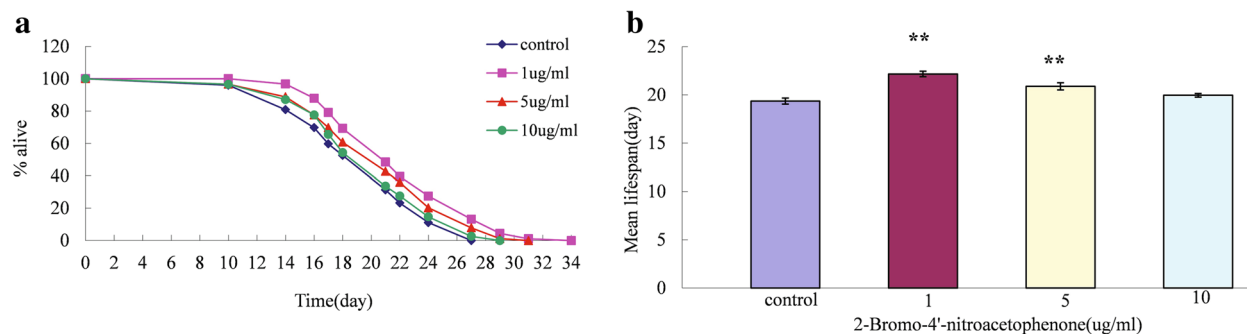


Fig. 6 The screened drug, 2-Bromo-4'-nitroacetophenone (PubChem CID000066840), effectively extends the lifespan of *C. elegans*. **a** Survival curves of the animals under control and treated by three different concentration, 1 $\mu\text{g/ml}$, 5 $\mu\text{g/ml}$ and 10 $\mu\text{g/ml}$, respectively. **b** Bar charts of the mean lifespans in nematodes. (** P -value <0.01)

To build the test set of drugs, we extracted 1,991 FDA-approved small molecule drugs and 207 FDA-approved biotech (protein/peptide) drugs from DrugBank database (V5.0). Totally, 2,198 FDA-approved drugs were included in the test set.

Aging-related genes

Aging-related genes were obtained from GenAge, which is the benchmark database of aging-related genes published by Human Ageing Genomic Resources (HAGR) [26]. GenAge covers the aging-related genes in human and four model organisms, including *Mus musculus*, *Drosophila melanogaster*, worm *Caenorhabditis elegans*, and yeast *Saccharomyces cerevisiae*. Each entry in GenAge is a manually curated by experts to ensure high-quality content. We obtained 681 gene entries for *Caenorhabditis elegans* from GenAge. As a gene may encode multiple different proteins due to the alternative splicing that is prevalent in eukaryote, there are 1,481 unique proteins encoded by these aging-related genes (The aging-related genes and the corresponding proteins are presented in Additional file 4).

Drug-protein bipartite network

Drug-protein interactions were downloaded from STITCH 4.0 [37]. STITCH is a comprehensive database that collects drug-protein interactions from four different sources: experiments, databases, text mining and predicted interactions. Meanwhile, STITCH has calculated an integrative confidence score for each drug-protein interaction, which indicates the confidence of the interaction supported by four types of evidence, i.e. experimental validation, manually curated databases, text mining and predicted interactions. To guarantee high confidence of the drug-protein interactions, we selected only the drug-protein interactions with experimentally supportive or external database evidences.

Besides, we found that most drugs in LOPAC¹²⁸⁰ have numbers of target proteins in human, but relatively less target proteins in *C. elegans*, through retrieval of STITCH. Considering that the scarcity of proteins associated to the effective and ineffective drugs will lead to a small number of training samples in the drug-protein network, we search for orthologous proteins in *C. elegans* of the known target proteins in human. We exploited OrthoList, a compendium of orthologous genes/proteins between *C. elegans* and human, to map the human proteins to orthologs in *C. elegans*. Orthologous genes in OrthoList are compiled from four orthology prediction methods, including InParanoid [44], OrthoMCL [45], HomoloGene and Ensembl Compara [46]. To ensure the quality of the orthologous proteins, we selected only the orthologous proteins if the orthologous proteins have been predicted by at least two methods. As a result, we got 2,518,944

high-confidence interactions between 397,258 drugs and 9,249 proteins in *C. elegans*.

The drug-protein bipartite network was constructed by choosing the high-confidence interactions of which drugs or proteins were included in the drug set or aging-related gene set mentioned above. After removal of small disconnected subnetworks, we got a drug-protein bipartite network with 18,317 drug-protein interactions between 813 unique drugs and 3,660 unique proteins (Detailed information is presented in Additional file 5). In particular, there are 450 known effective/ineffective drugs (34 effective drugs and 416 ineffective drugs) and 20 known aging-related genes among the drug-protein bipartite network, which is used as gold-standard set in the following empirical experiments to evaluate the performance of our method. The data resources and construction of the drug-protein network are illustrated in Fig. 7.

Methods

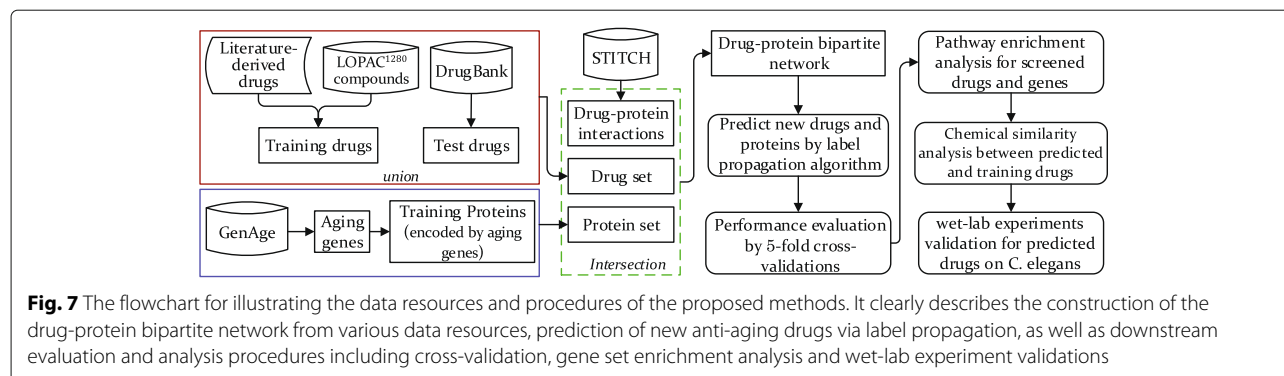
Random walk with restart on drug-protein network

We consider the problem of predict new lifespan-extending drugs in terms of label propagation on the drug-protein bipartite network. Our inspiration is that a small number of experimentally effective and ineffective drugs are labeled as positive and negative samples, and their labels might be propagated to other nodes along the edges of the drug-protein network. Figure 8 illustrates the label propagation starting from a few initially labeled nodes. The effective and ineffective drugs are respectively shown in red color and cyan color, and candidate drugs are shown in gray. The weight of the edges are proportional to the confidence scores of the drug-protein interactions.

Based on the drug-protein network, we start two independent random walks with restart from the effective and ineffective drugs, respectively. After convergence, we obtain a stationary distribution specifying the probability that a random walk with restart will arrive at each node. The drugs that can be reached with higher probability from effective drugs than from ineffective drugs are classified as effective ones, or ineffective ones otherwise.

Formally, suppose we have m drugs and n proteins of interest, and a set of drug-protein interactions with quantitative confidence scores. Denote by V_d and V_t the sets of drug and protein nodes, and $E = \{E(i, j)\}$ a set of edges connecting drugs and proteins, in which i belongs to V_d ($i = 1..m$) and j belongs to V_t ($j = 1..n$). Let $V = V_d \cup V_t$, we construct an undirected network $G = \langle V, E \rangle$. We define a symmetric affinity matrix A , where $A(i, j)$ is equal to the confidence score of the corresponding interaction if $E(i, j) \in E$, or 0 otherwise. Define D as the diagonal matrix of degrees of the nodes, i.e. $D(i, i) = \sum_j A(i, j)$, we constructed the transition matrix T as below:

$$T = D^{-1}A \quad (1)$$



To conduct two independent random walks, we introduce an initial label matrix L with $|V|$ rows and 2 columns, where the two column vectors are indexed according to the effectiveness of drugs that have been experimentally verified. The values of matrix L are given by: 1) $L(i, 1) = 1$ if drug i is ineffective in promoting longevity; 2) $L(i, 2) = 1$ if drug i is effective in promoting longevity; and 3) 0 for all other elements. Note that our method can simultaneously predict lifespan-extending drugs and aging-related genes, by setting the initial label matrix L to include both experimentally validated drugs and known aging-related genes, i.e. the element of second column $L(i, 2) = 1$ if the i -th node is effective drug or aging-related gene coding proteins.

Before starting the random walk, each column of the initial label matrix L was normalized to be a probability distribution:

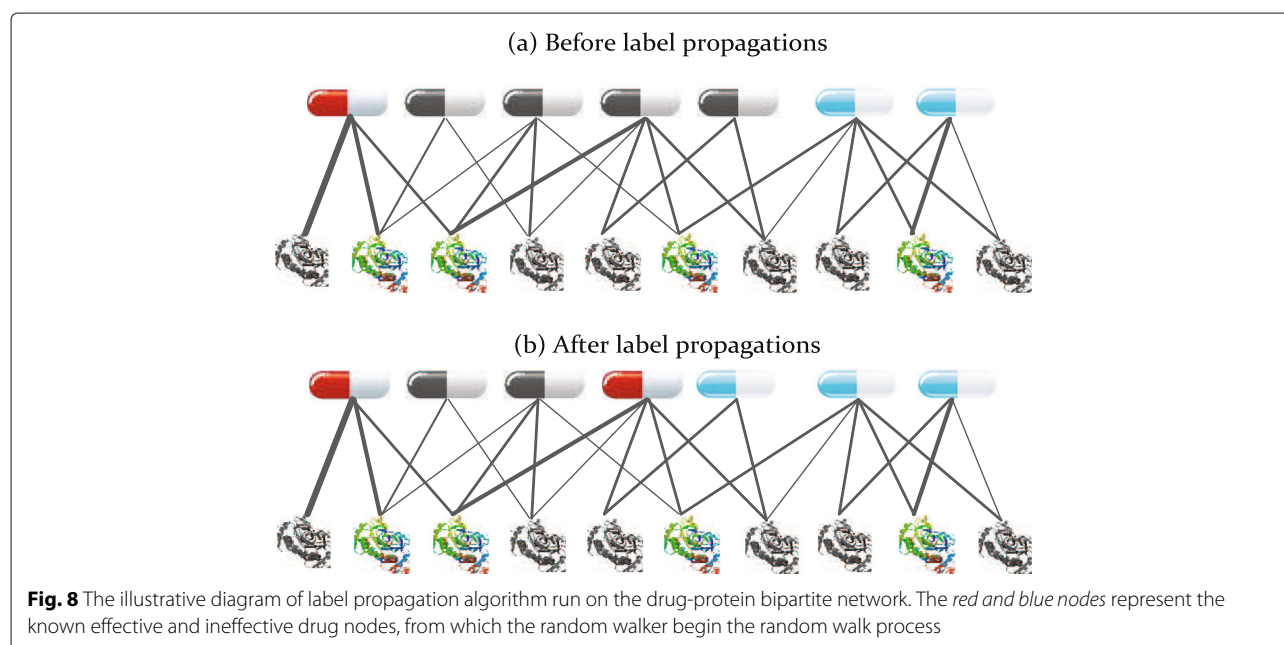
$$L' = (D_L^{-1}L^T)^T, \quad (2)$$

in which D_L is a diagonal matrix with $D_L(i, i) = \sum_j L(j, i)$. We allow the restart of the walk in every step at a source node with probability α , the random walk process can be formulated as:

$$P(t+1) = (1-\alpha)TP(t) + \alpha L', \quad (3)$$

where $P(t)$ is a $|V| \times 2$ matrix and $P(0) = L'$.

Let P^* be the matrix when the random walks converge, i.e. the change between $P(t)$ and $P(t+1)$ (measured by the L1 norm) is less than very small number ϵ . The two columns of P^* are the two stationary distributions that specify the probability to reach each node, corresponding to the two separately random walks with restart from the effective and ineffective drug nodes, respectively. Once we obtained P^* , we computed the odds ratio, called F -ratio as $F(i) = P^*(i, 2)/P^*(i, 1)$ for each candidate drug i . If $F(i)$ is greater than a predefined threshold, drug i is classified as effective one in promoting longevity. To build recall-precision curve, we ranked all candidate drugs



according to their F -ratio values and gradually increase the value of threshold θ used for classification decision. In Algorithm 1, we outline the steps of the label propagation algorithm on the drug-protein network. Moreover, an illustrative flowchart for our method is presented in Fig. 7.

Algorithm 1 Label propagation on drug-protein network via random walk with restart

Input: adjacent matrix A , diagonal matrix D , initial label matrix L and threshold θ

Output: Ranked drugs by F -ratio scores with classification label according to threshold θ

Algorithm: random walk with restart on drug-protein network

- 1: Build the transition matrix $T = D^{-1}A$;
 - 2: Normalize the initial label matrix $L' = (D_L^{-1}L^T)^T$ where D_L is a diagonal matrix with element $D_L(i, i) = \sum_j L(j, i)$
 - 3: Iterate $P(t + 1) = (1 - \alpha)TP(t) + \alpha L'$ to converge, where $P(t)$ is a $|V| \times 2$ matrix with $P(0) = L'$, and α is the restart probability at a source node;
 - 4: Compute F -ratio for each drug i as $F(i) = \frac{P^*(i,2)}{P^*(i,1)}$, where P^* is the convergence matrix.
 - 5: Classify drug i as effective if $F(i) > \theta$, or ineffective otherwise;
-

Drug similarity

For each drug, we obtained the SMILES from PubChem database [43] and then generated the chemical features and fingerprints using PaDEL software (release v2.21) [47]. PaDEL takes as input the SMILES of each drug to calculate 1D and 2D physicochemical descriptive features and fingerprints. For each drug, the attributes consist of 1,445 descriptive features, including such as atom count, bond count, molecular weight, xlogP, and 880 chemical fingerprints. Subsequently, we removed the features with the same attribute values across all drugs, and obtained 1,744 chemical attributes for each compound.

Based on the chemical attributes, we computed the chemical similarity between each pair of drugs. Formally, each drug was represented by a 1,744-dimension vector, in which the element is equal to the value outputted by PaDEL, or 0 otherwise. Denote by $\vec{c}_i = \{c_{i1}, c_{i2}, \dots, c_{iK}\}$ and $\vec{c}_j = \{c_{j1}, c_{j2}, \dots, c_{jK}\}$ the chemical attribute vectors of drugs i and drug j , we computed the *cosine similarity* as below:

$$S_d^{(1)}(i, j) = \frac{\sum_{k=1}^K c_{ik}c_{jk}}{\sqrt{\sum_{k=1}^K c_{ik}^2} \sqrt{\sum_{k=1}^K c_{jk}^2}}, \quad (4)$$

in which K is the length of the chemical attribute vector ($K=1774$).

Also, we computed the drug-drug similarities, referred to as *target sharing similarity* based on our built drug-protein interaction network. Each drug was represented by a 3,660-dimension binary vector whose element is 1 if the corresponding protein is targeted by the drug and otherwise 0. We built the target sharing similarity matrix based on the number of shared targets between two drugs. This similarity measure is defined as the ratio of number of common target proteins to the total number of target proteins of the two drugs. This measure is actually equivalent to *Jaccard score* that can be mathematically defined as:

$$S_d^{(2)}(i, j) = \frac{|t_i \cap t_j|}{|t_i \cup t_j|} \quad (5)$$

in which t_i and t_j are the binary vectors representing the proteins targeted by drug i and drug j , respectively.

Since we obtained two similarity measures derived from different attributes for each pair of drugs, we thus integrated the two similarity measures into a comprehensive similarity measure as below:

$$S_d(ij) = 1 - \prod_n (1 - S_d^{(n)}(ij)), \quad (6)$$

in which $S_d^{(1)}(ij)$ and $S_d^{(2)}(ij)$ represent the similarity measures between i and j -th drugs derived from chemical attributes and target proteins, respectively.

Prediction using classical classifiers

To calibrate the performance of our proposed method, two classical classifiers, k-Nearest Neighbor(kNN) and Support Vector Machine (SVM), were used to predict the lifespan-extending drugs. The k-Nearest Neighbor(kNN) was run by using Weka 3.7 [48], and SVM was run by using libsvm 3.17 [49]. To apply the kNN classifier, we convert the the similarity measure defined in Eq. (6) to the distance measure by computing $1 - S_dij$. The parameter k was set to 1, 3 and 5, respectively. As similar results were obtained for different k values, we reported only the results of $k=3$. For SVM, We concatenated the chemical attributes and the target proteins of each drug, and thus obtain a 5,434-dimension vector that is taken as input by libsvm. The radial basis function kernel was used, and other parameters were tuned by 10-fold cross validations. The experimentally verified drugs are labeled as positive and negative sample and used to train the SVM classifier, and the learned model was then used to classify the set of test drugs.

Reagents and worm strain maintenance

2-Bromo-4'-nitroacetophenone (PubChem CID000066840) was obtained from Sigma-Aldrich (St. Louis, MO, USA).

Wild-type *C. elegans* N2, which is originally obtained from Caenorhabditis Genetics Center, was maintained on nematode growth medium (NGM) plates seeded with *Escherichia coli* OP50 at 20°C as described previously [50]. Age synchronous populations of nematodes were obtained as described [51]. The compound was added to the NGM plates just before plating.

Lifespan assay

The lifespan assay was performed basically as described in [52]. 2-Bromo-4'-nitroacetophenone treatment was performed throughout the lifespan from the stage of L4-larvae. The worms were transferred daily for the first several days of adulthood during the assay. The nematodes were checked at intervals during the lifespan when they began to die and worms were considered as dead when they did not respond to the stimulation of a platinum wire.

Discussions and conclusions

It is of frequent occurrence that labeled samples, which mean their functions were experimentally validated, are extremely rare and unlabeled data is abundant. Therefore, many semi-supervised algorithms have been developed, including self-training, S3VMs and graph-based methods, which can fully utilize fewer labeled data and large unlabeled data in the learning and predicting processes. The merit of network-based semi-supervised algorithms lie in that they can take full advantage of the inherent structure between labeled and unlabeled data. In this paper, we employed the random walk with restart on drug-protein bipartite network to predict drugs with lifespan-extending effects on *C.elegans*. In fact, label propagation on bipartite networks has been widely used in various fields, such as drug repositioning [53], personal recommendation [54] and political polarity classification [55].

To the best of our knowledge, we are the first to propose the computational method for predicting lifespan-extending drugs in term of drug-protein interactions. To calibrate the performance of the proposed method, we employed two classical models, kNN and SVM, in the empirical evaluation of the performance, as we have not found any other existing computational methods for screening lifespan-extending drugs in term of drug-protein interactions. The 5-fold cross-validations on the gold-standard dataset showed that our method achieve higher performance than the two classical classifiers.

Our wet-lab experiments verified that one of the screened drugs, 2-Bromo-4'-nitroacetophenone (PubChem CID000066840), can significantly promote the longevity of *C.elegans*. In our future work, we will carry out systematic wet-lab experiments to verify what percentage of the screened drugs are effective in promoting longevity of *C. elegans*, and explore the biological

mechanism of anti-aging effect of the drugs in pathway and gene levels.

In summary, we proposed a semi-supervised classification algorithm, random walk with restart on bipartite graph, to predict drugs with lifespan-extending effects on *C.elegans*. Our method aims at predicting lifespan-extending drugs in a large scale, and narrowing down the scope of candidate drugs needed to be verified by wet-lab experiments. The results of empirical experiments and wet-lab experiment show that our computational screening approach achieves state-of-the-art performance.

Additional files

Additional file 1: Detailed information of the predicted drugs, including the probabilities derived from two random walks independently started from effective and ineffective drugs, as well as the *F*-ratio. (XLSX 281 kb)

Additional file 2: The screenshots of the pathway enrichment analysis results based on the 309 target proteins of drug ZINC218147572 and the set of 681 aging-related genes in GenAge, respectively. (DOCX 383 kb)

Additional file 3: Detailed information of the drugs collected from bioassay screening result of the LOPAC¹²⁸⁰, as well as the manually curated anti-aging drugs from literature review. (XLSX 180 kb)

Additional file 4: The aging-related genes of *C. elegans* extracted from GenAge database, together with the corresponding transcripts of each aging-related genes. (XLSX 99 kb)

Additional file 5: The drug-disease bipartite network used in the paper, which includes 18,317 drug-protein interactions between 813 unique drugs and 3,660 unique proteins. (XLSX 440 kb)

Acknowledgments

The authors wish to thank M.S. YingLong Song for the implementation of the computational method.

Declarations

This article has been published as part of *BMC Systems Biology* Volume 10 Supplement 4, 2016: Proceedings of the 27th International Conference on Genome Informatics: systems biology. The full contents of the supplement are available online at <http://bmcsystbiol.biomedcentral.com/articles/supplements/volume-10-supplement-4>.

Funding

This work was supported by the National Natural Science Foundation of China under grants No. 61672113 and No. 31300707, and the publication cost was supported by the National Natural Science Foundation of China under grants No. 31300707.

Authors' contributions

HL proposed the main idea and collected the data sets. GMM conducted the *in silico* experiments. TX conducted the wet-lab experiments on *C. elegans*. HL and ZHZ collaborated to write the manuscript. JHG gave helpful suggestions for improving the quality of this manuscript. LBL and ZZH supervised all aspects of the work. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹Changzhou University, Jiangsu 213164, China. ²Changzhou NO. 7 People's Hospital, Changzhou, Jiangsu 213011, China. ³Department of Computer Science and Technology, Tongji University, Shanghai 201804, China.

Published: 23 December 2016

References

- Fontana L, Partridge L, Longo VD. Dietary restriction, growth factors and aging: from yeast to humans. *Science*. 2010;328(5976):321–6.
- Anisimov V. Biology of aging and cancer. *Cancer Control*. 2007;14(1):23–31.
- Finkel T, Serrano M, Blasco MA. The common biology of cancer and ageing. *Nature*. 2007;448:767–74.
- Kennedy B, Steffen K, Kaeblerlein M. Ruminations on dietary restriction and aging. *Cell Mol Life Sci*. 2007;64:1323–8.
- Anderson R, Shanmuganayagam D, Weindruch R. Caloric restriction and aging: studies in mice and monkeys. *Toxicol Pathol*. 2009;37(1):47–51.
- Everitt A, Le Couteur D. Life extension by calorie restriction in humans. *Ann N Y Acad Sci*. 2007;1114:428–33.
- Bartke A. Long-lived klotho mice: new insights into the roles of igf-1 and insulin in aging. *Trends Endocrinol Metab*. 2006;17:33–5.
- Vellai T, Takacs Vellai K, Zhang Y, Kovacs A, Orosz L, Muller F. Genetics: influence of tor kinase on lifespan in *c. elegans*. *Nature*. 2003;426(6967):620.
- Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloan A, Even P, Cervera P, Le Bouc Y. Igf-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*. 2003;421(6919):182–7.
- Cohen E, Paulsson J, Blinder P, Burstyn-Cohen T, Du D, Estepa G, Adame A, Pham HM, Holzenberger M, Kelly J, Maslah E, Dillin A. Reduced igf-1 signaling delays age-associated proteotoxicity in mice. *Cell*. 2009;139(6):1157–69.
- Robida-Stubbs S, Glover-Cutter K, Lamming D, Mizunuma M, Narasimhan S, Neumann-Haefelin E, Sabatini D, Blackwell TK. Tor signaling and rapamycin influence longevity by regulating skn-1/nrf and daf-16/foxo. *Cell Metab*. 2012;15:713–24.
- Johnson S, Rabinovitch P, Kaeblerlein M. mtor is a key modulator of ageing and age-related disease. *Nature*. 2013;493:338–45.
- Kenyon C. The genetics of ageing. *Nature*. 2010;464:504–12.
- Harrison D, Strong R, Sharp Z, Nelson J, Astle C, Flurkey K, Nadon N, Wilkinson J, Frenkel K, Carter C, Pahor M, Javors M, Fernandez E, Miller R. Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature*. 2009;460(7253):392–5.
- Frankowski H, Alavez S, Spilman P, Mark K, Nelson J, Mollahan P, Rao R, Chen S, Lithgow G, Ellerby H. Dimethyl sulfoxide and dimethyl formamide increase lifespan of *c. elegans* in liquid. *Mech Ageing Dev*. 2013;134:69–78.
- Edwards C, Canfield J, Copes N, Rehan M, Lipps D, Bradshaw P. D-beta-hydroxybutyrate extends lifespan in *c. elegans*. *Aging (Albany NY)*. 2014;6(8):621–44.
- Honda Y, Fujita Y, Maruyama H, Araki Y, Ichihara K, Sato A, Kojima T, Tanaka M, Nozawa Y, Ito M, Honda S. Lifespan-extending effects of royal jelly and its related substances on the nematode *caenorhabditis elegans*. *PLoS ONE*. 2011;6(8):23527.
- Schriner S, Kuramada S, Lopez T, Truong S, Pham A, Jafari M. Extension of *drosophila* lifespan by cinnamon through a sex-specific dependence on the insulin receptor substrate chico. *Exp Gerontol*. 2014;60:220–30.
- Detienne G, De Haes W, Ernst U, Schoofs L, Temmerman L. Royalactin extends lifespan of *caenorhabditis elegans* through epidermal growth factor signaling. *Exp Gerontol*. 2014;60:129–35.
- Zhang Y, Lv T, Li M, Xue T, Liu H, Zhang W, Ding X, Zhuang Z. Anti-aging effect of polysaccharide from *bletilla striata* on nematode *caenorhabditis elegans*. *Pharmacogn Mag*. 2015;11(43):449–54.
- Wan Q, Zheng S, Wu G, Luo H. Aspirin extends the lifespan of *caenorhabditis elegans* via ampk and daf-16/foxo in dietary restriction pathway. *Exp Gerontol*. 2013;48(5):499–506.
- Anisimov V, Berstein L, Egormin P, Piskunova T, Popovich I, Zabezhinski M, Tyndyk M, Yurova M, Kovalenko I, Poroshina T, Semenchenko A. Metformin slows down aging and extends life span of female shr mice. *Cell Cycle*. 2008;7:2769–73.
- Bauer JH, Goupil S, Garber GB, Helfand SL. An accelerated assay for the identification of lifespan-extending interventions in *drosophila melanogaster*. *Proc Natl Acad Sci USA*. 2004;101:12980–5.
- Carretero M, Gomez-Amaro P, Petrascheck M. Pharmacological classes that extend lifespan of *caenorhabditis elegans*. *Front Genet*. 2015;6:77.
- Pan F, Chiu C, Pulapura S, Mehan M, Nunez-Iglesias J, Zhang K, Kamath K, Waterman M, Finch C, Zhou X. Gene aging nexus: a web database and data mining platform for microarray data on aging. *Nucleic Acids Res*. 2007;35:756–9.
- Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, de Magalhaes JP. Human ageing genomic resources: Integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res*. 2013;41(D1):1027–33.
- Tacutu R, Budovsky A, Fraifeld V. The netage database: a compendium of networks for longevity, age-related diseases and associated processes. *BioGerontology*. 2010;11(4):513–22.
- Pan F, Chiu CH, Pulapura S, Mehan MR, Nunez-Iglesias J, Zhang K, Kamath K, Waterman MS, Finch CE, Zhou XJ. Gene Aging Nexus: a web database and data mining platform for microarray data on aging. *Nucleic Acids Res*. 2007;35(Database issue):756–9.
- Leung M, Williams P, Benedetto A, Au C, Helmcke K, Aschner M, Meyer J. *Caenorhabditis elegans*: an emerging model in biomedical and environmental toxicology. *Toxicol Sci*. 2008;106(1):5–28.
- Kaletta T, Hengartner MO. Finding function in novel targets: *C. elegans* as a model organism. *Nat Rev Drug Discov*. 2006;5:387–99.
- Lucanic M, Lithgow G, Alavez S. Pharmacological lifespan extension of invertebrates. *Ageing Res Rev*. 2013;12(1):445–58.
- Gerber S, Assmus H, Bakker B, Klipp E. Drug-efficacy depends on the inhibitor type and the target position in a metabolic network—a systematic study. *J Theor Biol*. 2008;252(3):442–55.
- Pahikkala T, Airola A, Pietila S, Shakyawar S, Sz wajda A, Tang J, Aittokallio T. Toward more realistic drug-target interaction predictions. *Brief Bioinform*. 2015;16(2):325–7.
- Tang J, Karhinen L, Xu T, Sz wajda A, Yadav B, Wennerberg K, Aittokallio T. Target inhibition networks: predicting selective combinations of druggable targets to block cancer survival pathways. *PLoS Comput Biol*. 2013;9(9):1003226.
- Pang K, Wan Y, Choi W, Donehower L, Sun J, Pant D, Liu Z. Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics*. 2014;30:1456–63.
- Ye X, Linton JM, Schork N, Buck L, Petrascheck M. A pharmacological network for lifespan extension in *caenorhabditis elegans*. *Aging Cell*. 2014;13(2):206–15.
- Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher T, von Mering C, Jensen L, Bork P. Stitch 4: integration of protein chemical interactions with user data. *Nucleic Acids Res*. 2014;42:401–7.
- Huang D, Sherman B, Lempicki R. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
- McCarroll SA, Murphy CT, Zou S, Pletcher SD, Chin CS, Jan YN, Kenyon C, Bargmann CI, Li H. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat Genet*. 2004;36:197–204.
- McElwee J, Schuster E, Blanc E, Piper M, Thomas J, Patel D, Selman C, Withers D, Thornton J, Partridge L, Gems D. Evolutionary conservation of regulated longevity assurance mechanisms. *Genome Biol*. 2007;8(7):132.
- de Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*. 2009;25(7):875–81.
- Fontana L. Modulating human aging and age-associated diseases. *Biochim Biophys Acta*. 2009;1790(10):1133–8.
- Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem Substance and Compound databases. *Nucleic Acids Res*. 2016;44(D1):D1202–13.
- Sonnhammer EL, Ostlund G. Inparanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res*. 2009;37:234–9.
- Li L, Stoeckert C, Roos D. Orthomcl: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13(9):2178–89.
- Vilella AJ. Ensemblcompara genetrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. 2009;19:327–35.

47. Yap C. Padel-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem.* 2011;32(7):1466–74.
48. Hall M, Eibe F, Geoffrey H, Bernhard P, Peter R, Witten IH. The weka data mining software: an update. *SIGKDD Explorations.* 2009;11:10–18.
49. Chang C, Lin C. Libsvm: a library for support vector machines. *ACM Trans Intell Syst Technol.* 2011;2:1–27.
50. Zhuang Z, Lv T, Li M, Zhang Y, Xue T, Yang L, Liu H, Zhang W. The lifespan-extending effects of nymphaea hybrid root extract in the nematode *caenorhabditis elegans*. *Plant Foods Hum Nutr.* 2014;69(4):304–9.
51. Zhuang Z, Zhao Y, Wu Q, Li M, Liu H, Sun L, Gao W, Wang D. Adverse effects from clenbuterol and ractopamine on nematode *caenorhabditis elegans* and the underlying mechanism. *PLoS ONE.* 2014;9(1):85482.
52. Zhang W, Lv T, Li M, Wu Q, Yang L, Liu H, Sun D, Sun L, Zhuang Z, Wang D. Beneficial effects of wheat gluten hydrolysate to extend lifespan and induce stress resistance in nematode *caenorhabditis elegans*. *PLoS ONE.* 2013;8(9):74553.
53. Chen H, Zhang H, Zhang ZP, Cao Y, Tang W. Network-based inference methods for drug repositioning. *Comput Math Methods Med.* 2015;2015:130620.
54. Zhou T, Ren J, Medo M, Zhang YC. Bipartite network projection and personal recommendation. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2007;76(4 Pt 2):046115.
55. Zhou DX, Resnick P, Mei Q. Classifying the political leaning of news 738 articles and users from user votes. *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media.* 2011.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

