TÜBİTAK

Turkish Journal of Biology

http://journals.tubitak.gov.tr/biology/

Research Article

Turk J Biol
(2013) 37: 547-555
© TÜBİTAK
doi:10.3906/biy-1210-8

Global assessment of network inference algorithms based on available literature of gene/protein interactions

Gökmen ALTAY[1,*], Nejla ALTAY[2,3], David NEAL[2]

[1]Department of Biomedical Engineering, Bahçeşehir University, Beşiktaş, İstanbul, Turkey
[2]Cancer Research UK Cambridge Research Institute, Department of Oncology, University of Cambridge, Cambridge, UK
[3]Department of Mathematics, Bahçeşehir University, Beşiktaş, İstanbul, Turkey

Received: 04.10.2012 • **Accepted:** 04.03.2013 • **Published Online:** 06.09.2013 • **Printed:** 04.10.2013

Abstract: We propose a framework that uses the available gene/protein interaction databases of the literature as a universal benchmark in order to globally assess the inference performances of gene network inference algorithms. We also developed an R software package for convenient use of the framework, which can also be used in general as a quick tool to search in the literature for available validations of interactions. We applied the proposed approach to 2 publicly available prostate cancer gene expression datasets and a large breast cancer gene expression dataset. The results revealed different aspects and superiority of algorithms that had not been compared previously in the available literature. Our approach allowed the assessing and comparing of the algorithms on a real dataset of a size of around 30,000 probes, which showed the strengths and weaknesses of the algorithms from different points of view rather than conventional approaches. We further show that our approach provides a unique advantage in assessing the performance of an inference method when applied to a new dataset and thus sheds light on the results of a de novo application, which would be obscure without our approach.

Key words: Gene network inference, bioinformatics, gene/protein interaction databases

1. Introduction

Gene network inference (GNI) algorithms allow inferring gene interactions from an expression dataset, which estimates the genome-wide working mechanism of genes and helps elucidate cell physiology and pathogenesis (Rual et al., 2005; Schadt, 2009; Altay and Emmert-Streib, 2010a). Having an accurate estimate of the genome-wide network map allows more subtle drug development, as it specifies direct targets of gene interactions of interest. Nonetheless, this task is extremely difficult, as it is currently typical to come across gene expression datasets with around 50,000 probes and 1000 samples. This number gets higher when working on the intron and exon levels, which may have approximately 600,000 probes. Accumulated experimental and computational noise in such a large-scale dataset causes many false positives (FPs) and reduces the accuracy of the estimated gene network with a GNI algorithm. Therefore, it is essential to know the inference performance of a GNI method before employing it on a new dataset in a biological experiment. The inference performance of gene network inference algorithms is usually assessed with a synthetic and a few real biological datasets. This may cause variations in the performances of the methods over different datasets and makes it difficult to assess the algorithms. However, there is no consensus on a single algorithm that performs best in general because the performance may vary as the datasets vary (Altay and Emmert-Streib, 2010b; Narendra et.al, 2011). Therefore, a methodology is needed to be able to assess the performance of a GNI algorithm when applied to a de novo dataset. Similarly, it would be useful to be able to compare various GNI methods on a de novo dataset, as the performances may change.

In this paper, we set up a framework and provide an R software package with the ability to make this kind of global assessment via the available literature of interaction databases. Since the proposed framework allows a global assessment of networks predicted by any GNI algorithm independent from the dataset used, we call it *ganet* hereafter for ease of writing and reading.

Using *ganet*, one can assess the performance of any de novo application of a GNI method on any new real dataset of any size. We implemented *ganet* on 5 popular GNI methods that are known to infer large-scale gene networks and compared them based on the available literature. For the global assessments and comparisons of the GNI methods, we used 2 different gene expression

* Correspondence: gokmen.altay@bahcesehir.edu.tr

datasets of probes with sizes of around 30,000 and 12,000, which are the highest sizes of real datasets used for global assessments of GNI methods, although it is possible to use any size of dataset using the proposed approach. We elaborate the details of the comparisons and *ganet* in the following sections. In summary, our approach can be used for a global assessment of the predictions of gene/protein interactions of inference algorithms employing the literature.

2. Materials and methods

The block diagram of the proposed framework, *ganet*, for assessing the inference performances of GNI algorithms, is illustrated in Figure 1. A gene network is estimated by a GNI method and it is compared with the available interaction literature to assess its accuracy. Since the available literature is far from being complete, we use precision as well as the significance of the overlap between the 2 networks to get a complete picture of the relative accuracy of the predictions.

We are aware of the fact that the available literature might show interactions of completely different biological conditions, but predicting an edge with a significant correlation value between the gene pairs and having experimental support from any other biological condition provides a strong assumption for accepting these interactions as a true positive (TP) and computing the performance accordingly. In fact, these supposedly validated interactions of other biological conditions appear as the strongest candidate for performing experiments to validate and show if they are also interacting in the biological condition of the current dataset, which may then provide new insights on the biology of interest.

In order to determine that the overlapping interactions from the inferred set were not caused by chance alone, or in other words, to compute the significance of the overlap, we employed a widely used statistical method, Fisher's exact test (Fisher, 1934), that utilizes the hypergeometric distribution of overlapping probability (Fury et al., 2006). Figure 2 and Table 1 are presented to better explain the usage of Fisher's exact test for our global assessment of inference performances. As seen from the figures, given 2 networks (e.g., inferred and literature), we determine whether the overlap is statistically significant or not with respect to all the possible gene interactions in the universe of the dataset under consideration. Table 1 helps to better classify the regions shown in Figure 2 and used as entries in Fisher's exact test in R.

2.1. Databases in *ganet*

We combined all the major gene/protein interaction databases in the literature. Our ultimate goal is to unify all the available gene/protein interactions in the literature and provide a single combined database. The software also allows accessing and using any database combination available in the package in R, which is the common platform of the community. The interaction databases have different biological meanings. One reason for this is because interaction of genes might be obtained via various experimental methods. Regarding the GNI methods, we can classify the databases into 2 categories: first, the interaction databases that include gene/protein interactions obtained via a biochemical experiment for each gene pair; and second, the pathway databases that include groups of genes that share the same biological pathway and have been identified via a biochemical experiment. In the former set, the interactions do not necessarily mean that the gene pairs are causally interacting; they may be just sharing the same biological function. Since this is also a kind of interaction, although a weak one, we also include those databases as optional since they have been obtained

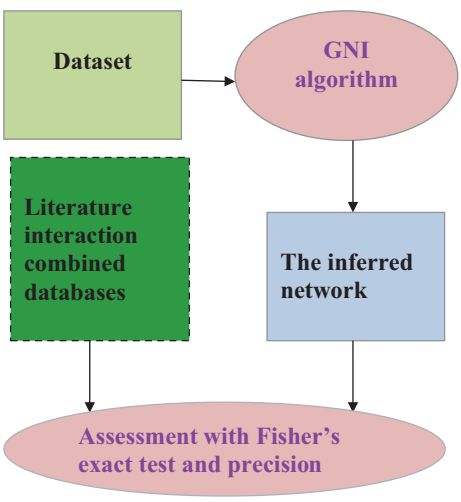

Figure 1. Block diagram of the proposed framework.

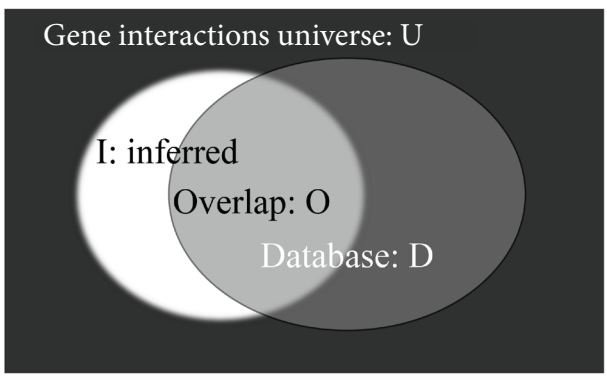

Figure 2. Illustration of the overlap for Fisher's exact test. Here, I denotes the number of inferred links of a GNI method. D is the number of links in the literature database to search for validations. O denotes the amount of overlap between I and D. U is the possible number of interactions considering the genes of the dataset under consideration.

Table 1. Parameters of Fisher's exact test.

	In the literature	Not in the literature
Inferred	O	(I – O)
Not inferred	(D – O)	(U – I – D + O)

I is the number of inferred links. O is the number of overlapping links. D is the number of links in the database with only the gene names in the dataset used for prediction.

through expensive biological experiments and might be useful for assessing GNI methods, too.

We do not include any computational databases in *ganet* that were obtained via estimates of gene/protein interactions, as they may be biased due to the estimator algorithm used and thus are not suitable for GNI method assessments. However, as a third classification, we also included the kind of database that mines the literature from the internet via a software program rather than manually. Below, we go briefly through all the included databases that we downloaded prior to November 2011.

2.1.1. BioGRID: Biological General Repository for Interaction Datasets (Stark et al., 2006)

BioGRID is an online interaction repository that includes protein and genetic interactions, with data compiled through comprehensive curation efforts. It has 258,644 unique links and can be downloaded from http://thebiogrid.org/, but it is continuously being updated. That will be reflected in *ganet*.

2.1.2. HPRD: Human Protein Reference Database (Keshava et al., 2009)

HPRD has 43,463 unique links and may be downloaded from http://www.hprd.org/ and http://www.pathwaycommons.org. All the information in HPRD has been manually extracted from the literature by expert biologists who read, interpret, and analyze the published data.

2.1.3. MINT: Molecular Interaction Database (Ceol et al., 2009)

MINT has 296,806 unique links and may be downloaded from http://www.pathwaycommons.org. MINT focuses on experimentally verified protein–protein interactions mined from the scientific literature by expert curators and is available at http://mint.bio.uniroma2.it/mint.

2.1.4. IntAct

IntAct provides protein interaction data that are populated by either curating from the literature or from direct data depositions (Kerrien et al., 2012). It has 174,011 unique links and may be downloaded from http://www.pathwaycommons.org; IntAct is also available at http://www.ebi.ac.uk/intact/.

2.1.5. BCI: B-Cell Interactome (Wang et al., 2006; Lefebvre et al., 2007)

Only biologically validated interactions are included in BCI. It has 14,547 unique links and may be downloaded from http://amdec-bioinfo.cu-genome.org/html/BCellInteractome.html.

These 5 databases are the major databases that we suggest using when assessing the global inference performance of a GNI algorithm that infers gene/protein causal interactions network rather than just an association network. In the following, we go through the other databases that are mainly pathway databases in the form of gene/protein interactions. When assessing a GNI algorithm that infers associations as well, combining all the available databases in the package is more suitable, and thus all are included in the package.

2.1.6. IMID: General Repository for Interaction Datasets

IMID has 1533 unique links and may also be downloaded from http://www.pathwaycommons.org (http://www.sbcny.org; November 2011).

2.1.7. CELL: The Cancer Cell Map

CELL contains selected cancer-related signaling pathways (http://cancer.cellmap.org/cellmap; November 2011).

NAT: Nature Pathway Interaction Database

NAT includes bimolecular interactions and cellular processes assembled into authoritative human signaling pathways. It has 12,944 unique links and may be downloaded from http://www.pathwaycommons.org (http://pid.nci.nih.gov/PID/index.shtml; November 2011).

2.1.8. REACTOME

REACTOME is an open-source, open-access, manually curated and peer-reviewed pathway database (http://www.reactome.org; November 2011). It is not a pathway, but rather is a curated database with an automated computational approach and is thus deemed to be less important.

2.1.9. ID-serve

ID-serve (Ramani et al., 2005) is an automated literature-mined human protein interaction dataset and denoted as UniHI in *ganet*. It has 31,182 unique links.

3. Results and discussion

The performances of GNI algorithms are mostly assessed based on synthetic networks via simulations of kinetic equations, which with arguable accuracy mimic all types of biological gene expression datasets. Synthetic analysis is useful as it allows the comparison of methods with a complete underlying true network of the synthetic dataset and is thus somewhat useful when any new GNI method is introduced. On the other hand, biologists are mostly interested in seeing performances based on real biological

datasets with their true interaction networks obtained via biochemical validations. In order to see the performance of methods in real applications, a few such biological datasets are being used individually (e.g., *E. coli* and *S. cerevisiae* species that can be downloaded from the M3D database of Faith et al. (2007)). However, these are only for a few species and are incomplete datasets, to be employed individually. Many algorithms with parameters set for these real networks may cause bias when applied to other biological datasets. Our global assessment approach, *ganet*, which combines the available gene/protein interaction databases, might help to assess the performances of GNI algorithms on a global level and might thus be considered as the most general approach for the most general interaction databases available in the literature. Bioinformatics tools are required to reach useful conclusions from biological datasets (Varışlı and Çen, 2007; Demir et al., 2012), and so we provide a bioinformatics tool for using *ganet*. Along with this paper, we also provide the R software package, also named *ganet*, as a convenient tool to apply the proposed framework. We have combined all major interaction databases, but we aim to continue adding more databases as we get feedback from the community and as updates of the current ones are released.

In order to demonstrate the benefits and contributions of *ganet*, we have applied it to 5 different GNI algorithms and have compared these algorithms based on the available literature or, to put it another way, compared with what is already known as global preknowledge in the literature. We selected mutual information based on popular GNI methods, as they are able to predict networks from very high-throughput datasets, and their applications are available in R software. These methods are CLR (Context Likelihood of Relatedness; Faith et al., 2007), Relevance Network (RELNET) (Butte et al., 2000), ARACNE (Margolin et al, 2006), MRNET (Meyer et al., 2007), and C3NET (Altay and Emmert-Streib, 2010a), which are widely cited in the literature and are available to use in the R platform (Meyer et al., 2008; Altay and Emmert-Streib, 2011). We refer interested readers to the work of Altay and Emmert-Streib (2010a, 2010b) for an overview of these methods.

Assessments of methods are performed using metrics, usually with area under receiver operating characteristics (AUROC), F-score, or area under precision-recall curve (AUPR) (Altay and Emmert-Streib, 2010b; Narendra et al., 2011). Nonetheless, in our approach, since the interaction databases of the literature are far from being complete, these may cause too many FPs and thus are not suitable metrics. The only somewhat suitable metric from among the conventional ones is the precision (precision = TP / (TP + FP)), as it shows the rate of accuracy (TPs) of the predicted network with respect to what is available in the literature. Therefore, in our approach, this metric is used along with a statistical overlapping test (e.g., Fisher's exact test) to decide if the number of TPs that overlap with the available literature is significant or whether it is due to chance alone. We illustrate the block diagram and describe the details of the proposed framework in Section 2. Below, we show the application results of the proposed approach, *ganet*, using its implementation software on the selected GNI methods, as presented in Table 2.

None of these GNI methods have previously been assessed with respect to what is available in the literature, apart from a few relatively small databases of a particular species. Our results reveal very different aspects of the inference performances of the methods, which complements the results of other conventional assessment methods. The application results in Table 2 are based on the prostate cancer dataset with 12,558 probes. Please see the details of the datasets in Section 4.2. C3NET followed by MRNET gave the best performances, and then ARACNE, CLR, and RELNET, in that order, when assessed based on conventional global approaches as presented by Altay and Emmert-Streib (2010a, 2010b). Here our results somewhat coincide with their observations using the precision metric, as the methods were ranked with MRNET, C3NET, CLR, ARACNE, and RELNET from best to worst. Very close scores of MRNET and C3NET also appeared in our case, but we additionally reveal that the number of predicted

Table 2. Performance results of *ganet* on various GNI methods with the first prostate dataset.

GNI method	Fisher's test P-value	Precision	#TP	#Predicted edges
RELNET	0	0.0075	40,101	5,322,600
ARACNE	2.64 × e-7	0.012	2455	196,683
MRNET	1.68 × e-29	0.027	207	7552
CLR	2.27 × e-109	0.018	2542	141,206
C3NET	3.43 × e-25	0.026	192	7339

The results are obtained by combining all the available databases in *ganet*.

links of MRNET is small, similar to C3NET, which suggests that MRNET is also a conservative inference algorithm, as C3NET is. We also show here that CLR has a better precision score than ARACNE and that they both predict a very large number of links. As expected, RELNET has the lowest precision since it infers the highest number of predicted links. Precision itself is widely used, but the precision scores we present here are novel, as they are ideally based on what is available in the current literature. Precision alone is not sufficient to assess performance, as we also need to know whether it is statistically significant. Having said that, when considering the significance of the overlap, the ranking of the performances changes to RELNET, CLR, ARACNE, MRNET, and C3NET. The dramatic change in the ranking of the performances reveals a very novel aspect of the performance of these algorithms, which has not been demonstrated before. We now know that even the worst method, RELNET, can infer more significant numbers of TPs at the cost of more FPs. We observe that if the number of inferred interactions increases, it becomes possible to infer a higher number of significant TP interactions. However, we show that not all the algorithms are able to provide this. We observed that ARACNE has the highest P-value, though significant, despite the fact that it has inferred the highest number of interactions after RELNET. We then observed the close P-values of MRNET and C3NET in the overlapping analysis, too, which confirms the close performance of these methods. The main conclusion of our performance analysis comes by considering the 2 performance scores together. Considering Table 2 in total, we have reached the following conclusions. First, we conclude that there are many more TP interactions to be inferred than the current methods allow, and thus there is much more to be done in the area of developing GNI algorithms. Second, if one wishes to be flexible and infer a very large gene network, then CLR appears to be better performing overall than ARACNE. Third, however, if one wishes to be conservative and infer a small network with higher accuracy, then MRNET and C3NET are similarly good alternatives, considering the relevant comparison results of Altay and Emmert-Streib (2010a) as well.

We now compare the GNI algorithms on a larger prostate cancer expression dataset with 26,107 probes, which is, to the best of our knowledge, the highest probe size dataset that has been used for comparing GNI algorithms' global inference performances. We illustrate the results in Table 3. The results have revealed very interesting conclusions, and it would not be wrong to say that the larger the size of the benchmark dataset, the better the performance comparisons are, as it magnifies the comparisons and discloses some hidden disadvantages of the GNI methods (Altay, 2012). The observations of CLR, ARACNE, and RELNET are similar, as in the previous finding, with better P-values for CLR and ARACNE. Again, in this application, C3NET provides the highest precision, but with a much higher P-value. However, we discovered that MRNET is not a practical GNI method to apply to a very large probe dataset, because the program did not give a result after 4 days. We terminated the program as it was not clear when we would get an output, and our focus is on practical algorithms that can be applied to real datasets with current technology. We have been using a MAC computer with 32 GB memory and 2.26 GHz processors. Regarding the duration that each algorithm takes, RELNET is the slowest of all the algorithms; it is not included in this comparison and that duration is not counted. For the rest of the required time, C3NET is the fastest, with a result in 1 min; CLR and ARACNE took 59 and 66 min, respectively.

In order to have more general results, we further evaluated the approach over a very recent breast cancer dataset, which is the largest available (Curtis et al., 2012). The results of simulations to measure performance using *ganet* are shown in Table 4. In fact, this new, very large-scale dataset showed that it is not possible to implement CLR on very large datasets as memory becomes insufficient, although we used a desktop computer with very high memory (32 GB). MRNET was seen again to be computationally impractical as it did not respond after a week of waiting to run the simulations. C3NET, ARACNE,

Table 3. Performance results of *ganet* on various GNI methods with the larger prostate dataset.

GNI method	Fisher's test P-value	Precision	#TP	#Predicted edges
RELNET	0	0.00133	3223	2,410,074
ARACNE	1.59 × e-20	0.0075	1209	161,057
MRNET	NA	NA	NA	NA
CLR	1.09 × e-163	0.018	765	41,755
C3NET	6.97 × e-79	0.025	240	9367

The results are obtained by combining all the available databases in *ganet*.

Table 4. Performance results of *ganet* on various GNI methods with the breast cancer dataset.

GNI method	Fisher's test P-value	Precision	#TP	#Predicted edges
RELNET	0	0.0145	3249	224,035
ARACNE	0	0.0031	361	113,090
C3NET	4.83e-192	0.0323	218	6730

The results are obtained by combining all the available databases in *ganet*.

and RELNET were able to run in reasonable durations and provided consistent results with previous performance-assessing simulations. The performance results matched the previous performances with respect to the success rank of the algorithms, considering only the 3 that were able to work on the very large-scale dataset. Additionally, using the same dataset, we provide individual performances of each of the databases in Tables 5–7 for various GNI algorithms.

Overall, considering the precisions and P-values of the presented tables together, we can conclude that if one wishes to be conservative with higher accuracy in inferring the gene network, then C3NET should be used for any dataset size. If one wishes to be less stringent and infer as large as possible a network, then we suggest using CLR.

As demonstrated in this study, our proposed approach, along with the easy-to-use R package *ganet*, is expected to be an essential additional framework for globally assessing the inference performance of any new GNI algorithm when introduced. It allows assessing the performance on any real dataset of any size based on ideally what is known in the literature. It also provides a common platform to compare all methods with a unique benchmarking network, i.e. the literature. Furthermore, it provides the unique opportunity to assess the performance of the application of a GNI method on a de novo dataset, which helps detect possible errors during the implementation, because the methods have many steps and parameters that are open to error.

Other efforts to provide a common platform to assess GNI methods include the DREAM conference series, which is a kind of competition (Stolovitzky et al., 2009). However, as argued by Narendra et al. (2011), this series has not provided a definitive answer as to what the best performing techniques are for genome-scale observational data. In addition, the assessments are mostly based on synthetic datasets. The only real gene expression data used for the assessment of the methods was in DREAM2 (Stolovitzky et al., 2009). It has been argued that as this competition involved a single dataset to which many methods had been applied, the results may be over-fitted and thus may not generalize to other datasets (Narendra et al., 2011).

Therefore, our proposed framework and the software *ganet* are a significant contribution to global assessment of networking methods on a single platform that is independent from the dataset.

3.1. Implementation: usage example of *ganet*

The R software package to implement *ganet* is currently available for download at https://sites.google.com/site/gokmenaltay/ganet, and it is planned for the software to also be available in CRAN following publication. The provided source file can be installed on all platforms (Mac, Windows, Linux). To install the latest R software, set the working directory in the same folder that *ganet* is placed. Then, in the R console, write the following: *install.packages("GAnet_1.0.tar.gz", type="source")*. Make sure first that the dependent *igraph* package is already installed. Then run: *library(GAnet)*, and you are ready to use *ganet.*

In order to run examples on *ganet*, we have given an exemplary list of interactions, named *ganet.ex.net*, that is used instead of the inferred gene network of a GNI algorithm. It is in the form of a matrix of 2 columns, and each row corresponds to an interaction of a gene pair. This matrix must consist of a unique set of rows while entering the functions of the software. Now we check to see if the predicted interactions are significantly validated by the HPRD (Human Protein Reference Database). We first load the networks from the package for the exemplary run in the R console as follows: *data(ganet.ex.net); data(hprd)*. Now the overlapped links are derived: *Validated <-ganet.ComLinks(netlist=as.matrix(ganet.ex.net), netdata=as.matrix(hprd))*. We now need the number of unique links in HPRD considering the genes available in our dataset. For this exemplary run, we label those genes from the package *data(ganet.ex.genes)*. We can now compute the universe of interactions possible with the genes of the dataset used. The maximum value of the universe can be computed with *nUniverse <- length(ganet.ex.genes)*(length(ganet.ex.genes) - 1)/2*. This number can be safely used while comparing the performance of GNI methods. However, if we are applying this to a de novo dataset with an algorithm and wish to see individual performance, then we suggest using at most half of this number, as it would not be correct to assume that all the genes might interact with each other in a biological condition. Now the significance of overlapping statistics of this validated list of interactions can be computed as

Table 5. Performance results of *ganet* using C3NET on the breast cancer dataset for each interaction database.

Database	Fisher's test P-value	Precision	Recall	#TP	#Predicted edges	#Edges in database
Combined	4.83e-192	0.0323	0.00047	218	6730	463,434
BCI	7.27e-45	0.0043	0.002	29	6730	14,270
BioGrid	2.68e-34	0.0054	0.00048	37	6730	76,629
HPRD	2.83e-71	0.008	0.00127	54	6730	42,484
INTACT	2.03e-11	0.0028	0.00019	19	6730	98,394
MINT	7.02e-78	0.0141	0.0003	95	6730	239,046
UniHI	8.14e-136	0.012	0.0029	81	6730	27,086
IMID	1	0	0	0	6730	1477
CELL	2.27e-06	0.0004	0.0031	3	6730	965
NAT	9.68e-37	0.0035	0.0019	24	6730	12,595
REACTOME	5.83e-43	0.0047	0.0012	32	6730	25,083

The results are obtained via the available databases in *ganet*.

Table 6. Performance results of *ganet* using ARACNE on the breast cancer dataset for each interaction database.

Database	Fisher's test P-value	Precision	Recall	#TP	#Predicted edges	#Edges in database
Combined	0	0.018	0.0043	2038	113,090	463,434
BCI	3.09e-205	0.0016	0.0131	188	113,090	14,270
BioGrid	3.82e-119	0.002	0.003	237	113,090	76,629
HPRD	3.97e-268	0.0027	0.0074	316	113,090	42,484
INTACT	3.27e-73	0.0018	0.002	204	113,090	98,394
MINT	0	0.0086	0.004	980	113,090	239,046
UniHI	0	0.0038	0.016	435	113,090	27,086
IMID	0.025	2.6e-05	0.002	3	113,090	1477
CELL	0.0008	3.5e-05	0.004	4	113,090	965
NAT	1.21e-243	0.0018	0.0163	206	113,090	12,595
REACTOME	0	0.0031	0.0143	361	113,090	25,083

The results are obtained via the available databases in *ganet*.

Table 7. Performance results of *ganet* using RELNET on the breast cancer dataset for each interaction database.

Database	Fisher's test P-value	Precision	Recall	#TP	#Predicted edges	#Edges in database
Combined	0	0.0145	0.007	3249	224,035	463,434
BCI	0	0.0014	0.0224	320	224,035	14,270
BioGrid	1.24e-158	0.0016	0.0049	378	224,035	76,629
HPRD	0	0.0024	0.0126	538	224,035	42,484
INTACT	5.08e-91	0.0014	0.0032	324	224,035	98,394
MINT	0	0.0066	0.0062	1494	224,035	239,046
UniHI	0	0.0027	0.0227	615	224,035	27,086
IMID	0.34	8.9e-06	0.0013	2	224,035	1477
CELL	0.00018	2.6e-05	0.0062	6	224,035	965
NAT	0	0.0025	0.0301	380	224,035	12,595
REACTOME	0	0.0031	0.022	569	224,035	25,083

The results are obtained via the available databases in *ganet*.

follows: *res <- ganet.FEtest(nOverlap=nrow(Validated), nPredicted=nrow(ganet.ex.net), nFocusedSet=nFocusedSet, nUniverse=nUniverse)*. In order to see the results, we write *res$stats$p.value*, which outputs 5.35e-43, and *res$precision*, which outputs 0.0725. In general, in order to consider the P-value to be statistically significant, it is expected to be lower than 0.05. In this example, the overlap is considered as very significant. The precision shows that 7.2% of the predicted interactions of the network are validated by the HPRD, and from the P-value we decide that this number is statistically significant and not by chance alone. In this example, if in considering the P-value we had concluded that the validated interactions were not significant enough, then we should check the GNI method used to determine if it has been properly applied, because those results may also be caused by an erroneous set-up of the GNI method while running. As we see, this example shows another unique advantage of *ganet*: it not only assesses the performance of GNI methods, but it also provides a feedback mechanism that gives the opportunity to check whether the implementation of a GNI method has been performed correctly for each new dataset. One can also combine some or all of the databases in the package to assess the performance of GNI methods. This is the ideal usage of *ganet*, as it allows for finding the validations of the prediction network from all of the literature available from the major interaction databases. In order to combine all the available databases in *ganet*, run the following: *CombinedDatabases <- ganet.combine(Bci=1, BIOGRID=1, HPRD=1, INTACT=1, MINT=1, UNIHI=1, IMID=1, CELL=1, NAT=1, REACT=1)*. Here, 1 means include and 0 means exclude the database while combining. We have now demonstrated the usage of the main functions of *ganet*. One can refer to the help section of the *ganet* functions for details of the package.

3.2. The biological dataset

In order to perform exemplary assessments of the inference performances of the popular GNI algorithms on real and large biological datasets, we first used the dataset of Singh et al. (2002) in Table 2. This dataset, which has 12,558 probes and 8802 unique genes and was used to study clinical prostate cancer behavior, is a widely used dataset in the literature. High-quality expression profiles of 52 prostate tumor samples were used in our application. The raw data were downloaded from http://www-genome.wi.mit.edu/MPR/prostate. The dataset was preprocessed as described in detail by Altay et al. (2011). We also used the larger and more recent prostate cancer dataset of Taylor et al. (2010) with 26,107 probes and 18,547 unique genes. We used 131 primary tumor samples of this dataset. Primary tumor samples of the recent breast cancer dataset of Curtis et al. (2012) were also used.

Acknowledgments

The R software packages C3NET (Altay et al., 2011) and minet (Meyer et al., 2008) are used for the applications of GNI methods in R (http://www.r-project.org). G Altay was funded by Cancer Research UK, and most of his work was conducted at the University of Cambridge during his postdoctoral studies. Most of the results of this manuscript are partly from the postdoctoral study of N Altay at the University of Cambridge.

References

Altay G (2012). Empirically determining the sample size for large-scale gene network inference algorithms. IET Syst Biol 6: 35–63.

Altay G, Asim M, Markowetz F, Neal DE (2011). Differential C3NET reveals disease networks of direct physical interactions. BMC Bioinformatics 12: 296.

Altay G, Emmert-Streib F (2010a). Inferring the conservative causal core of gene regulatory networks. BMC Syst Biol 4: 132.

Altay G, Emmert-Streib F (2010b). Revealing differences in gene network inference algorithms on the network-level by ensemble methods. Bioinformatics 26: 1738–1744.

Altay G, Emmert-Streib F (2011). Structural influence of gene networks on their inference: analysis of C3NET. Biol Direct 6: 31.

Butte A, Tamayo P, Slonim D, Golub T, Kohane I (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. P Natl Acad Sci USA 97: 12182–12186

Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G (2009). The molecular interaction database 2009 update. Nucleic Acids Res 37: D532–D539.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y et al. (2012). The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. Nature 486: 346–352.

Demir İ, Eryüzlü E, Demirbağ Z (2012). A study on the characterization and pathogenicity of bacteria from *Lymantria dispar* L. (Lepidoptera: Lymantriidae). Turk J Biol 36: 459–468.

Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biol 5: e8.

Fisher RA (1934). Statistical Methods for Research Workers. Edinburgh, UK: Oliver and Boyd.

Fury W, Batliwalla F, Gregersen PK, Li W (2006). Overlapping probabilities of top ranking gene lists, hypergeometric distribution, and stringency of gene selection criterion. Conf Proc IEEE Eng Med Biol Soc 1: 5531–5534.

Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U et al. (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Res 40: D841–D846.

Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A et al. (2009). Human Protein Reference Database - 2009 Update. Nucleic Acids Res 37: D767–772.

Margolin A, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7: S7.

Meyer P, Kontos K, Bontempi G (2007). Information-theoretic inference of large transcriptional regulatory networks. EURASIP J Bioinform Syst Biol 1: 79879.

Meyer P, Lafitte FK, Bontempi G (2008). minet: r/bioconductor package for inferring large transcriptional networks using mutual information. BMC Bioinformatics 9: 461.

Narendra V, Lytkin NI, Aliferis CF, Statnikov A (2011) A comprehensive assessment of methods for de-novo reverse-engineering of genome-scale regulatory networks. Genomics 97: 7–18.

Ramani AK, Bunescu RC, Mooney RJ, Marcotte EM (2005). Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. Genome Biol 6: R40.

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature 437: 1173–1178.

Schadt E (2009). Molecular networks as sensors and drivers of common human diseases. Nature 461: 218–223.

Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP et al. (2002). Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 1: 203–209.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. (2006). BioGrid: a general repository for interaction datasets. Nucleic Acids Res 34: D535–D539.

Stolovitzky G, Prill RJ, Califano A (2009). Lessons from the DREAM2 challenges. Ann NY Acad Sci 1158: 159–195.

Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B et al. (2010). Integrative genomic profiling of human prostate cancer. Cancer Cell 18: 11–22.

Varışlı L, Çen O (2007). Identification and characterization of the rat DVL2 gene using bioinformatic tools. Turk J Biol 31: 81–86.

Wang K, Banerjee N, Margolin AA, Nemenman I, Califano A (2006). Genome-wide discovery of modulators of transcriptional interactions in human B lymphocytes. Lect Notes Comput Sci 3909: 348–362.