BMC Bioinformatics

RESEARCH Open Access

AligNet: alignment of protein-protein interaction networks

Adrià Alcalá[1,2], Ricardo Alberich[1,2], Mercè Llabrés[1,2*], Francesc Rosselló[1,2] and Gabriel Valiente[3]

From 15th International Symposium on Bioinformatics Research and Applications (ISBRA'19)
Barcelona, Spain. 3-6 June 2019

*Correspondence:
merce.llabres@uib.es
[1] Department of Mathematics and Computer Science, University of the Balearic Islands, E-07122 Palma de Mallorca, Spain
[2] Balearic Islands Health Research Institute (IdISBa) E-07010 Palma de Mallorca, Spain
Full list of author information is available at the end of the article

Abstract

Background: All molecular functions and biological processes are carried out by groups of proteins that interact with each other. Metaproteomic data continuously generates new proteins whose molecular functions and relations must be discovered. A widely accepted structure to model functional relations between proteins are protein-protein interaction networks (PPIN), and their analysis and alignment has become a key ingredient in the study and prediction of protein-protein interactions, protein function, and evolutionary conserved assembly pathways of protein complexes. Several PPIN aligners have been proposed, but attaining the right balance between network topology and biological information is one of the most difficult and key points in the design of any PPIN alignment algorithm.

Results: Motivated by the challenge of well-balanced and efficient algorithms, we have designed and implemented AligNet, a parameter-free pairwise PPIN alignment algorithm aimed at bridging the gap between topologically efficient and biologically meaningful matchings. A comparison of the results obtained with AligNet and with the best aligners shows that AligNet achieves indeed a good balance between topological and biological matching.

Conclusion: In this paper we present AligNet, a new pairwise global PPIN aligner that produces biologically meaningful alignments, by achieving a good balance between structural matching and protein function conservation, and more efficient computations than state-of-the-art tools.

Keywords: Protein-protein interaction network, Global alignment, Network matching, Functional consistency

Background

One of the most difficult problems in systems biology is to discover protein-protein interactions as well as their associated functions. The alignment and analysis of protein-protein interaction networks (PPIN) has become a key ingredient to obtain functional orthologs as well as evolutionary conserved assembly pathways of protein complexes. With this purpose, several pairwise alignment algorithms have been proposed in the last 15 years. The early aligners [1–5] were aimed at finding *local* alignments between regions with similar structure in the networks under comparison. But since the alignments between regions of the pair of PPIN could be mutually inconsistent, it could be impossible to merge the alignments between regions into an alignment of the whole networks. In contrast, a *global* alignment algorithm is aimed at finding the best overall alignment between whole PPIN [6]. Several such global PPIN aligners have been proposed during the last years [4, 7–11].

Most PPIN aligners are based on the idea that "two nodes are similar when their corresponding neighbors are so," taking into account both the network topology and the biological features of the proteins in the definition of "similarity." The problem is that attaining the right balance between network topology and biological information is one of the most difficult and key points in any PPIN alignment algorithm. As it is shown in [12, 13], when an alignment process is guided by topological information only, it produces alignments with a high topological coherence but a low biological coherence, while when it is guided by sequence information only, the resulting alignments have a high biological coherence but a low topological coherence. This becomes specially inconvenient in those aligners where the user has to choose the value of a parameter that specifies the desired balance between the topological and the sequence similarities. In addition, most aligners are not efficient from the computational point of view.

Motivated by this lack of well-balanced and efficient algorithms, we have designed AligNet, a parameter-free pairwise PPIN alignment algorithm aimed at filling the gap between efficient topologically and biologically meaningful matchings. The overall idea of the algorithm is to obtain many local alignments that are combined and extended into a meaningful global alignment. The final alignment captures the benefits of considering both types of alignments: with the local alignments we capture the topological similarity between the networks and we speed up the running time of the algorithm, while with the final global alignment we solve the inconsistencies among the local alignments and yield an overall alignment of the pair of input PPIN. AligNet has been implemented in R [14], and the implementation is freely available from https://github.com/biocom-uib/AligNet.

A comparison of the results obtained with AligNet and with the best aligners assessed in [12, 13] shows that AligNet achieves indeed a good balance between topological and biological matching. In the tests reported in this paper, AligNet obtained high functional consistence scores between aligned proteins in most of the alignments and also a reasonable fraction of conserved interactions. In addition, AligNet, together with HubAlign [8], had the best running times among all the aligners considered in our tests.

Methods

In this paper, by a *graph* we understand an *undirected graph*, that is, a structure $G = (V, E)$ with V a finite set of *nodes* and E a family of 2-element subsets $\{u, v\}$ of V called the *edges* of the graph. A PPIN is modelled in a natural way as a graph, with its nodes representing the proteins and its edges, their interactions.

We introduce now some notations. Let $G = (V, E)$ be a graph. We say that an edge $e = \{u, v\}$ is *incident* to u and v. The nodes v such that $\{u, v\} \in E$ are the *neighbors* of u, and they form the set $N_G(u)$. The *degree* $\deg(u)$ of a node $u \in V$ is the number of edges incident to it. A *path* between two nodes $u, v \in V$ is a sequence of pairwise different edges $\{u, u_1\}, \{u_1, u_2\}, \ldots, \left\{u_{k-1}, u_k\right\}, \{u_k, v\}$ such that the first and last edges are incident to u and v, respectively, and every pair of consecutive edges share a node (different from u and v, in the case of the first and last edges, respectively). The *length* of a path is the number of edges forming it, and its *intermediate nodes* are $u_1, \ldots, u_k$. Two nodes are *connected* when there exists a path between them. For every pair of connected nodes $u, v \in V$, their *distance* $d_G(u, v)$ in G is the length of a shortest path connecting them. The *diameter* $D(G)$ of G is the maximum distance between any two connected nodes in G. The cardinality of a set X is denoted by $|X|$.

AligNet receives as input two graphs $G = (V, E)$ and $G' = (V', E')$ representing two PPIN (in particular, each node of them is injectively identified with a protein) and it produces, as output, a similarity score for them and a local and a global alignment between them. Figure 1 shows the pipeline of our algorithm AligNet. The main steps in AligNet that are described below are:

1 The computation of overlapping clusterings $C(G)$ and $C(G')$, respectively, of the input networks G and G'.
2 The computation of alignments between pairs of clusters in $C(G)$ and $C(G')$.
3 The computation of a matching between $C(G)$ and $C(G')$.
4 The computation of a local alignment of the input networks G and G'.
5 The extension of this local alignment to a meaningful global alignment.

Step 1. Overlapping clusterings. The first step in AligNet consists in computing an overlapping clustering of each input network. These clusterings are based on the following similarity score $s(u, v)$ between pairs of proteins (nodes) u, v in a PPIN G: If u, v are

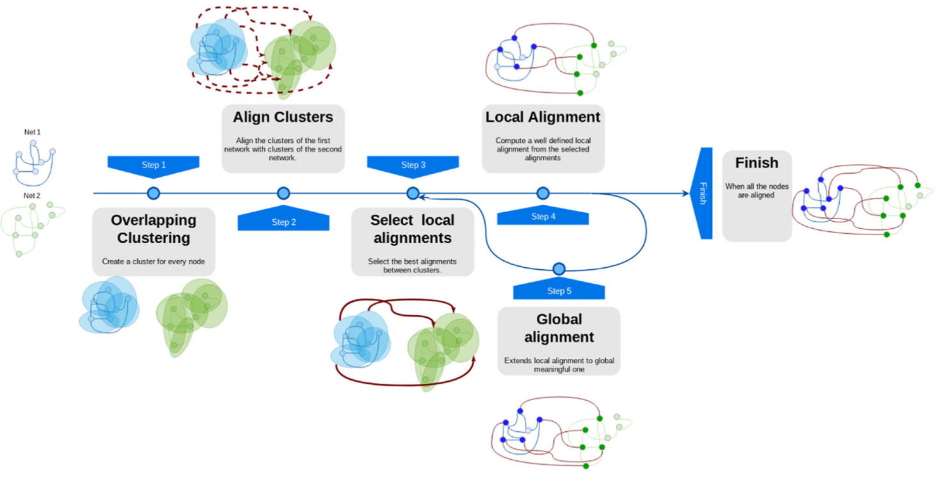
Net 1
Net 2
Step 1
Overlapping Clustering
Create a cluster for every node
Align Clusters
Align the clusters of the first network with clusters of the second network.
Step 2
Step 3
Select local alignments
Select the best alignments between clusters.
Local Alignment
Compute a well defined local alignment from the selected alignments
Step 4
Step 5
Global alignment
Extends local alignment to global meaningful one
Finish
Finish
When all the nodes are aligned

Fig. 1 Pipeline of AligNet algorithm

not connected by a path, then $s(u,v) = 0$, and if they are connected,

$$s(u,v) = \frac{B(u,v) + \frac{D(G)+1-d_G(u,v)}{D(G)+1}}{2},$$

where $B(u,v)$ is the *normalized bit score* of the proteins u and v, that is, the rescaled version of their alignment score obtained with BLAST+, which is independent of the size of the search space [15]. The intuition behind this similarity score is that two proteins are similar if they have similar sequences of nucleotides and they are relatively close to each other in the graph.

To obtain the overlapping clustering of an input network, we define a cluster centered at each node. To avoid the choice of a fixed and arbitrary cluster size, we considered the similarity score distribution and define the cluster centered at each node as follows. Let α be the third quartile of the distribution of the similarity score values of pairs of nodes, so that only 25% of the pairs of nodes (u,v) are such that $s(u,v) > \alpha$. Then, for every node $u \in V$, the *cluster* C_u in G *centered* at u is

$$C_u = \{v \in V \mid s(u,v) > \alpha\}.$$

Let $C(G) = \{C_u \mid u \in V\}$ and $C(G') = \left\{C_{u'} \mid u' \in V'\right\}$.

Figure 2 displays two toy PPIN that will be used as a running example throughout this section. The first network consists of 8 nodes and 9 edges, while the second network consists of 9 nodes and 17 edges. Figure 3 displays the PPI networks considered as a running example as well as its overlapping clustering. The first network consists of 8 nodes and 9 edges, so there are 8 clusters. The second network consists of 9 nodes and 17 edges, and its overlapping clustering has 9 clusters.

Step 2. Alignments between pairs of clusters. In this second step, AligNet computes an alignment between every pair of clusters $C_u \in C(G)$ and $C_{u'} \in C(G')$ such that $B(u,u') > 0$. These alignments define an alignment score between every such a pair of clusters that will be used in the third step to compute a matching between $C(G)$ and $C(G')$.

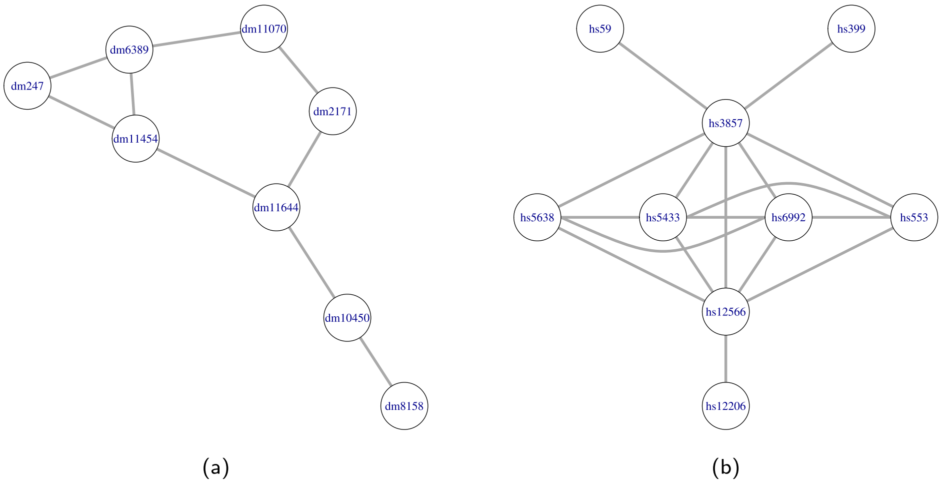

Fig. 2 **a** A subnetwork of the *Drosophila melanogaster* PPI network. **b** A subnetwork of the *Homo Sapiens* PPI network

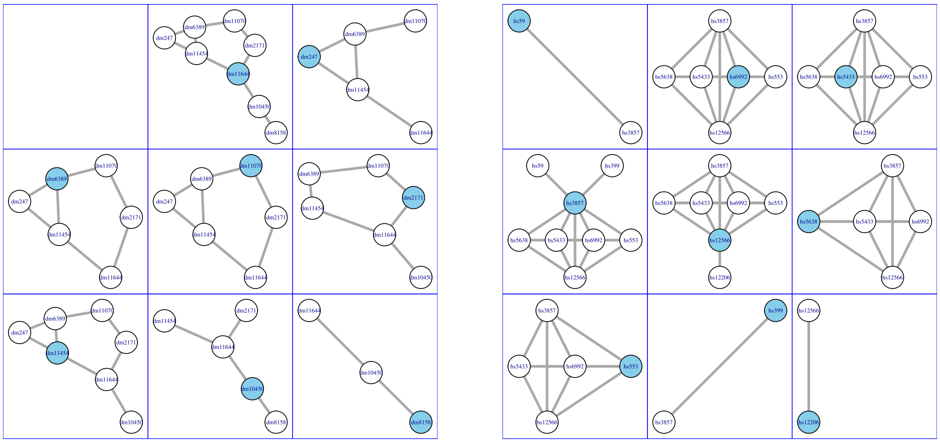

Fig. 3 Overlapping clusterings. This figure shows the overlapping clustering on the PPINs in Fig. 2 obtained by AligNet. We can see here the 8 clusters in the network in Fig. 2 on the left, and the 9 clusters in the network in Fig. 2 on the right. The center of every cluster is highlighted in blue. Since we have considered two small pieces of a PPIN, we obtain here that, the first cluster on the left is the entire piece of network. In the right, we obtain also the entire piece of network in the second cluster on the right. Notice that we obtain the whole piece of the network when we consider the cluster of a node that is in the center of the network

Formally, for every $u \in V$ and $u' \in V'$ such that $B(u,u') > 0$, the alignment between $C_u \in C(G)$ and $C_{u'} \in C(G')$ is obtained as follows:

(i) Match u with u'. Set $L_{u,u'} = \left\{(u,u')\right\}$, $L^{(1)}_{u,u'} = \{u\}$ and $L^{(2)}_{u,u'} = \{u'\}$.

(ii) For every $v \in C_u \cap N_G(u)$ and for every $v' \in C_{u'} \cap N_{G'}(u')$, let

$$F\left(v,v'\right) = |\deg(v) - \deg(v')| - B\left(v,v'\right) + 1.$$

Compute a matching $M_{u,u'} \subseteq (C_u \cap N_G(u)) \times (C_{u'} \cap N_{G'}(u'))$ that minimizes $\sum_{(v,v')\in M_{u,u'}} F(v,v')$ using the Hungarian algorithm [16]. Sort the pairs in $M_{u,u'}$ in decreasing order of their F value, and concatenate them to $L_{u,u'}$. Add their first coordinates to $L^{(1)}_{u,u'}$ and their second coordinates to $L^{(2)}_{u,u'}$.

(iii) Iterate step (ii), replacing (u,u') by the rest of the pairs in $L_{u,u'}$ and removing from C_u and $C_{u'}$ the nodes already aligned.

More specifically, in the k-th iteration, take the k-th element (v_0, v'_0) of $L_{u,u'}$. For every $w \in (C_u \setminus L^{(1)}_{u,u'}) \cap N_G(v_0)$ and every $w' \in (C_{u'} \setminus L^{(2)}_{u,u'}) \cap N_{G'}(v'_0)$, compute $F(w,w')$. Then, compute a matching

$$M_{v_0,v'_0} \subseteq \left((C_u \setminus L^{(1)}_{u,u'}) \cap N_G(v_0)\right) \times \left((C_{u'} \setminus L^{(2)}_{u,u'}) \cap N_{G'}(v'_0)\right)$$

that minimizes $\sum_{(v,v')\in M_{v_0,v'_0}} F(v,v')$. Sort the pairs forming M_{v_0,v'_0} in decreasing order of their F value, and concatenate them to $L_{u,u'}$. Add their first coordinates to $L^{(1)}_{u,u'}$ and their second coordinates to $L^{(2)}_{u,u'}$.

The resulting alignment $L_{u,u'}$ defines a partial injective mapping $\eta_{u,u'} : C_u \to C_{u'}$. The nodes in C_u that are matched to nodes in $C_{u'}$ form the domain of the mapping $\eta_{u,u'}$, which is denoted by $Dom\,\eta_{u,u'}$. Figure 4 shows an example of the alignment of a pair of clusters: one cluster from the first network and another cluster from the second network. The general idea behind this alignment procedure is that u is matched to u' and then a node $v \in C_u$ should be matched to a node v' in $C_{u'}$ when they have similar sequences and similar degrees, provided that, furthermore, there exist paths connecting u with v and u'

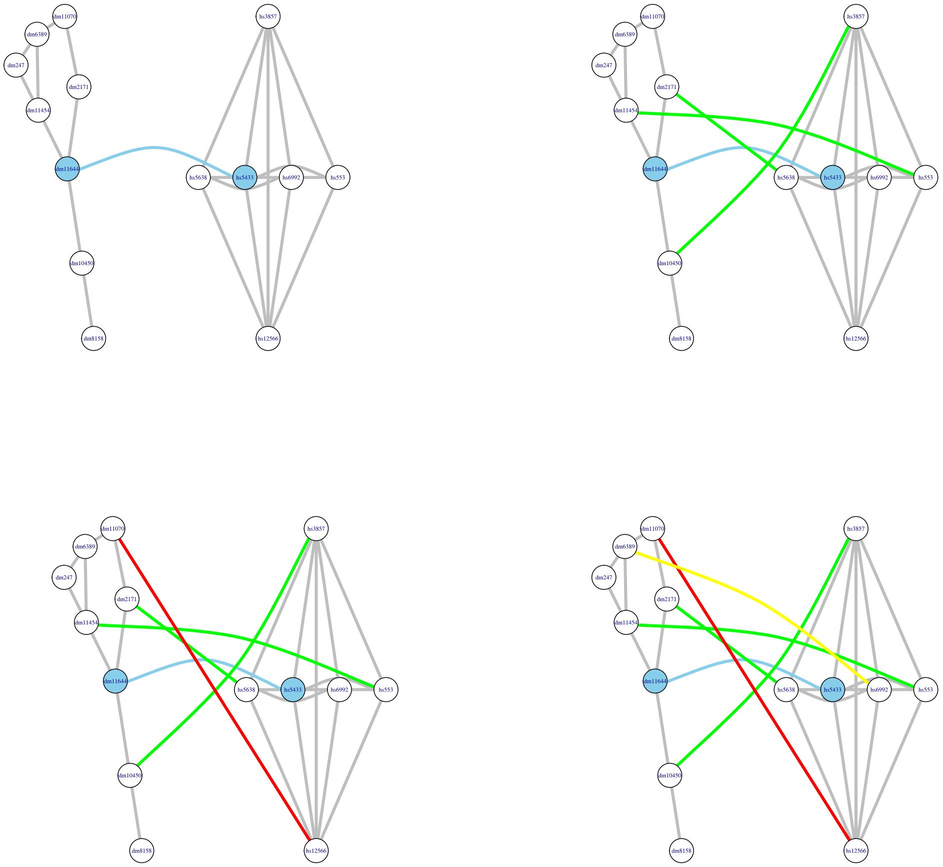

Fig. 4 Alignment of a pair of clusters. This figure shows how AligNet aligns two clusters which corresponds to Step 2 of our algorithm. The clusters in this example are, respectively, the first in the list of clusters of *G*, which are shown on the left in Fig. 3 and the seventh in the list of clusters of *G*′, which are shown on the right in Fig. 3. We show in the picture all the steps needed to align the cluster of *G* with the cluster of *G*′. From top to bottom in this figure, we can see that AligNet first aligns the centers of the clusters, which are the nodes highlighted in blue. Then, AligNet aligns the neighbors of the centers (second row). Next, AligNet aligns the neighbors of the neighbors. In each step we show in a different colour the nodes that are aligned in the present step. Notice that, in this example, there are two nodes that remain unmatched

with v' such that their intermediate nodes are already aligned in sequential order along the paths. The alignment procedure gives priority to matching neighbors of nodes x, x' at the possible shortest distance of the respective cluster centers and with $F(x, x')$ as large as possible among those pairs already matched at the same iterative step.

Step 3. Matching between families of clusters. Let

$$\mathcal{A} = \left\{\eta_{u,u'} \mid u \in V,\ u' \in V', B(u,u') > 0\right\}$$

be the set of alignments obtained in step 2. The *score* of each $\eta_{u,u'} \in \mathcal{A}$ is defined as

$$Score(\eta_{u,u'}) = \frac{\sum_{v \in Dom\, \eta_{u,u'}} B(v, \eta_{u,u'}(v))}{|Dom\, \eta_{u,u'}|} + \frac{|Dom\, \eta_{u,u'}|}{max_{\eta_{w,w'} \in \mathcal{A}} |Dom\, \eta_{w,w'}|}.$$

This score assesses simultaneously the average similarity of the sequences of the proteins matched by $\eta_{u,u'}$ and their number.

Once computed all these scores, AligNet obtains a matching between $C(G)$ and $C(G')$ by applying the maximum weighted bipartite matching algorithm to the bipartite graph whose nodes are the clusters in $C(G)$ and $C(G')$, whose edges connect pairs of clusters $C_u \in C(G)$ and $C_{u'} \in C(G')$ with $B(u,u') > 0$, and the weight of the edge connecting C_u with $C_{u'}$ is the score $Score(\eta_{u,u'})$. We shall denote by $\mathcal{C}$ the set of partial injective mappings $\eta_{u,u'}$ corresponding to pairs of clusters $(C_u, C_{u'})$ that are matched by this matching. Figure 5 shows the matching obtained in this step between the families of clusters in Fig. 3.

Step 4. Local alignment of PPIN. In this step, AligNet produces a local alignment between G and G' from the matching between $C(G)$ and $C(G')$ obtained in the previous step.

The main idea is to define this alignment by merging the partial injective mappings $\eta_{u,u'} \in \mathcal{C}$. The problem is that these mappings may be inconsistent. A first approach to overcome this problem would be to consider the weighted bipartite hypergraph with set of nodes $V \sqcup V'$ and where every mapping $\eta_{u,u'}$ defines a hyperarc with source its domain, target its image, and weight $Score\left(\eta_{u,u'}\right)$, and to solve on it the weighted bipartite hypergraph assignment problem, whose solution would provide a well-defined local alignment of the input networks.

However, in order to decrease the computation time of AligNet, we do not define this hypergraph from the whole $\mathcal{C}$, but just from a subset $\mathcal{R}$ of *best-scored* alignments built recursively as follows. Starting with $\mathcal{R} = \emptyset$, AligNet adds to $\mathcal{R}$ at each step a mapping $\eta_{w_0,w_0'} \in \mathcal{C}$ with w_0 not belonging to the union of the domains of the mappings $\eta_{w,w'}$ already in $\mathcal{R}$ and with maximum $Score\left(\eta_{w_0,w_0'}\right)$ among all such mappings. AligNet iterates this procedure until every node in $\bigcup_{\eta_{u,u'} \in \mathcal{C}} Dom\, \eta_{u,u'}$ belongs to the domain of some mapping in $\mathcal{R}$. In Fig. 6 we give the subset $\mathcal{R}$ of $\mathcal{C}$ for the networks in our running example.

Then, Alignet obtains from the directed hypergraph with nodes $V \sqcup V'$ and hyperarcs defined by the mappings $\eta_{u,u'} \in \mathcal{R}$ as explained above, a local well-defined alignment between G and G' as a solution of the corresponding weighted bipartite hypergraph assignment problem [17]. Figure 7 shows the local alignment obtained from the hypergraph corresponding to Fig. 6.

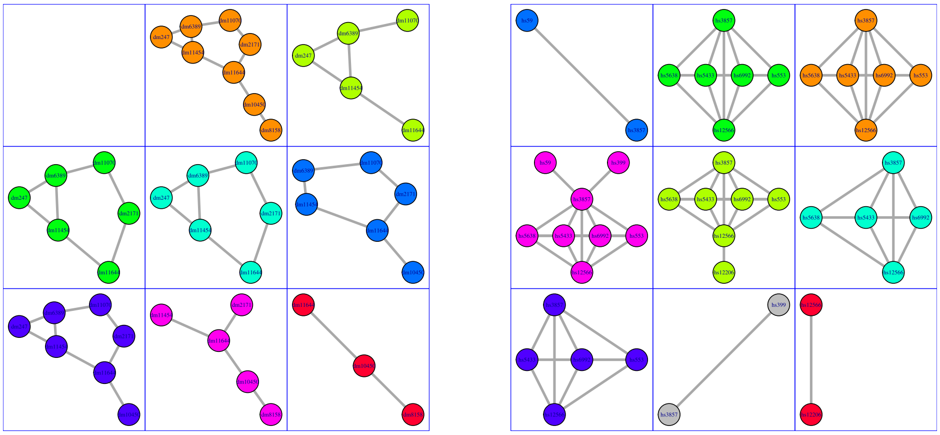

Fig. 5 Alignment of the clusterings. This figure shows the final assignment (same colour) between the clusters in Fig. 3 produced by AligNet, which corresponds also to Step 3. Each of the eight clusters obtained from G is aligned to one, and only one, of the nine clusters obtained from G'. Hence, one cluster from G' remains unmatched which is the second cluster in the third row on the right in Fig. 3. In this figure, we show the clusters from G on the left and its corresponding cluster image from G' on the right

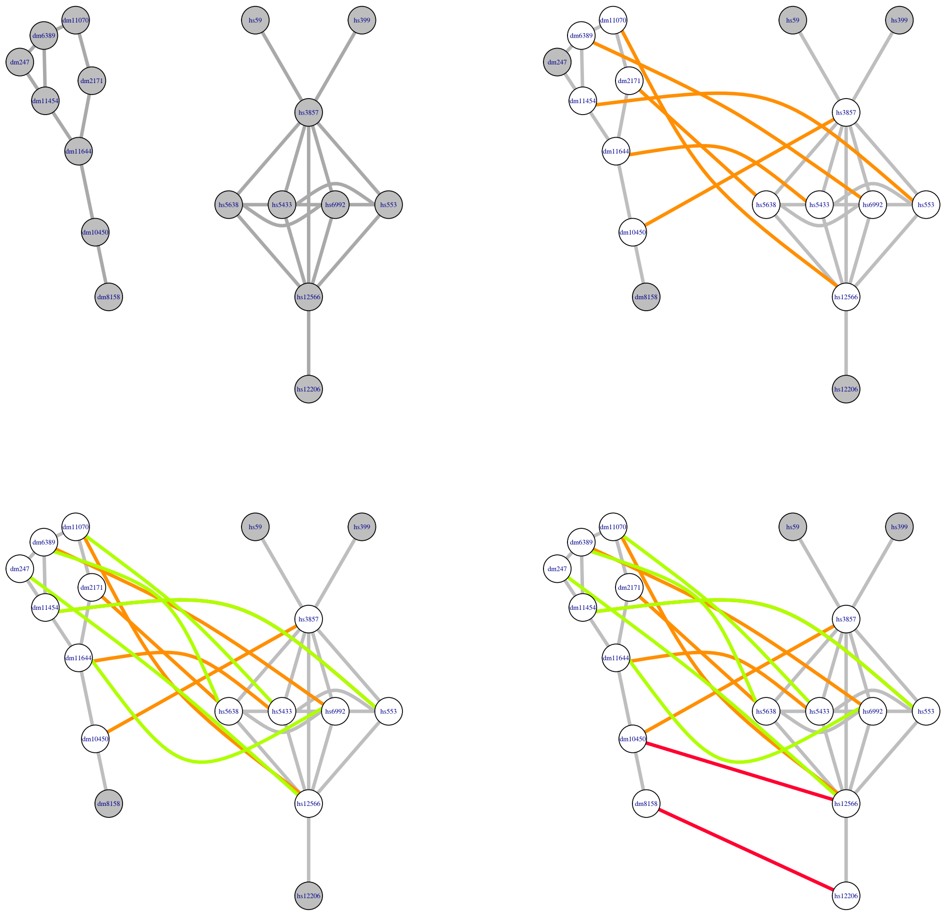

Fig. 6 Appropriate set of alignments. This figure shows how AligNet constructs an appropriate set of alignments considered to obtain a final local alignment. This corresponds to the Step 4 of our aligner. First of all, a maximum score alignment between a pair of clusters is chosen: in this case, this corresponds to the matching between the clusters in Fig. 4. Both clusters are shown in the second row of this figure. The shadowed nodes are the nodes that are not aligned. Next, a maximum score alignment of a pair of clusters with source a cluster centered at a shadowed node is chosen: it turns out to be the one in the second row in Fig. 5 and it is shown in the third row in this figure. Finally, the last alignment to be included in the appropriate set of alignments must be the one with source cluster centered at the remaining shadowed node: this corresponds to the alignment in the last row in Fig. 5 shown in the bottom of this figure. Notice that in the end, that is when we consider the three alignments together, there are four nodes in the source network with inconsistent assignments

Step 5. Global meaningful alignment of PPIN. In order to extend the local alignment produced in the previous step, AligNet iterates the following procedure:

- It removes the nodes in G and G' that have already been aligned, and it recomputes the score of each alignment $\eta_{u,u'}$ following the same definition as in step 3, but only taking into account the remaining nodes in its domain and image.
- It computes a new optimal matching $\mathcal{C}$ between $C(G)$ and $C(G')$, as in step 3, but using as edges those $\eta_{u,u'}$ whose updated score is positive, and weights these updated scores.
- It computes a new set $\mathcal{R}$ of best-scored alignments $\eta_{u,u'}$ with $Score(\eta_{u,u'}) > 0$, as in step 4.

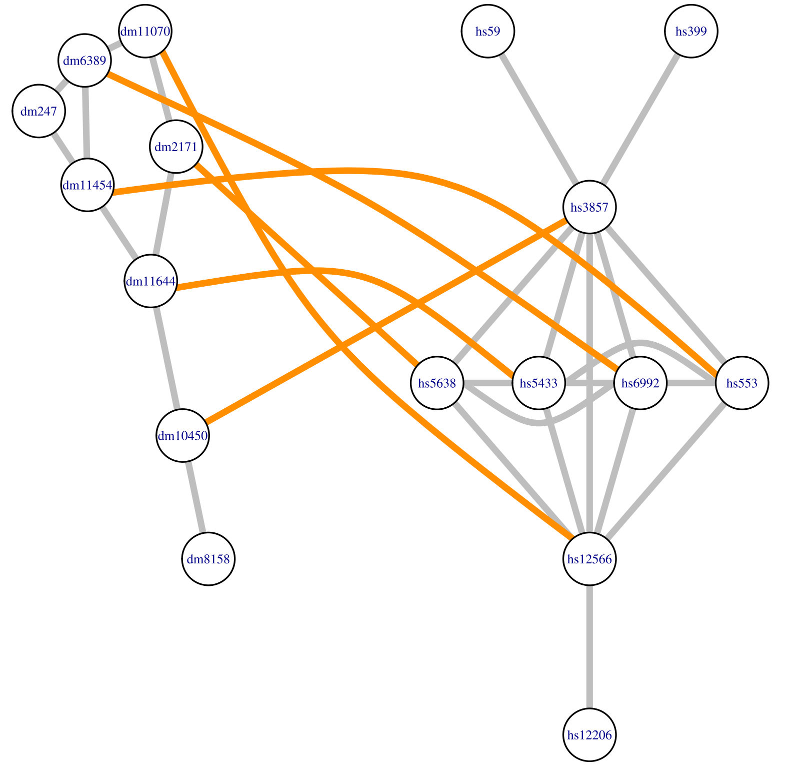

Fig. 7 Local alignment. This figure shows the local alignment of the original networks obtained by AligNet in its fourth step, once the inconsistent assignments have been solved. The coherent assignment of nodes is obtained as the solution to the weighted bipartite hypergraph assignment problem, for the hypergraph associated to the appropriate set of alignments described in Fig. 6. In this case, the hypergraph has three hyperarcs, corresponding to the three alignments considered in the appropriate set of alignments

- It defines a new directed hypergraph whose nodes are the nodes in $V \cup V'$ not yet aligned and hyperarcs the mappings $\eta_{u,u'}$ in the new set $\mathcal{R}$, understood as hyperarcs with source the still unaligned nodes in their domain and target the still unaligned nodes in their image.
- It computes a local alignment between unaligned nodes in V and V' by solving the weighted bipartite hypergraph assignment problem for this hypergraph, and it adds this local alignment to the alignment obtained so far.

This procedure is iterated while there exist nodes not aligned belonging to the domain or the image of some alignment $\eta_{u,u'}$ with (updated) positive score. In Fig. 8 we show the final global meaningful alignment obtained with AligNet for the networks in our running example.

Results

In this section we report the tests performed to assess the performance of AligNet. Following the comparisons published in [12, 13], we decided to compare AligNet with SPINAL [7], HubAlign [8], NATALIE [18], L-GRAAL [19], and PINALOG [20] on the dataset used in [12], which consists of the PPIN of *M. musculus* (mus), *C. elegans* (cel), *D. melanogaster* (dme), *S. cerevisiae* (sce), and *H. sapiens* (hsa), downloaded from the IsoBase database [21] (version 1.0.2); see Table 1. Unfortunately, we

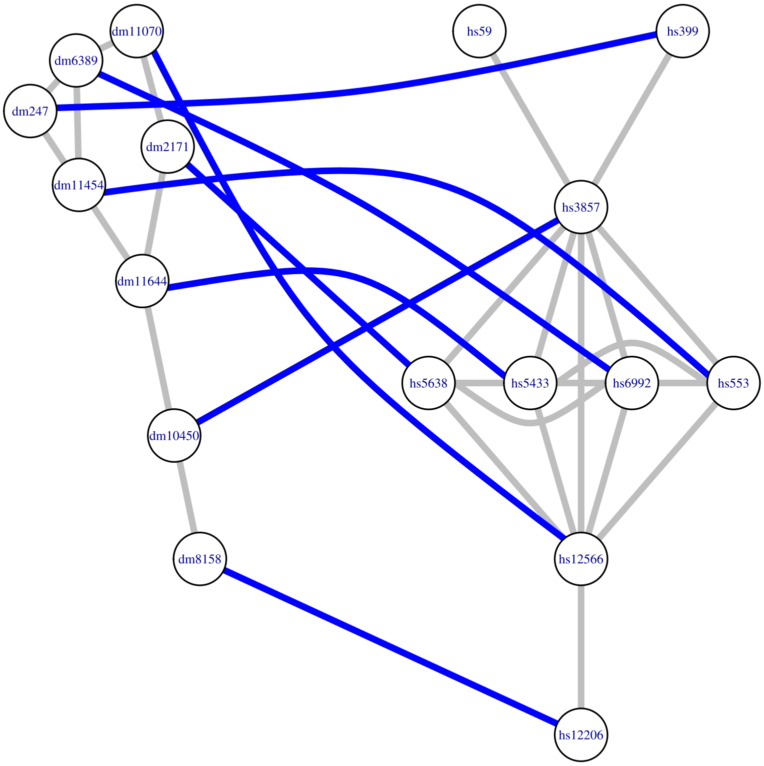

Fig. 8 Final global alignment. This figure shows the final global alignment of the original networks obtained by AligNet. Notice that, in the fifth step of AligNet, the previous alignment is extended to a global one. In this case, there were two unmatched nodes in the source network in Fig. 7 which are now assigned

had to discard the aligner NATALIE from our tests because some computations did not finish.

In a first assessment of the alignments, we used two quality measures: the *edge correctness ratio* (*EC*), which quantifies the amount of structure preserved by the alignment, and the *functional coherence value* (*FC*), which assesses the functional similarity of the aligned proteins by comparing their *Gene Ontology annotation*. More formally, let $G = (V, E)$ and $G' = (V', E')$ be two PPIN such that $|V| \leq |V'|$ and let $\mu : V \to V'$ be a mapping defining an alignment. The *edge correctness ratio* of μ is

$$EC(\mu) = \frac{\left|\left\{\{u,v\} \in E : \{\mu(u), \mu(v)\} \in E'\right\}\right|}{min\{|E|, |E'|\}}$$

Table 1 Number of nodes and edges (with and without loops) of the PPIN considered as input data in our tests

	Nodes	Edges (with loops)	Edges (without loops)
M. musculus	623	776	559
C. elegans	2,995	8,639	4,827
S. cerevisiae	5,524	164,718	82,656
D. melanogaster	7,396	49,467	24,937
H. sapiens	10,403	105,232	54,654

Table 2 Edge correctness ratio obtained in every alignment

Net1	Net2	AligNet	HubAlign	L-GRAAL	PINALOG	SPINAL
mus	cel	0.58	0.81	0.79	0.34	0.01
mus	sce	0.65	0.97	0.68	0.56	0.05
mus	dme	0.65	0.88	0.70	0.30	0.03
mus	hsa	0.76	0.95	0.77	0.62	0.24
cel	sce	0.24	0.83	0.38	0.30	0.06
cel	dme	0.31	0.68	0.53	0.18	0.01
cel	hsa	0.31	0.77	0.43	0.23	0.01
sce	dme	0.03	0.01	0.08	0.19	0.03
sce	hsa	0.04	0.03	0.13	0.19	0.04
dme	hsa	0.13	0.37	0.31	0.13	0.01
mean		0.37	0.63	0.48	0.30	0.05

and the *functional coherence value* of μ is

$$FC(\mu) = \frac{\sum_{u\in V} FS(u,\mu(u))}{|V|},$$

where the similarity score FS is defined by

$$FS(u,u') = \frac{|GO(u)\cap GO(u')|}{|GO(u)\cup GO(u')|},$$

with $GO(u)$ and $GO(u')$ the sets of GO annotations of the proteins u and u', respectively.

Tables 2 and 3, as well as Figs. 9 and 10, report the EC and FC scores of the alignments, respectively. These scores are produced by the aligners under consideration using the aligners' parameters suggested by default whenever it was needed. Because all alignments attained a very low FC score, to put these low scores in perspective, we estimated the maximum value FC_{max} of the FC score for every pair of networks. This maximum value FC_{max} was obtained solving the maximum weighted bipartite matching problem, where the complete bipartite graph had the proteins as nodes and the weight of each edge connecting one protein in a network to a protein in the other network was the FC score of the corresponding pair of proteins. These maximum values are listed in Table 3. We observe that they are very low, being around 0.2 in most computations. Also, we observe in Tables 2 and 3, that AligNet and HubAlign obtained the best balance between FC and EC scores followed by PINALOG and L-GRAAL.

Table 3 Functional coherence value obtained in every alignment

Net1	Net2	FC_{max}	AligNet	HubAlign	L-GRAAL	PINALOG	SPINAL
mus	cel	0.21	0.06	0.04	0.03	0.10	0.12
mus	sce	0.24	0.08	0.07	0.04	0.12	0.15
mus	dme	0.19	0.05	0.03	0.03	0.07	0.06
mus	hsa	0.54	0.23	0.26	0.10	0.48	0.10
cel	sce	0.20	0.06	0.03	0.04	0.13	0.19
cel	dme	0.23	0.04	0.02	0.02	0.09	0.09
cel	hsa	0.24	0.04	0.02	0.03	0.08	0.08
sce	dme	0.24	0.05	0.07	0.02	0.07	0.10
sce	hsa	0.26	0.06	0.08	0.02	0.09	0.11
dme	hsa	0.20	0.04	0.02	0.02	0.09	0.08
mean		0.26	0.07	0.06	0.04	0.13	0.11

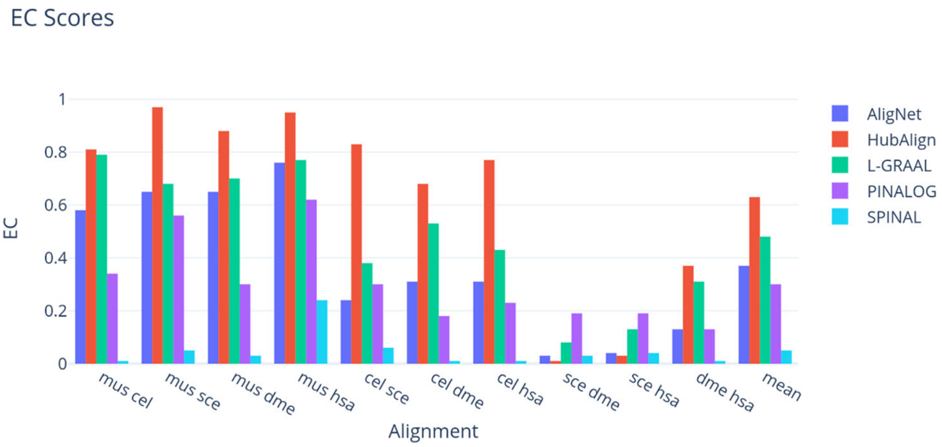

Fig. 9 Edge Correctness Scores. This figure shows the edge correctness score obtained for each aligner in every alignment. The different aligners are presented in different colours

In addition, in our first test and in order to measure the amount of variation or dispersion of the EC and FC scores used to evaluate the aligners, we introduced some *noise* to the networks by randomly adding and deleting 5% of the edges. For every aligner, we were able to compute 100 new pairwise alignments considering the perturbed networks of *M. musculus* mapped to the perturbed networks of *C. elegans*, *D. melanogaster*, and *S. cerevisiae*. In this way, for every aligner we ended up with a sample of 100 EC and FC scores for each of the alignments mus–cel, mus–sce and mus–dme. In Table 4, the mean of the EC and FC scores as well as their standard deviation are presented. Also, to visualise the scores distribution, we considered violin plots to present the results (See Figures 11,12 and 13). We conclude that small perturbations of the real networks produced small variations of the EC and FC scores.

As a second test, we compared the behavior of AligNet, PINALOG, HubAlign, and L-GRAAL in relation to the alignment of protein complexes (we excluded SPINAL from

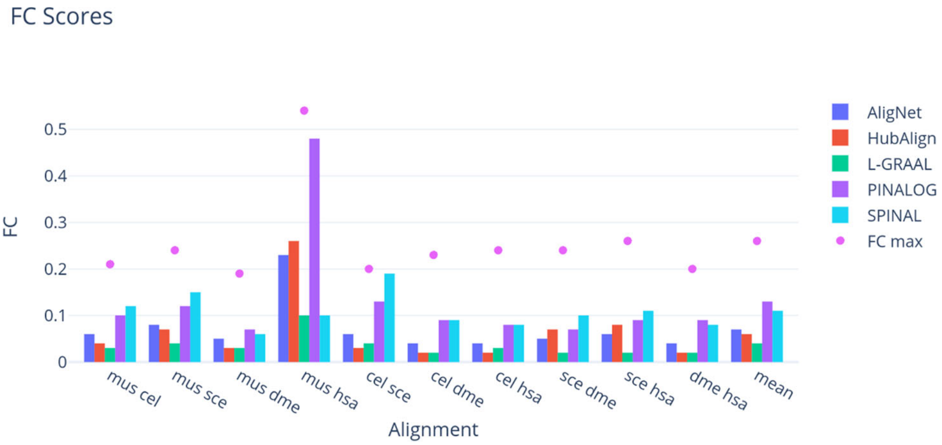

Fig. 10 Functional Coherence Scores. This figure shows the functional coherence score obtained for each aligner in every alignment. In a purple dot we show the maximal value expected for every The different aligners are presented in different colours

Table 4 Statistics of the EC and FC scores

Aligner	Nets	EC_mean	FC_mean	EC_min	EC_max	EC_sd	FC_sd
AligNet	mus_cel	0.5819678	0.0731654	0.5420394	0.6207513	0.0166303	0.0033529
AligNet	mus_dme	0.6583839	0.0516291	0.6118068	0.6923077	0.0158178	0.0023391
AligNet	mus_sce	0.6149481	0.0767457	0.5706619	0.6672630	0.0188964	0.0029397
HubAlign	mus_cel	0.8321288	0.0514488	0.7745975	0.8694097	0.0157104	0.0026933
HubAlign	mus_dme	0.9022209	0.0343045	0.8747764	0.9266547	0.0114605	0.0023421
HubAlign	mus_sce	0.9684609	0.0661089	0.9499106	0.9892665	0.0088473	0.0029502
L-GRAAL	mus_cel	0.7578712	0.0451042	0.7066190	0.8032200	0.0198398	0.0037184
L-GRAAL	mus_dme	0.7770627	0.0246251	0.7137746	0.8318426	0.0209408	0.0023098
L-GRAAL	mus_sce	0.6856139	0.0482183	0.5635063	0.8246869	0.0677820	0.0049054
PINALOG	mus_cel	0.0842755	0.1361287	0.0554562	0.1305903	0.0146922	0.0031677
PINALOG	mus_dme	0.1700990	0.0954297	0.1270125	0.2093023	0.0170064	0.0016534
PINALOG	mus_sce	0.4001862	0.1226443	0.3685152	0.4508050	0.0104970	0.0017084

this test because its results in the EC and FC tests were not convincing). Following the procedure explained in [20], we considered the database MIPS CORUM [22] as the gold standard for the human protein complexes and the information available in [23] as the gold standard for the yeast complexes. In addition, we considered the functional information available in MIPS CORUM for the human complexes and in MIPS FunCat [24] for the yeast complexes. To measure the quality of an alignment in terms of its behaviour on protein complexes, we used the *complex functional coherence value* (CFC), defined as the ratio of complexes that are aligned correctly with respect to the aligned complexes. More specifically, if we call a pair of complexes, one in each network, *coherent* when they share some biological function and *incoherent* otherwise, and if we denote by CP and NCP the numbers of coherent and incoherent pairs of aligned complexes, then $CFC = \frac{CP}{CP+NCP} \times 100$. We report the results obtained by all the aligners in Table 5 and Fig. 14. We observe there that AligNet obtained the highest CFC value (25.34) followed by PINALOG (24.48) whereas HubAlign and L- GRAAL obtained a very low CFC value (5, 4.75 resp.).

In order to further compare the results obtained by AligNet on protein complexes with those of the others aligners, we counted, for each other aligner A, the complexes that were not aligned either by AligNet or by A; the coherent and incoherent pairs among those complexes that were aligned by AligNet but not by A; and the coherent and incoherent pairs among those complexes that were aligned by A but not by AligNet. The

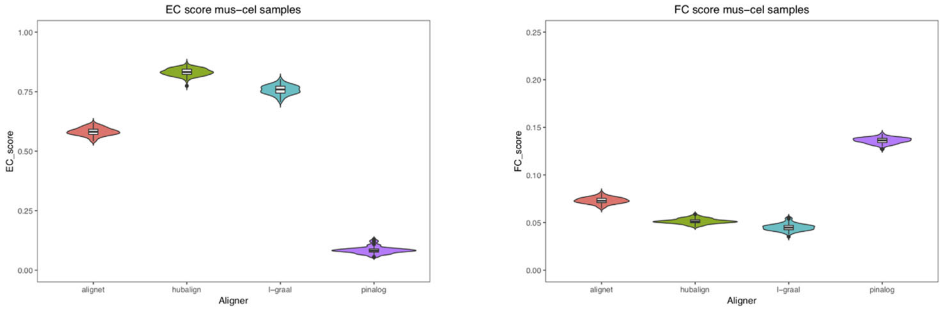

Fig. 11 Scores of mus–cel alignments. This figure shows as violin plots the distribution of the EC and FC scores obtained for every aligner in the alignments of the perturbed networks of mus and cel

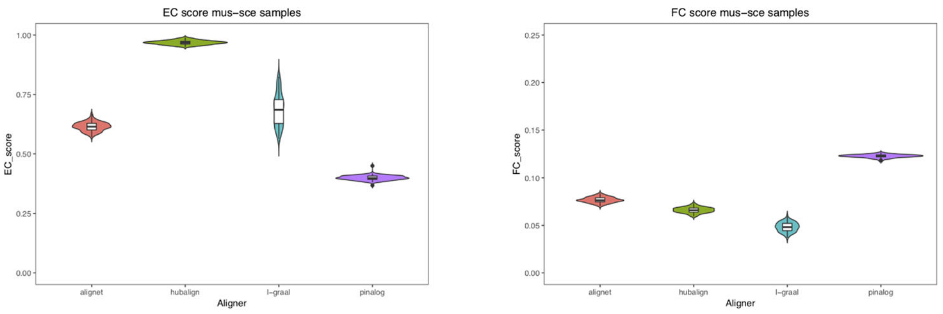

Fig. 12 Scores of mus–sce alignments. This figure shows as violin plots the distribution of the EC and FC scores obtained for every aligner in the alignments of the perturbed networks of mus and sce

results are given in Table 6 and Fig. 15. We observe there that the number of incoherent pairs by HubAlign, L-GRAAL and PINALOG versus AligNet nearly double the number of incoherent pairs by AligNet versus the others.

As a third test to evaluate the aligners, we considered the essential proteins, i.e. those proteins that are indispensable for the survival of an organism, again in the human and yeast PPINs. We evaluate the aligners performance assuming that essential proteins must be aligned to essential proteins. Thus, for every alignment between the PPIN of *S. cerevisiae* and *H. sapiens*, a true possitive (TP) is an essential protein matched to an essential protein while a false positive (FP) is an essential protein matched to a non essential one. In the same way, a true negative (TN) is a non essential protein matched to a non essential one and a false negative (FN) is a non essential protein matched to an essential one. The essential proteins information was retrieved from the DEG Database [25] (http://www.essentialgene.org/). We considered the following statistical measures to evaluate the aligners performance: *specificity* defined by TN/N, *precision* defined by TN/N, F_1*-score* defined by $2TP/(2TP + FP + FN)$, *accuracy* defined by $(TP + TN)/(P + N)$ and *balanced accuracy*, defined by $((TP/P) + (TN/N))/2$, where P and N are the number of essential and non essential proteins respectively in *S. cerevisiae*. Also, we calculated the Pearson correlation of this binary classification problem, called *MCC (Matthews Correlation Coefficient)* defined by

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

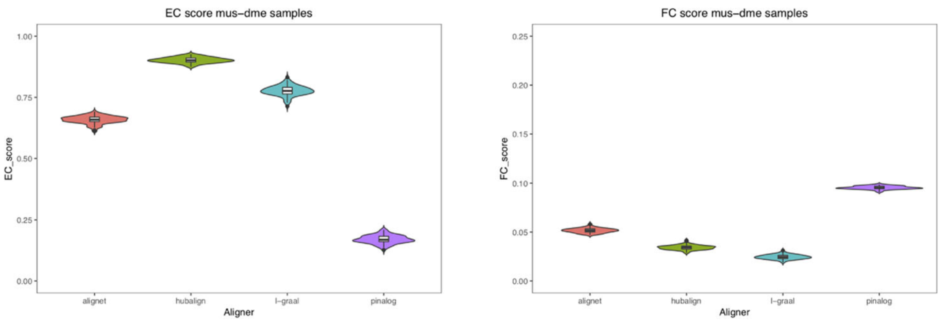

Fig. 13 Scores of mus–dme alignments. This figure shows as violin plots the distribution of the EC and FC scores obtained for every aligner in the alignments of the perturbed networks of mus and dme

Table 5 Number of not assigned, correctly assigned (CP), incorrectly assigned (NCP) protein complexes and the complex functional coherence value obtained for every aligner

	AligNet	PINALOG	HubAlign	L-GRAAL
Not Assigned	1269	945	1154	996
CP	128	203	31	37
NCP	377	626	589	741
CFC	25.34	24.48	5	4.75

and the *proficiency*, also called uncertainty coefficient or entropy coefficient. The uncertainty coefficient in this test is defined as follows: let $\{p_1, \ldots p_n\}$ be the set of proteins in *S. cerevisiae* and let η be an alignment between the two PPIN *S. cerevisiae* and *H. sapiens*. Two random variables X and Y are considered such that, X is a binary vector $X = (x_i)_{i=1,\ldots,n}$ such that x_i takes the value 1 if protein p_i is essential and the value 0 otherwise. Y is a binary vector $Y = (y_i)_{i=1,\ldots,n}$ such that y_i takes the value 1 if protein $\eta(p_i)$ is essential and the value 0 otherwise. Then, the uncertainty coefficient is defined by

$$UC = (H(X) - H(X|Y))/H(X)$$

where $H(X)$ is the entropy of X and $H(X|Y)$ is the conditional entropy. In this test, the uncertainty coefficient measures the capability to predict that a *S. cerevisiae* protein is essential provided that its image by η is essential. In Fig. 16 we show the values for each statistical measure obtained for every aligner. As we can observe there, all aligners have a similar value of accuracy and balanced accuracy. Concerning specificity, precision and F_1-score, HubAlign obtained the lowest value while the others aligners are comparable. The highest proficiency and MCC values were obtained by AligNet while the lowest one was obtained by PINALOG.

Finally, in order to study the efficiency of the considered aligners, we observed their running time and memory space needed to perform an alignment. We run our implementation of AligNet on a server with 4 processors at 2.6 GHz and 20 GB of RAM and we also run the latest implementations of PINALOG (downloaded from http://www.sbg.bio.ic.ac.uk/~pinalog/), SPINAL (downloaded from http://code.google.com/p/spinal/), HubAlign (downloaded from" https://github.com/hashemifar/HubAlign) and also L-GRAAL (downloaded from http://www0.cs.ucl.ac.uk/staff/natasa/L-GRAAL/). NATALIE could not align the two smallest networks,

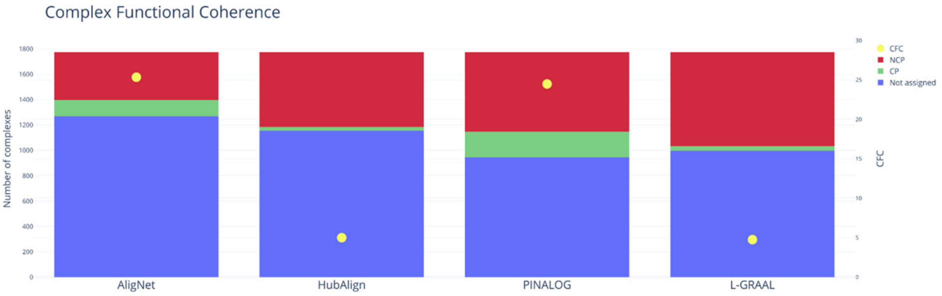

Fig. 14 Complex Functional Coherence. This figure shows the number of non-assigned complexes (in blue), the number of coherent pairs (in green), the number of incoherent pairs (in red) and the complex functional coherence value (yellow dot). The number of complexes is shown on the left axis, while the complex functional coherence value is shown on the right axis

Table 6 Numbers of complexes assigned by AligNet and not assigned by the other aligners, and conversely

	AligNet vs HubAlign	HubAlign vs AligNet	AligNet vs L-GRAAL	L-GRAAL vs AligNet	AligNet vs PINALOG	PINALOG vs AligNet
Not Assigned	891	891	763	763	815	815
Assigned	263	378	233	506	130	454
Incoherent	175	357	167	477	105	375
Coherent	88	21	66	29	25	79
Precision	33.5%	5.6%	28.3%	5.7%	19.2%	17.4%

C. elegans and *D. melanogaster*, on a computer with 64 GB of RAM. PINALOG, SPINAL, HubAlign and L-GRAAL were able to complete all the alignments. In order to visualize their running times, we show the running time of every finished computation for each aligner in the top barplot in Fig. 17. We can observe that AligNet is considerably faster than PINALOG and SPINAL, with a running time of less than 1,000 seconds in most of the alignments. However,it is difficult to see the running times in some alignments because SPINAL needed more than 20,000 seconds for the alignment between *S. cerevisiae* and *H. sapiens*. Thus, in order to visualize the results in the cases where the aligners consumed less than 3,500 seconds, we describe in Fig. 18 the running times cutting at ten minutes. We observe there that HubAlign is the fastest aligner followed by AligNet.

In Fig. 19, we present the running times ordered by the networks size. We observe that in the case of AligNet the running time increases as so do the networks. However, this is not the case of L-GRAAL, SPINAL and HubAlign. On the other hand, PINALOG presents a correlation between networks sizes and running times but it is the slowest aligner. Thus, AligNet is the aligner that present the strongest correlation between running time and networks size.

Discussion

We performed three tests to evaluate and compare our tool AligNet to the best aligners according to [12, 13]. In the first test we assessed the alignment correctness by calculating the EC and FC scores. We present the results in Tables 2 and 3, as well as Figs. 9 and 10. We can observe there that the alignments of small networks with a small number of edges, such as *M. musculus*, produced alignments with high EC scores, especially when the target network has a large number of edges. However, we can also observe that, when the

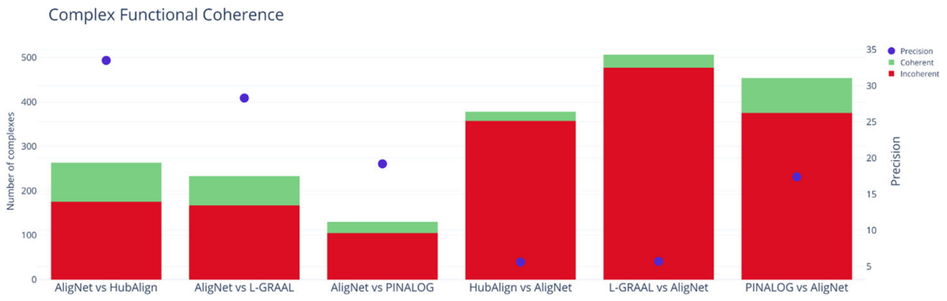

Fig. 15 Complex Functional Coherence Precision. This figure shows the number of coherent pairs (green) and incoherent pairs (red) obtained with one aligner versus the other

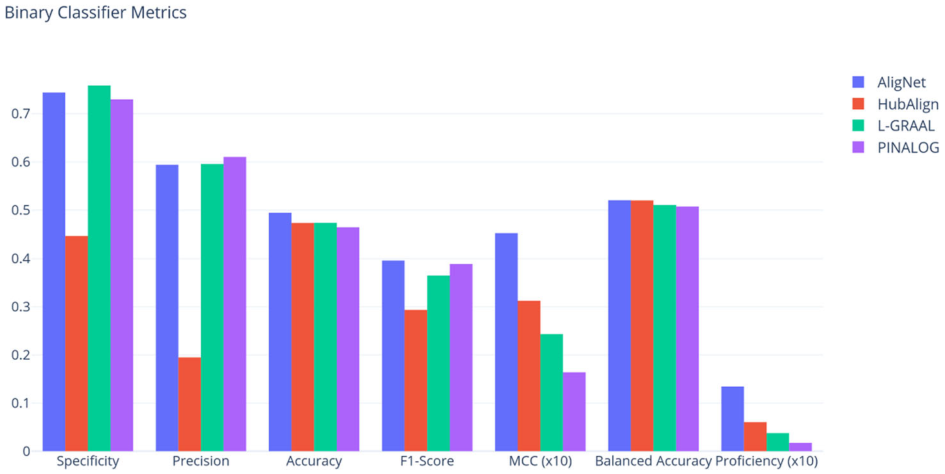

Fig. 16 Binary Classifier Metrics. This figure shows the results obtained for each aligner in the essential proteins alignment test, for every statistical measure

number of edges in the source network increased, the EC scores decreased dramatically even in the case of HubAlign. As far as the functional coherence goes, we can observe in Table 3 and Fig. 10 that all aligners attained a very low FC score whose value in most of the computations is around 0.2 points below the maximum score that can be obtained. An overview to Figs. 9 and 10 reflects that the order from the highest to the lowest EC scores is almost the opposite to the order from the highest to the lowest FC score. That is, the alignment with the highest EC score gets the lowest FC score, being AligNet and HubAlign the aligners that obtained the best balance between FC and EC scores followed by PINALOG and L-GRAAL.

In this first test, we also measured the amount of variation or dispersion of the EC and FC scores used to evaluate the aligners. We introduced some *noise* to the networks by

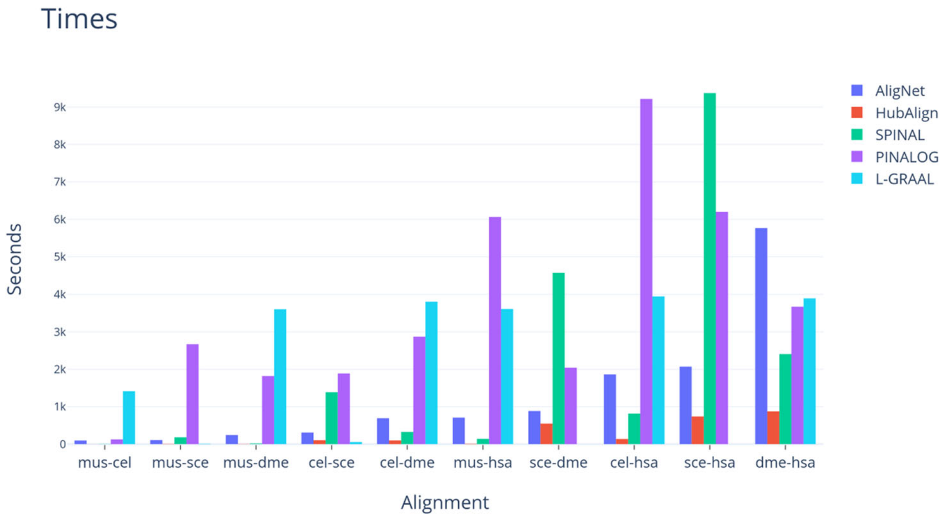

Fig. 17 Running times. This figure shows the running times (in seconds) we obtained when we performed all the alignments for every pair of the considered networks. In this figure we present the results obtained with the aligners AligNet, PINALOG, SPINAL, HubAlign and L-GRAAL

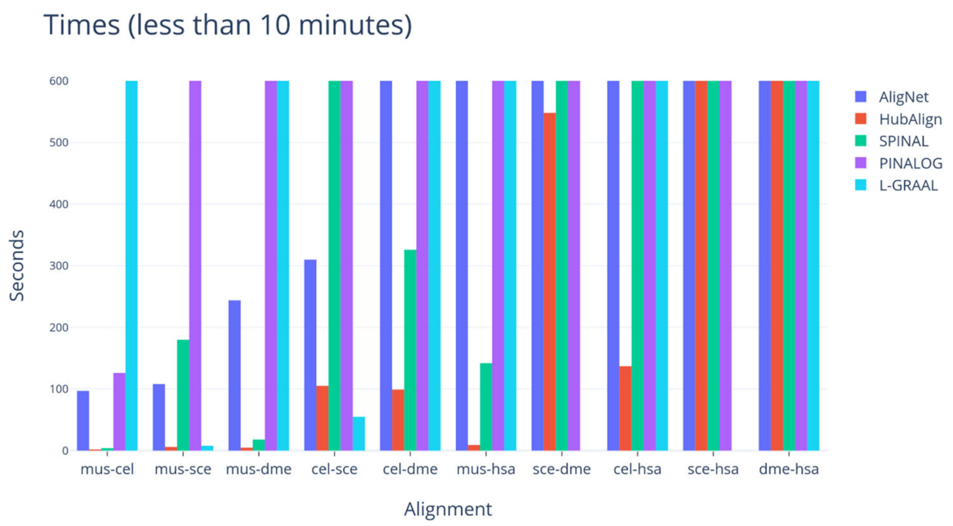

Fig. 18 Running times cut at 10 minutes. We show in this figure the running times for those alignments that took lees than 10 minutes

randomly adding and deleting 5% of the edges. In this way, for every aligner we ended up with a sample of 100 EC and FC scores for each of the alignments mus–cel, mus–sce and mus–dme. In Table 4, the mean of the EC and FC scores as well as their standard deviation are presented. Notice that the differences between the mean of the EC scores obtained by AligNet and HubAlign is around 0.3 being HubAlign the aligner with highest EC scores, while the differences between the mean of the FC scores obtained by AligNet and PINALOG is at maximum 0.05 being PINALOG the aligner with highest FC scores but the lowest EC scores. Thus, the goal of AligNet is accomplished since it clearly obtained the best balance between EC and FC scores. Also, to visualise the scores distribution, in Figures 11,12 and 13 we present the results considering violin plots. In violin

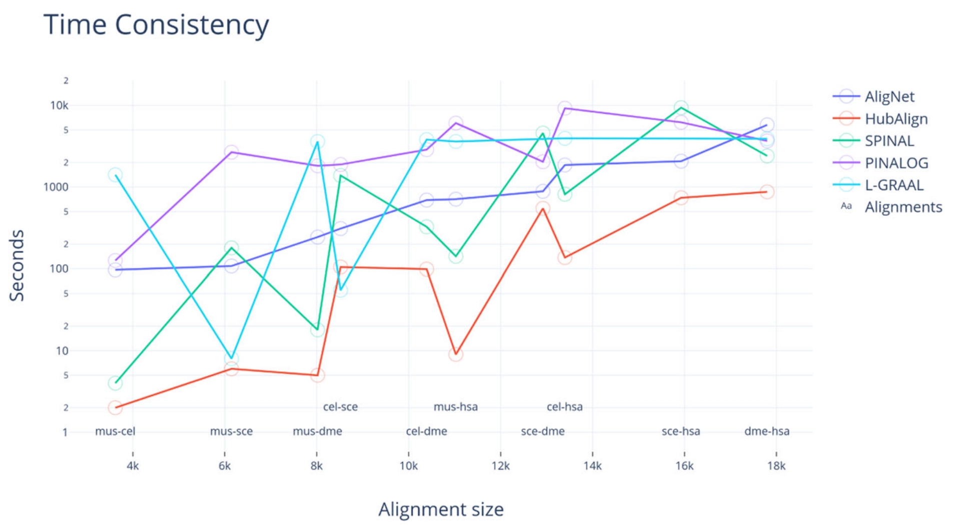

Fig. 19 Time Consistency. This figure shows the running times in seconds obtained for every pairwise alignment and every aligner. We ordered the pairwise alignments considering the size (number of nodes) of the networks

plots we can observe the probability density of the EC and FC scores as well as all the data that is shown in a box plot. As we can observe in these figures, HubAlign and L-GRAAL obtained the highest EC scores but the lowest FC scores in contrast to PINALOG that obtained the lowest EC scores but the highest FC scores. Notice that, the violin's shapes show the scores distribution, that is, flat and wide violins indicate that most of the values are near to the mean in contrast to vertical and narrow violins where the values are dispersed away from the mean. There is only a vertical violin corresponding to the EC scores in the alignments of L-GRAAL between mus and sce. This entails that except for this vertical violin case, small perturbations of the real networks produced small variations of the EC and FC scores.

In the second test we evaluated the alignment of protein complexes. We used the *complex functional coherence value* (CFC) to measure the quality of an alignment in terms of its behaviour on protein complexes. The CFC score is defined as the ratio of complexes that are aligned correctly with respect to the aligned complexes. In Table 5 and Fig. 14 we show the results obtained by all the aligners. AligNet obtained the highest *CFC* value (25.34) followed by PINALOG (24.48) and HubAlign but L- GRAAL obtained a very low *CFC* value (5, 4.75 resp.). In order to further compare the results obtained by AligNet on protein complexes with those of the others aligners, we counted, for each other aligner *A*, the complexes that were not aligned either by AligNet or by *A*; the coherent and incoherent pairs among those complexes that were aligned by AligNet but not by *A*; and the coherent and incoherent pairs among those complexes that were aligned by *A* but not by AligNet. The results are given in Table 6 and Fig. 15. In its first two numerical columns we can see that 891 complexes were not aligned neither by AligNet nor by HubAlign; 263 complexes were aligned by AligNet but not by HubAlign, of which 88 were correctly aligned (coherent pairs) and 175 were incorrectly aligned (by AligNet); and 378 complexes were aligned by HubAlign but not by AligNet, of which 21 were correctly aligned and 357 were incorrectly aligned (by HubAlign). Therefore, HubAlign aligned more complexes than AligNet, but AligNet achieved a higher precision in the alignment of complexes than HubAlign: 33.5% vs 5.6%. In a similar way, AligNet also showed a higher precision than L-GRAAL and a slightly higher precision than PINALOG (19.2% vs 17.4%), although PINALOG aligned more complexes than AligNet. Our interpretation is that AligNet is more conservative than PINALOG.

In the third test we evaluated the alignment of essential proteins in the human and yeast PPINs. We evaluated the aligners performance assuming that essential proteins must be aligned to essential proteins and we compute seven statistical measures. In Fig. 16 we show the values for each statistical measure obtained for every aligner. As we can observe there, all aligners have a similar value of accuracy and balanced accuracy. Concerning specificity, precision and F_1-score, HubAlign obtained the lowest value while the others aligners are comparable. The highest proficiency and MCC values were obtained by AligNet while the lowest one was obtained by PINALOG.

Finally, one of the weak points of PPIN aligners is their lack of efficiency. Indeed, as we have already mentioned, although NATALIE was suggested as a good aligner, it could not align the two smallest networks, *C. elegans* and *D. melanogaster*, on a computer with 64 GB of RAM. With respect to PINALOG, SPINAL, HubAlign and L-GRAAL, we were able to complete all the alignments. In order to visualize their running times, we show the running time of every finished computation for each aligner in Fig. 17. We can observe there that SPINAL is, with a big difference, the slowest one to compute the alignment between *H. sapiens* and *S. cerevisiae*, and also between *D. melanogaster* and *S. cerevisiae*. PINALOG is the slowest one, also with a big difference, to compute the alignment between *C. elegans* and *H. sapiens*, as well as the alignment between *H. sapiens* and *M. musculus*. AligNet is considerably faster than PINALOG and SPINAL, with a running time of less than 1,000 seconds in most of the alignments. Only in one computation, the alignment between *D. melanogaster* and *H. sapiens*, AligNet is slower than PINALOG and SPINAL, and the difference is less than 2,000 seconds. Because SPINAL needed more than 20,000 seconds for the alignment between *S. cerevisiae* and *H. sapiens*, it is difficult to see the running times in some alignments . Thus, in order to visualize the results in the cases where the aligners consumed less than 3,500 seconds, we show in Fig. 18 the running times cutting at ten minutes. We observe there that HubAlign is the fastest aligner. Thus, we conclude that HugAlign is the fastest aligner followed by AligNet.

We also present the running times ordered by the networks size in Fig. 19. It should be expected that the running time increases as so do the networks, and this is the case of AligNet. However, this is not the case of L-GRAAL, SPINAL and HubAlign. Actually, L-GRAAL shows an unpredictable running time related with the networks size. In sum, HubAlign is clearly the faster aligner but the correlation between the networks size and running times is not clear. PINALOG presents a correlation between networks sizes and running times but it is the slowest aligner. And AligNet present the strongest correlation between running time and networks size and it is faster than PINALOG.

Conclusions

In this paper we present AligNet, a new method and software tool for the pairwise global alignment of PPIN aimed to produce biologically meaningful alignments by achieving a good balance between structural matching and protein function conservation. AligNet is a parameter-free algorithm that, given two PPIN, produces a consistent alignment from the smaller network, in terms of number of nodes, to the larger network. Its implementation in R is freely available from https://github.com/biocom-uib/AligNet.

In order to assess the correctness of AligNet, we have evaluated the quality of the alignments obtained with it and with the 4 best aligners established in [12, 13], namely: PINALOG, SPINAL, HubAlign, and L-GRAAL. As a result of the comparison between the aligners, we obtained again, as it was the case in [12, 13], that the agreement of the alignments obtained with different aligners is very low. Most global aligners achieved a high node coverage, meaning that the average number of assigned nodes in the source network is high, but all of them obtained a very low biological coherence value. With respect to the topological coherence value, some aligners were able to obtain a high score but it was associated with a low biological coherence score. Overall, we can conclude that AligNet is the aligner that obtained the best balance between topological coherence (it preserves 60% of the edges) and functional coherence (relative function coherence values between 20% and 40% and the highest complex functional coherence score, 25.34). PINALOG obtained similar functional coherence scores than those of AligNet, lower topological coherence scores and the lowest proficiency value. HubAlign and L-GRAAL obtained high topological coherence scores but very low CFC values. SPINAL surprisingly obtained a very low topological coherence value. Thus, we recommend Alignet to preserve the biological function, and to preserve the topological structure.

Abbreviations

PPIN: Protein-Protein Interaction Networks; EC: Edge Correctness; FC: Functional Coherence; CFC: Complex Functional Coherence; CP: Coherent Pairs; NCP: Incoherent Pairs; TP: True Positive; FP: False Positive; TN: True Negative; FN: False Negative; MCC: Matthews Correlation Coefficient; UC: Uncertainty Coefficients

Acknowledgements

We thank Gabriel Riera for the technical support.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 21 Supplement 6, 2020: Selected articles from the 15th International Symposium on Bioinformatics Research and Applications (ISBRA-19): bioinformatics*. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-6.

Authors' contributions

ML, FR and GV conceived and coordinated the study, performed data analysis and drafted the manuscript. RA and AA performed all the bioinformatic analyses. All authors read and approved the final manuscript.

Funding

This research was partially supported by the Spanish Ministry of Science, Innovation and Universities and the European Regional Development Fund through projects DPI2015-67082-P and PGC2018-096956-B-C43 (FEDER/ MICINN/AEI). Publication costs are funded by Spanish Ministry of Economy and Competitiveness and European Regional Development Fund project PGC2018-096956-B-C43 (FEDER/ MICINN/AEI). The funding body did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All R and Python scripts used in this paper, as well as all data used in it, are available at the GitHub repository page https://github.com/biocom-uib/AligNet and http://bioinfo.uib.es/~recerca/AligNet/

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Mathematics and Computer Science, University of the Balearic Islands, E-07122 Palma de Mallorca, Spain. [2]Balearic Islands Health Research Institute (IdISBa) E-07010 Palma de Mallorca, Spain. [3]Algorithms, Bioinformatics, Complexity and Formal Methods Research Group, Technical University of Catalonia, E-08034 Barcelona, Spain.

Received: 13 August 2019 Accepted: 16 April 2020 Published: 18 November 2020

References

1. Kelley BP, Yuan B, et al. PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res. 2004;32(Web Server issue):W83–88.
2. Koyutürk M, Kim Y, et al. Pairwise alignment of protein interaction networks. J Comput Biol. 2006;13(2):182–199.
3. Li Z, Wang Y, et al. Alignment of protein interaction networks by integer quadratic programming. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2006. p. 5527–30.
4. Liang Z, Xu M, Teng M, Niu L. NetAlign: a web-based tool for comparison of protein interaction networks. Bioinformatics. 2006;22(17):2175–7.
5. Narayanan M, Karp RM. Comparing protein interaction networks via a graph match-and-split algorithm. J Comput Biol. 2007;14(7):892–907.
6. Elmsallati A, Clark C, Kalita J. Global alignment of protein-protein interaction networks: A survey. IEEE/ACM Trans Comput Biol Bioinforma. 2016;13(4):689–705.
7. Aladağ AE, Erten C. SPINAL: Scalable protein interaction network alignment. Bioinformatics. 2013;29(7):917–24.
8. Hashemifar S, Xu J. HubAlign: an accurate and efficient method for global alignment of protein-protein interaction networks. Bioinformatics. 2014;30(17):i438–44.
9. Neyshabur B, Khadem A, Hashemifar S, Arab SS. NETAL: a new graph-based method for global alignment of protein-protein interaction networks. Bioinformatics. 2013;29(13):1654–62.
10. Patro R, Kingsford C. Global network alignment using multiscale spectral signatures. Bioinformatics. 2012;28(23): 3105–14.
11. Singh R, Xu J, Berger B. Global alignment of multiple protein interaction networks with application to functional orthology detection. PNAS. 2008;105(35):12763–8.
12. Clark C, Kalita J. A comparison of algorithms for the pairwise alignment of biological networks. Bioinformatics. 2014;30(16):2351–9.
13. Malod-Dognin N, Ban K, Pržulj N. Unified alignment of protein-protein interaction networks. Sci Rep. 2017;7(953):.
14. Alain FZ, Elena NI, Erik HWG. Meesters. A Beginner's Guide to R: Springer; 2009.

15. Camacho C, Coulouris G, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):1.
16. Kuhn HW. The Hungarian method for the assignment problem. Naval Res Logist. 2005;52(1):7–21.
17. Borndörfer R, Heismann O. The hypergraph assignment problem. Discrete Optim. 2015;15:15–25.
18. Klau GW. A new graph-based method for pairwise global network alignment. BMC Bioinformatics. 2009;10(1):S59.
19. Malod-Dognin N, Pržulj N. L-GRAAL: Lagrangian graphlet-based network aligner. Bioinformatics. 2015;31(13):2182–9.
20. TPhan HT, Sternberg MJE. PINALOG: A novel approach to align protein interaction networks—implications for complex detection and function prediction. Bioinformatics. 2012;28(9):1239–45.
21. Park D, Singh R, et al. IsoBase: a database of functionally related proteins across PPI networks. Nucleic Acids Res. 2011;39(suppl 1):D295–D300.
22. Ruepp A, Brauner B, et al. CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Res. 2008;36(suppl 1):D646–D650.
23. Gavin A-C, Aloy P, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440(7084): 631–6.
24. Ruepp A, Zollner A, et al. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res. 2004;32(18):5539–45.
25. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. DEG 10, an update of the Database of Essential Genes that includes both protein-coding genes and non-coding genomic elements. Nucleic Acids Res. 2014;42:D574–D580.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.