ClustEx2: Gene Module Identification using Density-Based Network Hierarchical Clustering

Conference Paper · November 2018

DOI: 10.1109/CAC.2018.8623442

3 authors:

Zijian Ding
Tsinghua University
5 PUBLICATIONS **285** CITATIONS
SEE PROFILE

Wenbo Guo
Tsinghua University
13 PUBLICATIONS **119** CITATIONS
SEE PROFILE

Jin Gu
Tsinghua University
85 PUBLICATIONS **2,099** CITATIONS
SEE PROFILE

ClustEx2: gene module identification using density-based network hierarchical clustering

Zijian Ding
MOE Key Laboratory of Bioinformatics,
BNRIST Bioinformatics Division,
Department of Automation
Tsinghua University
Beijing, China
dingzijian881218@126.com

Wenbo Guo
MOE Key Laboratory of Bioinformatics,
BNRIST Bioinformatics Division,
Department of Automation
Tsinghua University
Beijing, China
gwb17@mails.tsinghua.edu.cn

Jin Gu
MOE Key Laboratory of Bioinformatics,
BNRIST Bioinformatics Division,
Department of Automation
Tsinghua University
Beijing, China
jgu@tsinghua.edu.cn

***Abstract*—With the fast accumulation of large-scale omic data in cancer, it is easy to get a list of seed genes associated with any clinical phenotype change or anti-cancer drug perturbation. However, it remains a challenging task to functionally interpret these genes. Many studies indicate that genes work cooperatively as functional modules in complex cellular processes. More biological insights can be obtained by clustering these genes and their closely interacting neighbors in molecular networks as gene modules. Unlike the traditional network community detection, the gene modules involve in the seed genes that carry context specific information and their closely connected neighbors in the static molecular networks. We developed a new method ClustEx2, which can identify gene modules based on a set of user-defined seed genes in a given gene network. The method formulates the module identification in a unified framework according to a density-based hierarchical clustering method. ClustEx2 can incorporate both network topology and context specific information of seed genes along with their interactions, such as differential expressions and co-expressions. Its performance was systematically investigated for a known biological process, tumor necrosis factor induced inflammation, and it also helped obtain potential biological functions by analyzing the anti-cancer drug response associated modules using the TCGA data.**

Keywords—gene module, molecular networks, context specific information, density-based hierarchical clustering

I. INTRODUCTION

With the fast development and reducing cost of high throughput technologies, more different omics data of cancer are available, such as gene mutation, DNA methylation and gene expression, etc. The Cancer Genome Atlas (TCGA) and International Cancer Genome (ICGC) [1] provide large scale of omic data and clinical diagnosis and treatment information. The Cancer Cell Line Encyclopedia (CCLE) [2], Genomics of Drug Sensitivity in Cancer (GDSC) [3] and the Library of Network-based Cellular Signatures (LINCS) offer anti-cancer drug responses and corresponding molecular data before or after drug perturbations. Based on these datasets, we can obtain the lists of seed genes that are associated with different disease phenotypes and molecular perturbations by anti-cancer drugs. Usually we can interpret these lists with applications of function enrichment tools such as DAVID [4] and GSEA [5].

However, genes are not isolated but work cooperatively to perform different cellular functions. The interactions among genes reflect their collaboration and are very useful for interpreting the functions and regulatory structures of the seed genes. Frequently interacting genes involved in common biological processes and signaling pathways usually form a gene module, which is an important characteristic of functional gene network [6].

Discovering gene modules based on the seed genes and gene-gene interaction networks is important for understanding the molecular mechanisms of cancer. Different from the traditional strategies to detect network community [7], which are mainly based on the network topology, the module identifications here aim to find the dense sub-networks (or commonly called as gene modules) containing the seed genes and their closely connected neighbors. Therefore, the module identification methods need to integrate the dynamic context specific information of omic data and the static network topology. Several computational methods have been developed to identify gene modules, including active subnetworks [8], [9], network biomarkers [10], overlapping modules [11], responsive modules [12], and significantly altered subnetworks in cancer [13].

In our previous work, we proposed a module identification framework ClustEx, which can identify responsive gene modules by clustering differentially expressed genes and their closely interacting neighbors in a given gene network [12], [14]. Although it performed better than several existing methods, there remains some shortcomings: 1) it treats every differentially expressed gene equally though these seed genes varies a lot in expression change levels, and 2) its framework divides the seed gene clustering and non-seed gene extending into separate procedures, which makes it hard to control the output module sizes. To overcome these shortcomings, we propose a more unified framework ClustEx2. In order to evaluate its performances, we also took into account a tool called SubNet [15] which includes several classic methods, such as nearest neighbor, shortest path, diffusion kernel [16] and PageRank [17].

ClustEx2 aims to identify context-specific modules in a gene network based on any user-defined seed genes. It has two major advantages than the previous methods: 1) it integrates the seed gene clustering and non-seed gene extending in a unified procedure based on a density-based hierarchical clustering framework; and also, 2) it is able to incorporate the context specific information of seed genes and gene-gene interactions, such as fold changes and co-expressions. The density-based clustering method of ClustEx2 takes into account both the dynamic context information and the static network topology. In addition, this

Research supported by National Natural Science Foundation of China [61370035 and 81630103] and Tsinghua University Initiative Scientific Research Program.

method makes it more convenient to control output module sizes by offering a principle, that when module sizes are smaller, the overall densities of each module are higher and thus the genes in modules are more likely to be associated with the phenotype under study.

We first evaluated ClustEx2 on a known biological process, i.e. tumor necrosis factor (TNF) induced inflammation, according to a time-course microarray dataset of vascular endothelial cells after TNF treatment used in our previous work [12]. Results showed that under equal module sizes, the performances of ClustEx2 were generally better than the recommended method of SubNet (PageRank) and ClustEx. We then applied ClustEx2 to annotate the gene modules associated with clinical responses to cisplatin, according to the results of our work on evaluating the molecule-based prediction of clinical drug responses in cancer [18]. Results showed that ClustEx2 can discover biological processes known to mediate cisplatin response.

In the part of METHOD, we proposed a novel gene network module identification method, namely ClustEx2. In the part A of RESULTS, we validated its performance based on a well-known biological process, tumor necrosis factor induced inflammation. In the part B of RESULTS, we applied ClustEx2 to identify clinical drug response associated gene network modules, in order to explore the biological mechanisms of anti-cancer drug responses.

II. METHOD

A. Overview

Given an undirected gene network and a list of user-defined seed genes, ClustEx2 addresses the problem of gene module identification using a density-based hierarchical clustering framework. Notably, ClustEx2 can consider context specific information of genes and their interactions derived from a specific gene expression data set, such as fold changes of differentially expressed genes, and co-expression correlations of gene-gene interactions in the network. After mapping seed genes onto the network, random walk and diffusion kernel were applied to diffuse this information from the seed genes to their neighbors in the network. To this end, random walk and diffusion kernel were applied to calculate the gene importance scores and gene-gene similarities, which are the basis of the density-based clustering. Specifically, gene important scores function as the "densities". Then, modules were identified by progressively clustering the genes in a decreasing order of the importance scores or densities (Fig. 1). With modules expanding, the overall density of each module decreased. Users can adjust the sizes of output modules and visualize them in the Cytoscape [19]. We implemented ClustEx2 in C++ and provided a command line interface on both Linux and Windows. All resources, including the additional files mentioned in the following parts of the paper, are available at http://bioinfo.au.tsinghua.edu.cn/software/clustex2/.

B. Gene importance and gene-gene similarity scores

Seed genes with higher context specific scores are more likely to reflect the molecular perturbation or phenotype change, and non-seed genes located around seed genes may be involved in common biological functions [20]. We took this context specific scores into consideration, and assumed

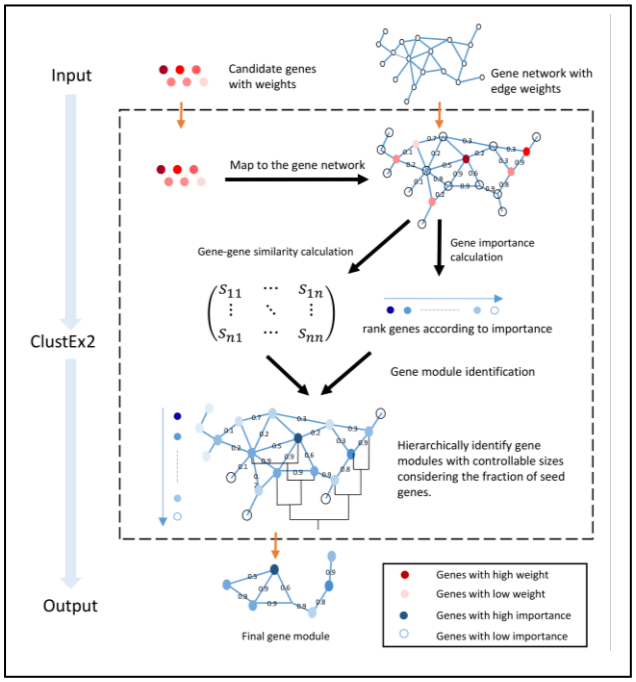

Fig. 1. Workflow of ClustEx2

that these scores carried by seed genes could influence other non-seed genes through the network interactions. Therefore, gene importance scores were calculated to represent the influence of the context specific information smoothed by the gene network [21], using random walk with restart which was applied to identify disease-associated genes by Köhler S. et al [22]. The equation of random start with restart is as follows:

$$p^{t+1} = (1-r)Ap^{t} + rp^{0} \quad (1)$$

where p^0 is the normalization of the context specific information (such as the fold changes), p^t is the probability of the random walker appearing at each gene at the time point t, p^{t+1} is the probability at time point $t+1$, and A is the column normalized adjacent matrix that represents the network topology (could be weighted by gene-gene co-expressions). Practically, gene importance scores formulates a steady-state probability distribution of random walk. It represents the idea of density based on the context information in p^0 and the network topology in A. The restart probability r controls the balance of these two information.

We suppose that genes in the same module are closely connected. In order to measure the extent to which each pair of genes were closely connected in the network, gene-gene similarity scores were calculated. The diffusion kernel which conceptually performs a lazy random walk based on matrix exponentiation and diffuses the context specific information from each gene [16], was used to calculate the pairwise similarities of genes in the network. The diffusion kernel based network similarity has been shown to outperform other measurements such as shortest path length [23], as in

$$K = e^{\beta H}, H = A - D. \quad (2)$$

The element K_{ij} in matrix K represents the similarity score between gene i and gene j. The negative Laplacian matrix H was calculated by subtracting the adjacent matrix A from the diagonal matrix D with D_{ij} as the degree of gene i.

C. Gene module identification

To use the gene importance and gene-gene similarity scores at the same time, a density-based hierarchical clustering method [24, 25] was developed. The original idea of density-based clustering is that samples with high densities are clustered into modules, and different modules are separated by samples with low densities. In our method, genes with high importance scores and high gene-gene similarity scores were clustered into separate modules.

In the proposed clustering method, a module incorporated genes with high importance scores first and then genes with lower scores gradually in a hierarchical manner. Different modules were separated by genes with limited importance scores. To be specific, firstly all genes were sorted in descending order according to their importance scores. Secondly, the top gene with the highest score was initiated as a new module. Then, module clustering went on by checking the remaining genes one by one: if the next gene's similarity to any existing modules was larger than the pre-given threshold, i.e. neighborhood, this gene and the modules were merged into one module; otherwise, this gene was initiated as another new module. The approach can clearly display the process of modules expansion, which helps users flexibly control the output module sizes by checking the fractions of seed genes in the largest modules. A description of the framework with formula was presented in the Additional file 1.

An important characteristic of hierarchical clustering is that a proper cutoff determines the output modules, otherwise all samples will be incorporated into one module at last. Practically we can regard this as finding a proper break point for the hierarchical clustering algorithms. Similarly, users can define a break point for the proposed clustering method. To help users define this break point, ClustEx2 provides several visualized features, including the sizes of modules, the number and fraction of seed genes in modules under different neighborhood parameters (Fig. 2). The cutoff is defined as the size of the largest module. Users can determine this cutoff based on the output figures as shown in Fig. 2. For example, if a user wants that the largest module size is not too large, and the fraction of seed genes is not too small, the user can determine the cutoff as 100/200/300 each time and obtain the final modules. Since there may appear a dozen of gene modules during clustering, it is unrealistic to monitor these features in all modules. As a compromise, ClustEx2 provides the features in the largest two modules. The process of density-based hierarchical clustering is stopped when the module size and the fraction of seed genes are satisfactory.

There are three parameters that can affect the final results. Fortunately we find a simple way for the parameterization. Users only have to decide the value of one parameter, i.e. neighborhood. It controls whether a gene is close enough to some module. A smaller neighborhood results in modules in which genes are more closely connected. In contract, a larger neighborhood results in modules in which genes are not that

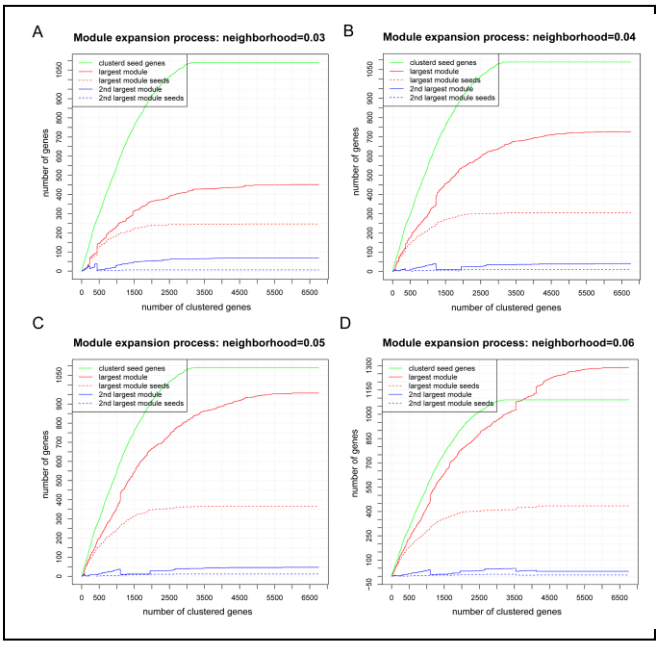

Fig. 2. Display of features during module expansion

close. The principle is that with the same number of genes clustered, a smaller neighborhood results in more modules and higher intra-cluster similarity of each module. The range of this parameter is from 0 to 1. Practically, We suggest choosing an appropriate neighborhood that is small enough to result in high intra-cluster similarity of each module but not too small to avoid that every gene forms a module by itself. For example, as in Fig. 2, if we want the largest module size to be 100, we suggest a smaller neighborhood as 0.03, which results modules to be more compact. Besides the neighborhood, uses can use the default settings. Details of the default parameterization were presented in Additional File 1.

III. RESULTS

A. Responsive gene modules of TNF induced HUVECs

In order to test whether ClustEx2 improves performances on identifying gene modules, we took advantage of a well-annotated biological process, i.e. TNF induced inflammation, using a time course gene expression dataset for TNF treated HUVECs (human umbilical vein endothelial cells), on which ClustEx was proved to perform better than several benchmark methods [12]. To begin with, we confirmed that ClustEx2 incorporated genes with high densities, namely seed genes with large fold changes and non-seed genes with large degrees. Then we applied three reference functional gene sets from public databases to prove that ClustEx2 can better identify genes related to TNF induced inflammation. Finally, we used functional enrichment method to verify that ClustEx2 can identify more biological processes and signaling pathways. We also compared ClustEx2 to a classic method PageRank [17] recommended among several classic methods by SubNet [15].

Here we show how to apply ClustEx2 to identify gene modules. After filtering lowly expressed genes in the time series gene expression data of TNF induced inflammation, we mapped all genes to the PPI network curated by STRING

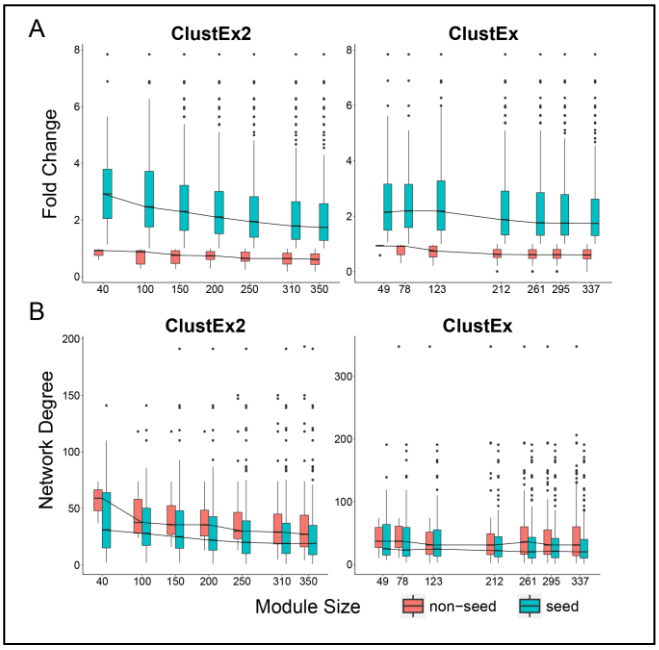

Fig. 3. The distributions of fold changes and network degrees in responsive gene modules identified by ClustEx2 and ClustEx

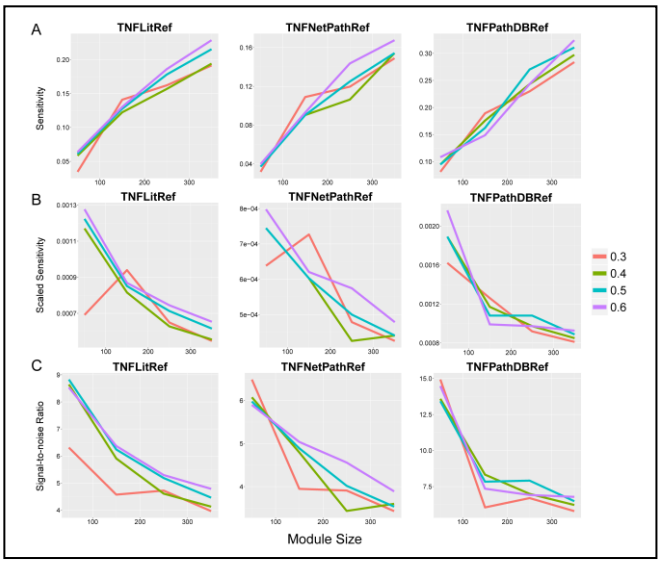

Fig. 4. The sensitivities and signal-to-noise ratios of responsive gene modules identified by ClustEx2

database. The resulting network contained 6,756 genes, 42,905 edges and 1,089 seed genes (>=2 fold, fold change measures the extent to which each gene differentially expressed, more details of the data process is in the additional file 1). Gene modules responsive to TNF inducement in HUVECs were identified for these seed genes in the network by the three methods. The detailed parameter settings were provided in Additional file 2. The basic information of the identified modules were provided in Additional file 1: Table S1.

The density-based clustering framework of ClustEx2 brought several advantages over ClustEx. ClustEx2 takes into account the fold changes of seed genes and network topology simultaneously. Higher fold changes of seed genes are more likely to be caused by the TNF treatment, and their closely connected non-seed genes with higher network degrees might be more likely to be involved in the common cellular processes. Therefore we assumed that the modules with high overall fold changes and network degrees were likely to be closely associated with biological functions induced by TNF treatment. As the identified modules by ClustEx2 expanded, the overall fold changes of seed genes and the network degrees of non-seed genes decreased (Fig. 3). What's more, since both ClustEx2 and ClustEx need break points to determine final modules, an early stopping of ClustEx2 can result in modules with higher overall fold changes and network degrees. Finally, compared to the separated procedures of clustering and extending of ClustEx, the more unified framework of ClustEx2 made it more convenient to control output module sizes.

We then assessed how accuracy ClustEx2 can identify genes known to be related to TNF induced inflammation in different modules with various sizes. We collected three reference functional gene sets from (1) TNFNetPathRef, short for "TNF/NF-kB signaling pathway" from NetPath database, (2) TNFPathDBRef, short for "TNF signaling pathway" from PID/BioCarta/Reactome and (3) TNFLitRef, short for literature mining from PubMed abstracts. Sensitivity and signal-to-noise ratio were applied to evaluate the accuracy and significance level. This strategy was used by ClustEx [12] and the three gene sets were updated here. As shown in Figure 4, under no matter what parameterizations of neighborhood or what reference gene sets, with the module sizes increasing, the sensitivities are increasing, the sensitivities normalized by module sizes are also decreasing, and the signal-to-noise ratios are decreasing. This may imply that when the module size is smaller and genes in the module have larger fold changes or network degrees, the identified module is more likely to be associated with the molecular perturbation, namely TNF inducement. When comparing the performance of the three different methods, ClustEx2 consistently achieved better performances than PageRank under no matter what parameters or reference gene sets, as shown in Figure 5. Based on the Literature mining reference set, ClustEx2 was better than ClustEx under all parameters, and they were comparable according to the other two gene sets.

Higher fraction of seed genes and closely connected non-seed genes with higher degrees in the responsive module were supposed to be enriched in more TNF related KEGG signaling pathways. DAVID [4] was applied to identify enriched biological functions and pathways. We used the module with 350 genes which contained 257 seeds under the 0.03 neighborhood, since a smaller neighborhood means that genes in modules are more closely connected. The genes in this module were significantly enriched in 27 KEGG pathways (Bonferroni corrected p-value<=0.05), including many pathways known to be affected by TNF, such as Apoptosis, Notch signaling pathway, JAK-STAT signaling pathway and Toll-like receptor signaling pathway (Additional file 2). Two pathways in the module of ClustEx2 were not found by ClustEx, including adherens junction and leukocyte transendothelial migration. It has been reported that adherens junction is increased in response to TNF [26] and leukocyte TEM is induced by TNF stimulation [27]. Though PageRank outperformed ClustEx2 on the fraction of seed genes and the number of significantly enriched KEGG pathways, some well-known activated pathways such as JAK-STAT pathway were not enriched in the identified modules (Additional file 2).

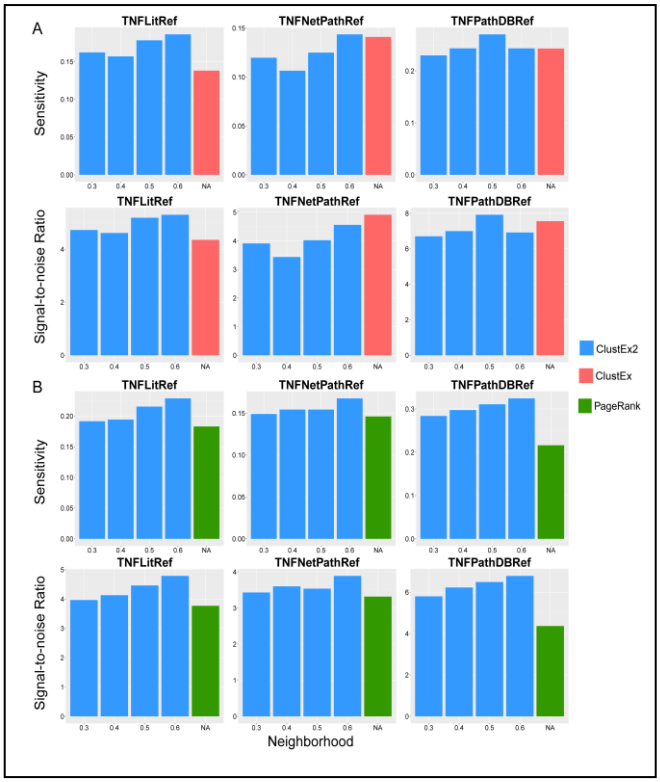

Fig. 5. The comparision of sensitivities and signal-to-noise ratios of responsive gene modules identified by ClustEx2, ClustEx and PageRank

To sum up, ClustEx2 has the characteristic that with increasing module sizes, the overall densities are decreasing. We used the TNF induced inflammation to prove that this trend is associated with the studied phenotype, namely with increasing module sizes, ClustEx2 can identify more phenotype associated genes though with a lower significance level. In addition, ClustEx2 performs better than either ClustEx or PageRank under the equal module sizes.

B. Differential gene modules of clinical drug response

After confirming that ClustEx2 performed better than its previous version ClustEx, we applied it to obtain new biological findings of drug responses in cancer. Predicting clinical drug responses in cancer is an important issue in the era of precision medicine [28]. In our previous work, we systematically evaluated the performance of different types of genomic data on predicting clinical responses of different drugs in different cancer types [18]. In the analysis, due to small sample sizes and far more molecular features, we assumed that genes independently predicted drug response, ignoring the fact that genes interact with each other to perform biological functions. Here, we applied ClustEx2 to identify gene modules which consist of differentially expressed genes between the responders and non-responders. We defined these modules as differential gene modules and focused on BLCA (bladder cancer) and CESC (cervical cancer), since the gene expression data of these two cancer patients predict cisplatin response significantly better than random classifiers [18].

Here we show how to apply ClustEx2 to identify drug response differential gene modules. In order to cover as many genes as possible measured by TCGA RNA-seq data in a gene network, we integrated different public protein-protein interaction databases including OPHID, MINT, HPRD, BioGrid and DIP. The network contained 21,980 genes and 377,670 interactions, which can cover almost all genes in TCGA data. The genes with absolute values of average fold changes larger than 2 were defined as seed genes, and the weights of edges were derived from gene co-expressions. ClustEx2 identified two gene modules with 50 genes in each for BLCA and CESC (Additional file 1: Fig. S6, Additional file 4).

Through functional enrichment analysis, we found that the BLCA or the CESC module was significantly related to immune response and oxidation reduction, respectively (FDR 5.47e-27 and 1.22e-9). In the BLCA module, 32 out of the 50 genes are directly involved in the immune response, in which CCL5 gene has been reported to be related to cisplatin resistance [29]. However, it's been seldom reported that the immune response as a whole are related to cisplatin response in BLCA. As a result, ClustEx2 may help postulate a gene network basis for the mechanism of cisplatin resistance. In the CESC module, 19 out of the 50 genes are directly involved in the oxidation-reduction process. It's known that cisplatin, which includes metal platinum, can cause oxidation-reduction. And it's been reported recently that this process are responsible for cisplatin resistance [30], [31]. To sum up, ClustEx2 can provide us with new information about gene regulatory networks of clinical drug responses in cancer.

IV. Discussion

ClustEx2 is developed to identify context-specific gene modules based on a list of seed genes and closely connected neighbors in a gene network. It applies the density-based clustering method to take into consideration the dynamic context specific information of seed genes and static network topology. As a result, with module sizes increasing, the overall density decreases, which is implied by the decreasing overall context specific scores such as fold changes and network degrees. And therefore ClustEx2 finds more phenotype associated genes under an increasing significance level when the module sizes are smaller.

Based on a known biological process, TNF induced inflammation, ClustEx2 was proved to outperform the PageRank method recommended by SubNet from a set of module identification methods [15] and ClustEx [12]. Finally, ClustEx2 was applied to analyze the drug response data derived from TCGA [18] and identified gene modules that are significantly associated with known oncogenic processes, such as oxidation reduction in CESC patients.

Considering its improved performances and ease-of-use, we offer source codes and binary files of ClustEx2 on both Windows and Linux. For broader usage, we would like to clarify some minor problems here. The clustering can lead to one or two large modules and dozens of small modules. This may be caused by the reason that the gene incorporation and module merging is very similar to the single linkage criteria, which favors modules to merge or incorporate genes by the minimum shortest distance. In future work, we would like to adopt average linkage or complete linkage criteria. Another problem is that the calculation of diffusion kernel is time consuming. We recommend that users can keep the diffusion kernel result stored in a text file for parameter tuning.

Acknowledgment

Thanks Dongfang Wang for the design of web-server version. Thanks Ting Wang for writing advices.

REFERENCES

[1] T. J. Hudson, W. Anderson, A. Aretz, and A. D. Barker, "International network of cancer genome projects," Nature, vol. 464, no. 7291, pp. 993–8, 2010.

[2] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. a. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. a. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, and L. a. Garraway, "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," Nature, vol. 483, pp. 603–307, 2012.

[3] W. C. Reinhold, S. Varma, V. N. Rajapakse, A. Luna, F. G. Sousa, K. W. Kohn, and Y. G. Pommier, "Using drug response data to identify molecular effectors, and molecular 'omic' data to identify candidate drugs in cancer.," Hum. Genet., vol. 134, no. 1, pp. 3–11, 2015.

[4] D. W. Huang, R. a Lempicki, and B. T. Sherman, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.," Nat. Protoc., vol. 4, no. 1, pp. 44–57, 2009.

[5] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," Proc Natl Acad Sci U S A, vol. 102, no. 43, pp. 15545–15550, 2005.

[6] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology.," Nature, vol. 402, no. 6761 Suppl, pp. C47–C52, 1999.

[7] M. Girvan, M. E. J. Newman, M. Girvan, M. E. J. Newman, and M. E. J. Newman, "Community structure in social and biological networks," Proc. Natl. Acad. Sci. U. S. A., vol. 99, no. 12, pp. 7821–7826, 2002.

[8] T. Ideker, O. Ozier, B. Schwikowski, and A. F. Siegel, "Discovering regulatory and signalling circuits in molecular interaction networks.," Bioinformatics, vol. 18 Suppl 1, pp. S233–S240, 2002.

[9] S. Nacu, R. Critchley-Thorne, P. Lee, and S. Holmes, "Gene expression network analysis and applications to immunology.," Bioinformatics, vol. 23, no. 7, pp. 850–858, 2007.

[10] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis.," Mol. Syst. Biol., vol. 3, no. 140, p. 140, 2007.

[11] M. Szalay-Beko, R. Palotai, B. Szappanos, I. A. Kovács, B. Papp, and P. Csermely, "ModuLand plug-in for Cytoscape: Determination of hierarchical layers of overlapping network modules and community centrality," Bioinformatics, vol. 28, no. 16, pp. 2202‑2204, 2012.

[12] J. Gu, Y. Chen, S. Li, and Y. Li, "Identification of responsive gene modules by network-based gene clustering and extending: application to inflammation and angiogenesis.," BMC Syst. Biol., vol. 4, p. 47, 2010.

[13] F. Vandin, E. Upfal, and B. J. Raphael, "Algorithms for detecting significantly mutated pathways in cancer," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, vol. 6044 LNBI, pp. 506–521.

[14] Y. Chen, J. Gu, D. Li, and S. Li, "Time-course network analysis reveals TNF-alpha can promote G1/S transition of cell cycle in vascular endothelial cells," Bioinformatics, vol. 28, no. 1, pp. 1–4, 2012.

[15] Q. Zhang and Z. D. Zhang, "SubNet: A Java application for subnetwork extraction," Bioinformatics, vol. 29, no. 19, pp. 2509–2511, 2013.

[16] J. Lafferty and R. I. Kondor, "Diffusion Kernels on Graphs and Other Discrete Input Spaces," ICML '02 Proc. Ninet. Int. Conf. Mach. Learn., pp. 315--322, 2002.

[17] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," World Wide Web Internet Web Inf. Syst., vol. 54, no. 1999–66, pp. 1–17, 1998.

[18] Z. Ding, S. Zu, and J. Gu, "Evaluating the molecule-based prediction of clinical drug responses in cancer," Bioinformatics, p. btw344, 2016.

[19] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software Environment for integrated models of biomolecular interaction networks," Genome Res., vol. 13, no. 11, pp. 2498–2504, 2003.

[20] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function.," Mol. Syst. Biol., vol. 3, no. 88, p. 88, 2007.

[21] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, and R. Sharan, "Associating genes and protein complexes with disease via network propagation," PLoS Comput. Biol., vol. 6, no. 1, 2010.

[22] S. Köhler, S. Bauer, D. Horn, and P. N. Robinson, "Walking the Interactome for Prioritization of Candidate Disease Genes," Am. J. Hum. Genet., vol. 82, no. 4, pp. 949–958, 2008.

[23] X. Ma, T. Chen, and F. Sun, "Integrative approaches for predicting protein function and prioritizing genes for complex phenotypes using protein interaction networks," Brief. Bioinform., vol. 15, no. 5, pp. 685–698, 2013.

[24] H. P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 1, no. 3, pp. 231–240, 2011.

[25] D. Wishart, "Mode analysis: A generalization of nearest neighbor which reduces chaining effects," Numer. Taxon., vol. 76, pp. 282–311, 1969.

[26] J. Millán, R. J. Cain, N. Reglero-Real, C. Bigarella, B. Marcos-Ramiro, L. Fernández-Martín, I. Correas, and A. J. Ridley, "Adherens junctions connect stress fibres between adjacent endothelial cells.," BMC Biol., vol. 8, no. 1, p. 11, 2010.

[27] A. L. de Jong, D. M. Green, J. A. Trial, and H. H. Birdsall, "Focal effects of mononuclear leukocyte transendothelial migration: TNF-alpha production by migrating monocytes promotes subsequent migration of lymphocytes," J Leukoc Biol, vol. 60, no. 1, pp. 129–136, 1996.

[28] F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," N. Engl. J. Med., vol. 363, no. 1, pp. 1–3, 2015.

[29] C. G. Zhou B, Sun C, Li N, Shan W, Lu H, Guo L, Guo E, Xia M, Weng D, Meng L, Hu J, Ma D, "Cisplatin-induced CCL5 secretion from CAFs promotes cisplatin-resistance in ovarian cancer via regulation of the STAT3 and PI3K/Akt signaling pathways," Int. J. Oncol., vol. 48, no. 5, pp. 2087–2097, 2016.

[30] L. Galluzzi, I. Vitale, J. Michels, C. Brenner, G. Szabadkai, A. Harel-Bellan, M. Castedo, and G. Kroemer, "Systems biology of cisplatin resistance: past, present and future," Cell Death Dis., vol. 5, no. 5, p. e1257, 2014.

[31] S. Dasari and P. Bernard Tchounwou, "Cisplatin in cancer therapy: Molecular mechanisms of action," Eur. J. Pharmacol., vol. 740, pp. 364–378, 2014.