

RESEARCH ARTICLE

CytoGTA: A cytoscape plugin for identifying discriminative subnetwork markers using a game theoretic approach

S. Farahmand^{1,2}, M. H. Foroughmand-Araabi³, S. Goliaei⁴, Z. Razaghi-Moghadam^{1*}

1 Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran, **2** College of Science and Mathematics, University of Massachusetts Boston, Boston, Massachusetts, United States of America, **3** Department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran, **4** University of Tehran, Tehran, Iran

* razzaghi@ut.ac.ir



OPEN ACCESS

Citation: Farahmand S, Foroughmand-Araabi MH, Goliaei S, Razaghi-Moghadam Z (2017) CytoGTA: A cytoscape plugin for identifying discriminative subnetwork markers using a game theoretic approach. PLoS ONE 12(10): e0185016. <https://doi.org/10.1371/journal.pone.0185016>

Editor: Julio Vera, Friedrich-Alexander-Universitat Erlangen-Nurnberg, GERMANY

Received: February 15, 2017

Accepted: September 4, 2017

Published: October 2, 2017

Copyright: © 2017 Farahmand et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying this study are third party data. Interested researchers may apply for access to transcriptome data through publicly available database of GEO and the following link: <http://bioinformatics.nki.nl/data>. The interactome data are accessible by contacting the corresponding author of doi:10.1038/nbt1295. The authors confirm that they do not have special access privileges to these data and that interested researchers may apply to access these data in the manner described.

Abstract

In recent years, analyzing genome-wide expression profiles to find genetic markers has received much attention as a challenging field of research aiming at unveiling biological mechanisms behind complex disorders. The identification of reliable and reproducible markers has lately been achieved by integrating genome-scale functional relationships and transcriptome datasets, and a number of algorithms have been developed to support this strategy. In this paper, we present a promising and easily applicable tool to accomplish this goal, namely CytoGTA, which is a Cytoscape plug-in that relies on an optimistic game theoretic approach (GTA) for identifying subnetwork markers. Given transcriptomic data of two phenotype classes and interactome data, this plug-in offers discriminative markers for the two classes. The high performance of CytoGTA would not have been achieved if the strategy of GTA was not implemented in Cytoscape. This plug-in provides a simple-to-use platform, convenient for biological researchers to interactively work with and visualize the structure of subnetwork markers. CytoGTA is one of the few available Cytoscape plug-ins for marker identification, which shows superior performance to existing methods.

Introduction

It is commonly acknowledged that genetic perturbations in human cells are the main reason of cancer initiation and progression[1]. Using potential cancer biomarkers, as an objective evidence for the diagnosis and monitoring cancer in earlier stages, provides a valuable opportunity for researchers to detect, cure, or at least delay the progression of cancer in human body [2]. Genetic mutations may bring diverse consequences, including conformational alteration in protein structure, loss or serious changes in protein function and also deregulation of gene expression. There have been numerous studies suggesting differentially expressed genes (DEGs) in cancer versus normal samples, as biomarkers [3–5]. However, the efficiency and reproducibility of identified biomarkers rely extensively on sample size, data quality and heterogeneity of experimental platforms utilized for analysis[6].

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

To enhance the efficiency and reproducibility of identified biomarkers, more recent methods have leveraged the assumption that “genes associated to a specific disease, tend to have physical interactions among their products or situate in a particular pathway” [7, 8]. According to this hypothesis, a number of methods confined their analysis to the pathway genes and reported dysregulated pathways with high representation of the DEGs. For example, in an approach proposed by Gue et al. [9] each potentially dysregulated pathway was quantified by the mean and the median of the expression levels of its participated genes. In another study, Tomfohr et al. [10] calculated the activity score of a given pathway by the singular value decomposition technique. Furthermore, Su et al. [11] used a log-likelihood approach to identify potential pathway-based markers.

Regarding the reliability and the accuracy of identified biomarkers, the pathway-based methods mainly overtake the ones based on single genes. However, the main deficiency of pathway-based methods is that, a limited number of genes and significant DEGs may be covered by known pathways. To address this shortcoming, several methods have utilized valuable information embedded in protein-protein interaction networks (PPINs) and integrated gene expression profiles with them [12–18]. The first method, aimed at identifying network-based markers using a simulated annealing algorithm, was proposed by Ideker et al. [16]. Network-based markers which are also referred to as subnetwork markers are a set of mainly significant DEGs with high interconnectivity in a PPIN. The identification of subnetwork markers, with remarkable efficiency and reproducibility, has recently attracted much attention and many studies have been motivated to improve the markers [12–18].

Recently, Farahmand et al. [13] proposed a game theoretic approach (GTA) with the objective of extracting reliable subnetwork markers with potential prognostic utility. The identified subnetwork markers by GTA showed significantly high discriminative power and the proposed strategy of GTA would be applicable to acquire makers linked to a given phenotype. In this paper, a new plugin for Cytoscape [19] (CytoGTA) is presented which uses all GTA features and strengthen to identify optimized subnetwork markers. Cytoscape is an open source tool for integrating, analyzing and visualizing data in the context of networks. CytoGTA plugin takes a PPIN and a gene expression profile of two phenotype classes as input, and returns subnetwork markers that can discriminate two classes. The plugin has been implemented in Cytoscape application, a known graphical environment that is convenient for researchers in systems biology. The outputs of this plugin can be used by other analysis tools within Cytoscape, for further investigations.

Materials and methods

As we aimed to provide an easy-to-use tool for identifying potential subnetwork markers, CytoGTA has been developed as a Cytoscape app, written in Java programming language. CytoGTA is based on GTA algorithm which has been approved for its superior reproducibility and reliability as compared with some well-known algorithms. CytoGTA takes as input a weighted (or possibly unweighted) PPIN and a gene expression profile, and results in numbers of connected subnetworks, which are highly associated with the given expression profile.

The game theoretic approach used in GTA, models the task of extracting subnetwork markers as a static game. In the modeled game, nodes of the PPIN are considered as players trying to optimize their utility by choosing the strategy of joining or leaving subnetworks, which are respectively represented by 1 and 0 in the subnetwork state vector. The utility of each player is quantified by a payoff function, which is a combination of a gain function and a loss function. The Nash equilibrium is then calculated and the result state vector represents optimized subnetwork marker. In the following, a brief description of the gain and loss functions is given.

Gain function

Given any subnetwork $G_s(V_s, E_s)$, there are $|V_s| = n$ proteins corresponding to n genes (Genes = $\{g_1, g_2, \dots, g_n\}$). Also, an expression profile for m different samples of two different phenotypes is available. The expression vector $x_i = (x_i^1, x_i^2, \dots, x_i^m)$ contains expression values of gene $g_i (i = 1, 2, \dots, n)$, where x_i^j is the expression level of gene g_i in sample j . In order to relate the sample phenotypes to their given expression levels, the GTA algorithm computes the log-likelihood ratio (LLR) of each gene to obtain a standard feature vector for that gene. This data is then used in association with phenotype data to determine whether statistically significant correlation can be identified between the expression levels for each gene and the given disease [11]. The LLR of gene g_i for j^{th} sample, due to two different phenotype classes is defined by:

$$LLR_i(x_i^j) = \log \left[\frac{f_i^1(x_i^j)}{f_i^2(x_i^j)} \right] \tag{1}$$

where $f_i^1(x_i^j)$ and $f_i^2(x_i^j)$ are the conditional probability density function (PDF) of the expression level of gene g_i under phenotype 1 and phenotype 2, respectively. The *density* function in Apache Commons Math [20] is used to calculate the PDF of the expression levels under different distributions of two phenotypes. Commons Math is a self-contained Java library provided by Apache Foundation, which contains components for mathematical and statistical routines. The vector $LLR_i = (LLR_i(x_i^1), LLR_i(x_i^2), \dots, LLR_i(x_i^m))$ for gene g_i contains log-likelihood values of gene g_i for m different samples.

A local scoring (LS) function is also defined for each gene g_i . By this scoring, GTA tries to find the role of each protein in the subnetwork in connecting DEGs. The LS function for gene g_i with joining strategy is defined as Eq (2),

$$LS_i = \sum_{t=1}^k t - score(LLR_i) \tag{2}$$

where $g_{i_1}, g_{i_2}, \dots, g_{i_k}$ are k neighbors of gene g_i in G_s , and t -score is t -test statistic score (independent test) for LLR values. Libraries provided by Apache Commons Math is exploited to perform t -test statistical score for two unpaired sample data sets assuming unequal variances. Since a gene with leaving strategy has no neighbor in the subnetwork, its LS value is set to be zero.

Furthermore, to score the connectivity of each subnetwork, a density value is assigned. Suppose $G_{join} = (V_{join}, E_{join})$ is an induced graph of G_s on nodes (players) with joining strategy (V_{join}), then the density value is defined by:

$$DE(G_{join}) = \frac{\sum_{e \in E_{join}} w(e)}{\binom{|V_{join}|}{2}} \tag{3}$$

where $w(e)$ is the weight of edge e (for unweighted PPINs, $w(e)$ is set to be 1 for all edges in E_{join}). The value of $DE(G_{join})$ is assigned to all genes in G_s , irrespective of their chosen strategy.

Finally, the gain function (GF) is determined as Eq (4) for gene g_i in G_s , in which α, β and γ are constants:

$$GF(i, G_s, G_{join}) = \alpha \cdot t - score(LLR_i) + \beta \cdot LS_i + \gamma \cdot DE(G_{join}) \tag{4}$$

In above equation, α, β and γ are weighting parameters to imply each function's importance. $t - score(LLR_i)$ is the t -test statistics score of the LLR_i , which is considered in Eq (4) for gene g_i with any chosen strategy.

Loss function

The loss function (LF) for gene g_i with joining or leaving strategy is defined in Eq (5), where δ is a constant.

$$LF(i, G_s, G_{join}) = \delta \cdot (|V_{join}| - 1) \quad (5)$$

Payoff function

Eventually, the payoff function (PF) for a given agent g_i and the subnetwork G_s is calculated as follows:

$$PF(i, G_s, G_{join}) = GF(i, G_s) - LF(i, G_s) \quad (6)$$

By using a numerical method, GTA algorithm examined different values for constants in payoff function, and the most powerful discriminatory markers were achieved by setting $\alpha = 1.24$, $\beta = 1$, $\gamma = 1$ and $\delta = 2$. To keep the consistency with GTA, in our developed plugin, the same values are assigned to the constants.

Main algorithm

In the developed algorithm, proteins (gene products) within the given PPIN are sorted in decreasing order, regarding their $t - score(LLR)$ values. Starting from the highest ranked protein, each of them is chosen as a seed, excluding the ones having degree less than the average degree of the PPIN. Once a seed is chosen, all proteins at distance at most two from the seed, are identified in the PPIN. The starting seed and its close nodes (at most two steps away from it) are considered as the players of the algorithm and form a candidate subnetwork. In the candidate subnetwork, all one-step neighbors of the seed, are selected and sorted in increasing order, regarding their degrees, and are then iteratively removed from the subnetwork while their removal increases the local clustering coefficient of the seed. The local clustering coefficient of a node is defined by the proportion of edges between its first neighbors divided by the number of all possible edges between them. Subsequently, the same exact procedure is applied on the two-step neighbors, the ones which are connected to the remaining one-step neighbors from the previous step.

A Nash equilibrium in a game describes a state with a set of strategies for players in which no player gain a better payoff by changing its own strategy alone[14]. In our case, Nash equilibrium occurs when all nodes play their best responses to the strategies of others. In other words, in this Nash equilibrium, no node can get a better payoff by changing its chosen strategy (change from joining to leaving or vice versa), unilaterally. For a given subnetwork of size $|V_s|$ and with two available strategies for each of its nodes, there are $2^{|V_s|}$ possible states to be evaluated. Subsequently, for each state, payoffs for every node are calculated, and eventually an equilibrium state is chosen for the subnetwork. In the case of multiple Nash equilibria, the subnetwork with the highest average absolute $t - score(LLR)$ of its genes, is reported as the result.

To overwhelm computational limitations, caused by the number of possible states, each game (candidate subnetwork) is divided into a number of sub-games, where each sub-game has a relatively low number of players (nodes). The Nash equilibrium is computed for each sub-game, and then merged, to form a single optimized subnetwork. Although the identified subnetwork may not represent a global Nash equilibrium of the game, it can be considered as a suitable alternative to it. The overview of the game theoretic approach used in GTA is depicted in Fig 1.

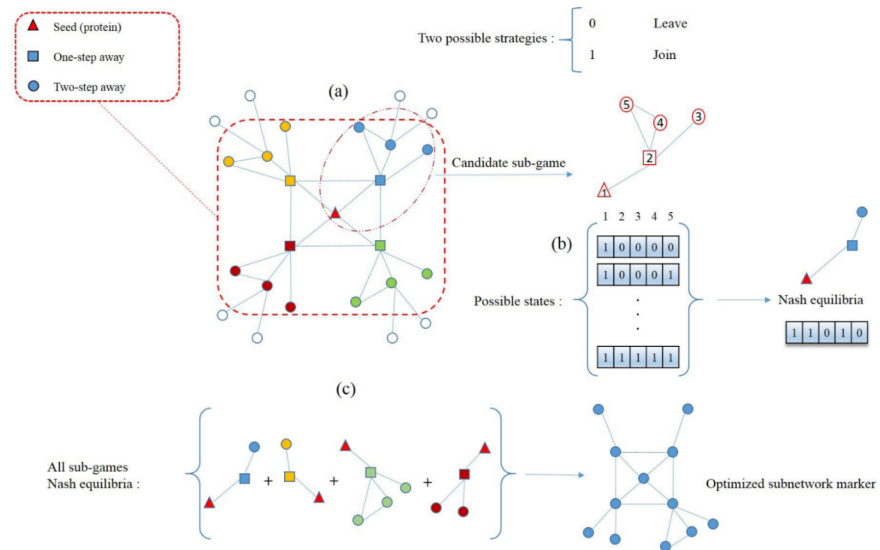


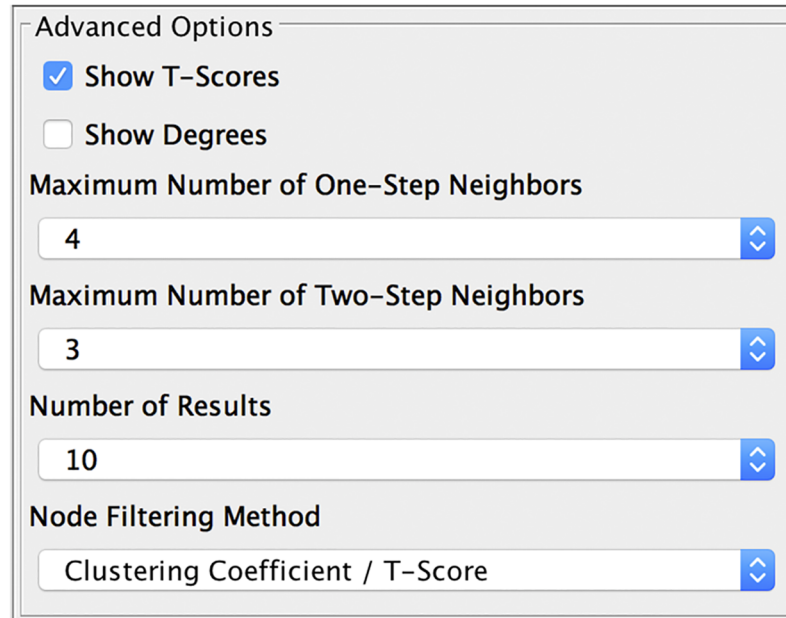
Fig 1. Visual overview of the approach used in GTA algorithm. A candidate subnetwork (game) is shown in (a). The selected seed of this subnetwork (game) has 4 one-step and 12 two-step neighbours. Therefore, the number of possible states to be evaluated, is 2^{16} . To reduce the complexity of the game, it is divided into 4 sub-games due to the number of the first-step neighbours (one of these sub-games is surrounded by a dotted line as a representative). Here in this example, each sub-game has 2^4 possible states. The possible states of the representative sub-game are shown in (b). Since, the seed inevitably takes part in the resulting subnetwork, the value of 1 is assigned to the seed in each possible state. The Nash equilibria of all sub-games are shown in (c) and are then integrated into a single optimized subnetwork marker.

<https://doi.org/10.1371/journal.pone.0185016.g001>

To develop CytoGTA, GTA algorithm has been written in Java programming language. Computing the local clustering coefficient and determining the equilibrium for all possible states are two of the most time-consuming parts in CytoGTA. CytoGTA leverages some pervasive characteristics of biological networks, such as small-world property and sparsity, to speed up the computation time in GTA and improve its efficiency. As explained before, in GTA algorithm, for a particular candidate subnetwork, the local clustering coefficient of its seed needs to be updated when any of its one- or two-step neighbors is removed from the subnetwork. In CytoGTA, the algorithm is adapted so that updating the local clustering coefficient is performed more efficiently. Moreover, some minor optimization changes are made to accelerate computing the equilibrium, which leads to the slightly better performance of CytoGTA.

Results and discussion

To run CytoGTA, two types of data need to be imported into Cytoscape, a PPIN and an expression profile. Using the “Import Network From File” option, the input PPIN (weighted or unweighted network) can be loaded and build in Cytoscape. The expression profile of two phenotype classes is also importable through the “Import Table From File” option. Once the input PPIN and the expression profile are successfully loaded, a summary will appear in “Table Panel”. Moreover, by choosing “CytoGTA” in “Control Panel”, the loaded network is accessible through the “Reference Network” icon, and the list of sample labels in the expression profile are presented in the “Control” and the “Case” icons. To get the mappings between the expression profile of each gene and its corresponding protein in the PPIN, the same system of nomenclature needs to be used in the expression profile and in the PPIN. A number of options are provided in “Advanced Options”, which are expected to be specified before running CytoGTA. Fig 2 shows all CytoGTA icons in “Control Panel”.



Advanced Options

Show T-Scores

Show Degrees

Maximum Number of One-Step Neighbors

4

Maximum Number of Two-Step Neighbors

3

Number of Results

10

Node Filtering Method

Clustering Coefficient / T-Score

Fig 2. Advanced options in the control panel of CytoGTA. Attributes which can be specified by the user.

<https://doi.org/10.1371/journal.pone.0185016.g002>

To run CytoGTA, it is necessary to select a PPIN from the drop-down list in the “Reference Network” icon, and some samples of different phenotype classes from “Control” and “Case” to be compared by their expression patterns. The further step is to specify the options in “Advanced Options”, which are already initialized with default values. Selecting the checkboxes next to “Show T-Scores” and “Show Degrees” enables CytoGTA to present the t-score and the degree for each gene in the result subnetwork marker. As it is mentioned before, the one-step and the two-step neighbors of the selected seed are pruned iteratively, with the aim of increasing its local clustering coefficient. Furthermore, CytoGTA allows the users to set the maximum values for the numbers of one-step and two-step neighbors, which can vary from 1 to 19. Moreover, the number of result subnetwork markers is selectable through the drop-down list in the “Number of Results”. Finally, by “Node Filtering Method” icon, users are provided with the option to select the pruning strategy used by CytoGTA to exclude less informative nodes.

By clicking on “Find Modules”, CytoGTA starts running the algorithm on the loaded input, considering user-requested options. When the execution of CytoGTA is finished, a list of top optimized subnetwork markers appears in the “Results” section of the “Control Panel”. The result list contains the desired number of subnetworks, requested by the user. Each subnetwork marker has given a name, which is descriptive of its seed gene. CytoGTA provides the users with the visualization of the result subnetwork markers, as well as displaying the t-score and the degree of their comprising genes (only if their corresponding checkboxes are ticked). In the graphical visualization of the subnetworks, the seed, the one-step neighbors and the two-step neighbors are distinguished by different colors to enhance the interpretation of the results. The result subnetworks are exportable, and are usable by other analysis tools within Cytoscape. Fig 3 shows different panels obtained from the execution of CytoGTA on a sample data, by setting the maximum numbers of one-step and two-step neighbors to be 4 and 3, respectively.

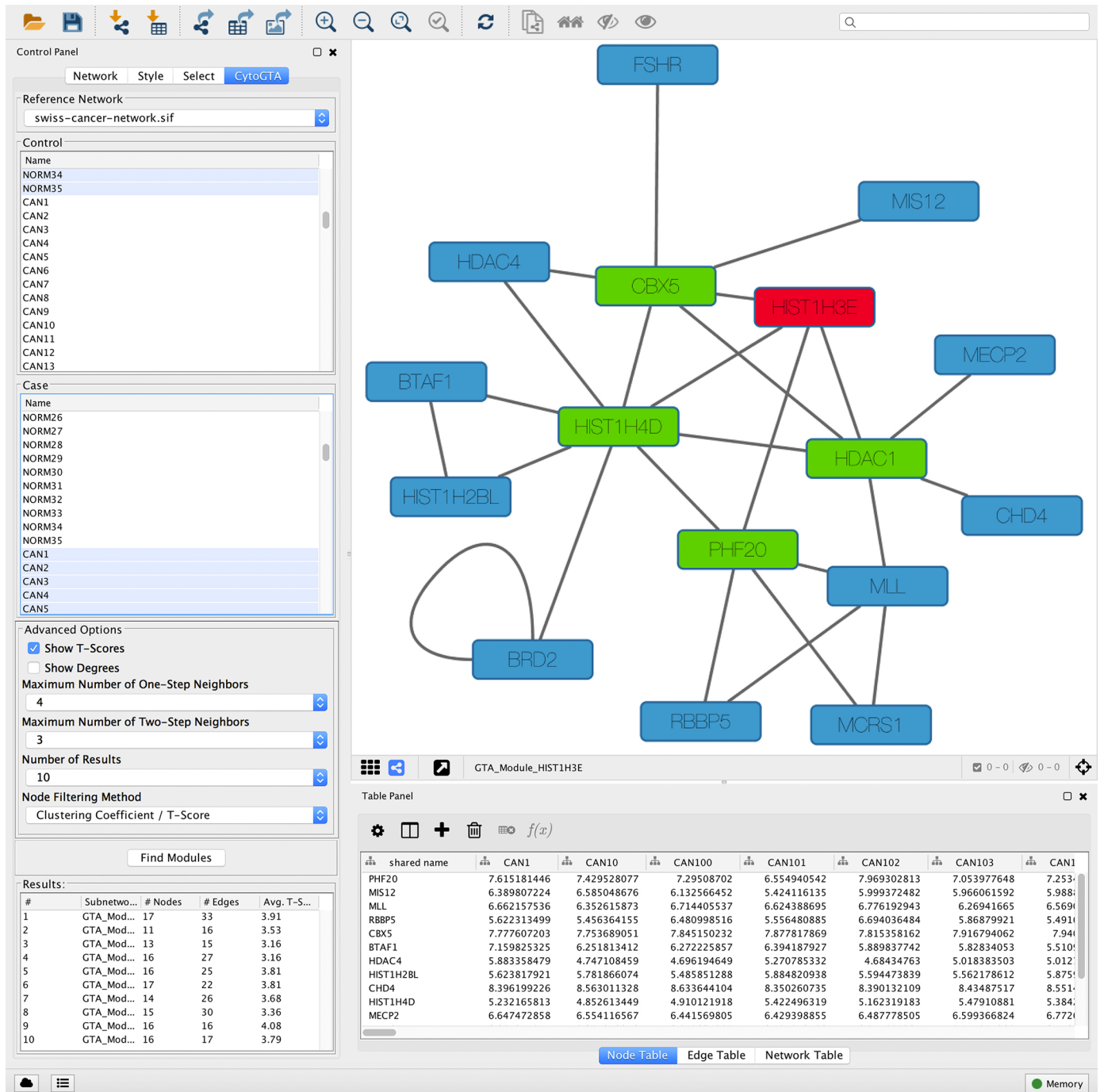


Fig 3. A sample of CytoGTA result. In this sample the maximum numbers of one-step and two-step neighbors are set to be 4 and 3, respectively.

<https://doi.org/10.1371/journal.pone.0185016.g003>

Experiments

To evaluate the efficiency of CytoGTA, it was applied on three independent gene expression profiles of metastatic and non-metastatic breast cancer samples, and subsequently, the classification performance of result markers was evaluated by the support vector machine (SVM)

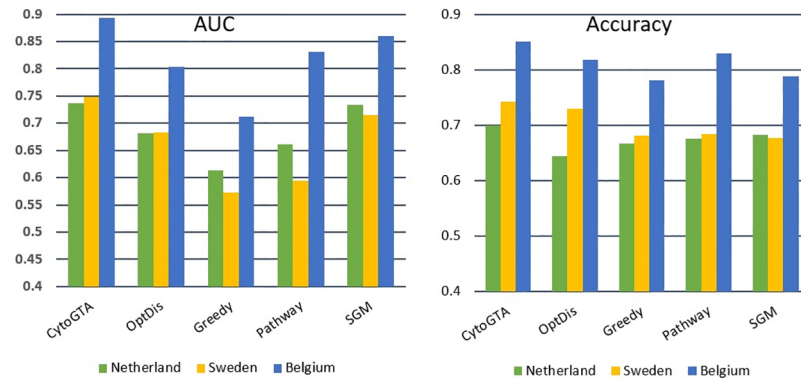


Fig 4. Classification performance comparison of the within-dataset experiment.

<https://doi.org/10.1371/journal.pone.0185016.g004>

method. Two of the expression profiles were retrieved from the Affymetrix U133A platform and are accessible through the Gene Expression Omnibus (GEO) database with GEO accession number GSE7390 [21] (the Belgium dataset) and GSE1456 [22] (the Sweden dataset). The third dataset (the Netherlands dataset) was profiled on the Agilent microarray platform, and obtained from the study by van de Vijver et al. [4]. To account for the magnitude differences among transcript levels, all expression profiles were normalized before importing into Cytoscape. The normalization process was performed with the RMA algorithm [23] using the Affy package in R. Furthermore, the human PPIN used in this evaluation, was taken from the study by Lage et al. [24], which contains 169,810 high confidence interactions on 12,879 proteins.

Approaches used in this study for evaluating the efficiency of CytoGTA, can be categorized into single gene-based, pathway-based and network-based (the greedy and the optimally discriminative (OptDis) methods). For the single gene-based approach, the 50 genes with the best absolute $t - score(LLR)$ values were chosen as the markers. In the pathway-based approach, among 1320 reported pathways in the C2 curated gene sets in the Molecular Signature Database (MsigDB), the top 50 were extracted based on the $t - score(LLR)$ values of their member genes. The top 50 optimized subnetworks were chosen in network-based approaches employed in this study (greedy, OptDis and CytoGTA). All aforementioned marker types were used to train and test the SVM classifier.

Both within-dataset and cross-dataset analyses were conducted to evaluate the classification performance of the markers. The within-dataset classification evaluation was performed by five-fold cross-validation, repeated ten times. The classification performance was measured in terms of accuracy and the area under the curve (AUC). As it is shown in Fig 4, based on both criteria, CytoGTA shows better classification performance compared to other methods.

In addition, the reproducibility of the markers was estimated by the cross-dataset analysis. To this aim, the top 50 markers extracted from one dataset, were used in five-fold cross-validation experiment on the other two datasets. In this analysis, the validation experiment was also repeated ten times. As it is presented in Figs 5 and 6, the classifier constructed on CytoGTA, mostly outperformed those based on other methods, which suggests the superior reproducibility of markers obtained by CytoGTA.

Furthermore, to examine the biological relevance of identified markers with cancer, Gene Ontology (GO) analysis was performed using PANTHER [25]. Interestingly, the result markers were significantly enriched with GO terms (see Fig 7), most of which are well-known in cancer pathology and are consistent with several studies on this field [12, 21, 26].

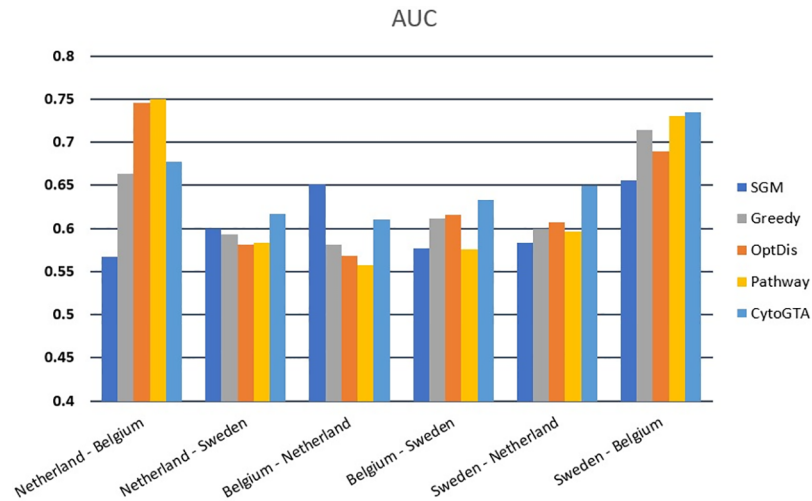


Fig 5. Classification performance comparison of the cross-dataset experiment in term of AUC criterion.

<https://doi.org/10.1371/journal.pone.0185016.g005>

Conclusion

In in-silico systems biology, there has been an increasing interest in developing tools, with the objective of facilitating the study of complex diseases and traits. Here, CytoGTA as a new Cytoscape plugin is introduced, in which a PPIN and a gene expression profile of two phenotype classes are given as input, and a number of discriminative markers are returned as output. The high performance of CytoGTA, in terms of reproducibility and reliability, makes it extremely convenient for the purpose of identifying biomarkers in different phenotypes. In this light, embedding CytoGTA in Cytoscape will contribute with the overall profitability of Cytoscape as a comprehensive tool for the analysis of biological networks.

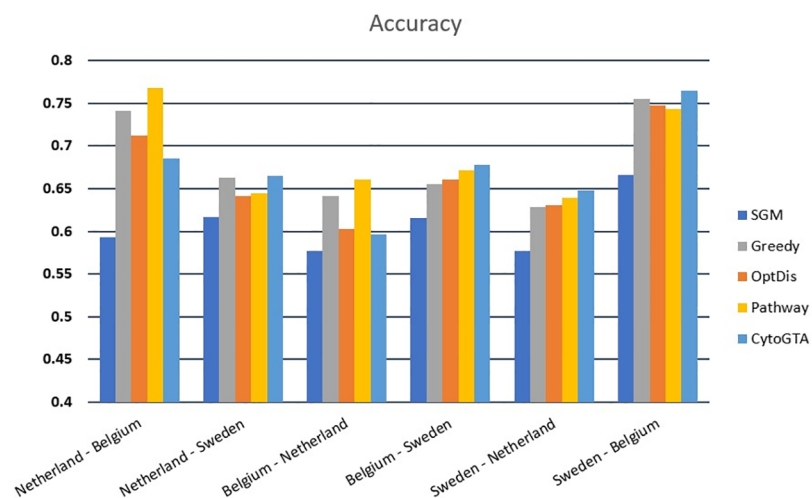


Fig 6. Classification performance comparison of the cross-dataset experiment in term of accuracy criterion.

<https://doi.org/10.1371/journal.pone.0185016.g006>

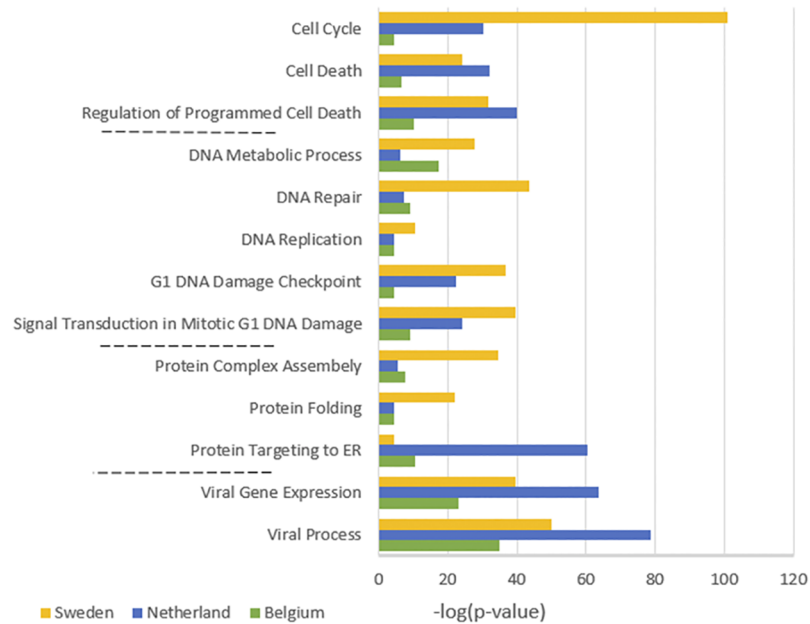


Fig 7. Enriched biological processes of the identified subnetwork markers for Sweden, Netherland and Belgium datasets.

<https://doi.org/10.1371/journal.pone.0185016.g007>

Availability

The tutorial, full package and related examples are available at <https://github.com/cocoamilk/GTA>.

Author Contributions

Conceptualization: S. Goliaei, Z. Razaghi-Moghadam.

Data curation: S. Farahmand, Z. Razaghi-Moghadam.

Investigation: S. Goliaei, Z. Razaghi-Moghadam.

Methodology: S. Farahmand, M. H. Foroughmand-Araabi, Z. Razaghi-Moghadam.

Project administration: Z. Razaghi-Moghadam.

Software: M. H. Foroughmand-Araabi.

Supervision: Z. Razaghi-Moghadam.

Validation: S. Farahmand, Z. Razaghi-Moghadam.

Writing – original draft: S. Farahmand.

Writing – review & editing: M. H. Foroughmand-Araabi, Z. Razaghi-Moghadam.

References

1. Sadikovic B, Al-Romaih K, Squire JA, Zielenska M. Cause and consequences of genetic and epigenetic alterations in human cancer. *Curr Genomics*. 2008; 9(6):394–408. <https://doi.org/10.2174/138920208785699580> PMID: 19506729

2. Dao P, Colak R, Salari R, Moser F, Davicioni E, Schonhuth A, et al. Inferring cancer subnetwork markers using density-constrained biclustering. *Bioinformatics*. 2010; 26(18):i625–31. <https://doi.org/10.1093/bioinformatics/btq393> PMID: 20823331
3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000; 403(6769):503–11. <https://doi.org/10.1038/35000501> PMID: 10676951.
4. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002; 347(25):1999–2009. <https://doi.org/10.1056/NEJMoa021967> PMID: 12490681.
5. Al WYe. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005; 365:671–9. [https://doi.org/10.1016/S0140-6736\(05\)17947-1](https://doi.org/10.1016/S0140-6736(05)17947-1) PMID: 15721472
6. Jahid MJ, Ruan J. A Steiner tree-based method for biomarker discovery and classification in breast cancer metastasis. *BMC Genomics*. 2012; 13 Suppl 6:S8. <https://doi.org/10.1186/1471-2164-13-S6-S8> PMID: 23134806.
7. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell*. 2011; 144(6):986–98. <https://doi.org/10.1016/j.cell.2011.02.016> PMID: 21414488.
8. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature reviews Genetics*. 2011; 12(1):56–68. <https://doi.org/10.1038/nrg2918> PMID: 21164525.
9. Guo Z, Zhang T, Li X, Wang Q, Xu J, Yu H, et al. Towards precise classification of cancers based on robust gene functional expression profiles. *BMC bioinformatics*. 2005; 6:58. <https://doi.org/10.1186/1471-2105-6-58> PMID: 15774002.
10. Tomfohr J, Lu J, Kepler TB. Pathway level analysis of gene expression using singular value decomposition. *BMC bioinformatics*. 2005; 6:225. <https://doi.org/10.1186/1471-2105-6-225> PMID: 16156896.
11. Su J, Yoon BJ, Dougherty ER. Accurate and reliable cancer classification based on probabilistic inference of pathway activity. *PloS one*. 2009; 4(12):e8161. <https://doi.org/10.1371/journal.pone.0008161> PMID: 19997592.
12. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol*. 2007; 3:140. <https://doi.org/10.1038/msb4100180> PMID: 17940530.
13. Farahmand S, Goliaei S, Ansari-Pour N, Razaghi-Moghadam Z. GTA: a game theoretic approach to identifying cancer subnetwork markers. *Mol Biosyst*. 2016; 12(3):818–25. <https://doi.org/10.1039/c5mb00684h> PMID: 26750920.
14. Holt CA, Roth AE. The Nash equilibrium: a perspective. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(12):3999–4002. <https://doi.org/10.1073/pnas.0308738101> PMID: 15024100.
15. Hwang T, Park T. Identification of differentially expressed subnetworks based on multivariate ANOVA. *BMC bioinformatics*. 2009; 10:128. <https://doi.org/10.1186/1471-2105-10-128> PMID: 19405941.
16. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*. 2002; 18 Suppl 1:S233–40. PMID: 12169552.
17. Mitra K, Carvunis AR, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. *Nature reviews Genetics*. 2013; 14(10):719–32. <https://doi.org/10.1038/nrg3552> PMID: 24045689.
18. Qiu YQ, Zhang S, Zhang XS, Chen L. Detecting disease associated modules and prioritizing active genes based on high throughput data. *BMC bioinformatics*. 2010; 11:26. <https://doi.org/10.1186/1471-2105-11-26> PMID: 20070902.
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003; 13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658.
20. Developers CM. Apache commons math. Forest Hill M, USA: The Apache Software Foundation. 2017.
21. Pawitan Y, Bjohle J, Amler L, Borg AL, Eghazi S, Hall P, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast cancer research: BCR*. 2005; 7(6):R953–64. <https://doi.org/10.1186/bcr1325> PMID: 16280042.
22. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003; 31(4):e15. PMID: 12582260.
23. RDC T. R: A language and environment for statistical. R Foundation for Statistical Computing. 2004.
24. Lage K, Karlberg EO, Stirling ZM, Olason PI, Pedersen AG, Rigina O, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*. 2007; 25(3):309–16. <https://doi.org/10.1038/nbt1295> PMID: 17344885.

25. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013; 41(Database issue):D377–86. <https://doi.org/10.1093/nar/gks1118> PMID: 23193289.
26. Yao C, Li H, Zhou C, Zhang L, Zou J, Guo Z. Multi-level reproducibility of signature hubs in human interactome for breast cancer metastasis. *BMC systems biology.* 2010; 4:151. <https://doi.org/10.1186/1752-0509-4-151> PMID: 21059271.