Citation: CPT Pharmacometrics Syst. Pharmacol. (2017) 6, 168–176; doi:10.1002/psp4.12167

ORIGINAL ARTICLE

IODNE: An Integrated Optimization Method for Identifying the Deregulated Subnetwork for Precision Medicine in Cancer

S Mounika Inavolu[1,2], J Renbarger[3], M Radovich[1,2], V Vasudevaraja[1,2], GH Kinnebrew[1,2], S Zhang[1,2] and L Cheng[1,2,3]*

Subnetwork analysis can explore complex patterns of entire molecular pathways for the purpose of drug target identification. In this article, the gene expression profiles of a cohort of patients with breast cancer are integrated with protein-protein interaction (PPI) networks using, simultaneously, both edge scoring and node scoring. A novel optimization algorithm, integrated optimization method to identify deregulated subnetwork (IODNE), is developed to search for the optimal dysregulated subnetwork of the merged gene and protein network. IODNE is applied to select subnetworks for Luminal-A breast cancer from The Cancer Genome Atlas (TCGA) data. A large fraction of cancer-related genes and the well-known clinical targets, *ER1*/*PR* and *HER2*, are found by IODNE. This validates the utility of IODNE. When applying IODNE to the triple-negative breast cancer (TNBC) subtype data, we identified subnetworks that contain genes such as *ERBB2*, *HRAS*, *PGR*, *CAD*, *POLE*, and *SLC2A1*.
CPT Pharmacometrics Syst. Pharmacol. (2017) **6**, 168–176; doi:10.1002/psp4.12167; published online 7 March 2017.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ The topic was to design an optimization-searching algorithm, IODNE, which uses the intersection of gene and protein networks to obtain a dysregulated subnetwork for drug target selection in a cohort of patients.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ How to combine networks information of gene-gene and protein-protein? How to search an optimum gene subnetwork for drug treatment for a cohort of patients?

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE
☑ Networks' combination between gene regulatory networks and PPI networks (PPIs). A gene subnetwork searching of drug targets for breast cancer subtypes Luminal-A and TNBC.

HOW THIS MIGHT CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS
☑ Integrating gene and protein network and pharmacology, an optimizing subnetwork searching for drug targets selection holds the promise of expanding the current opportunity space and a cohort of patient's therapeutics in the clinic systematically.

Systematic network analysis on cancer has multiple potential biological and clinical applications. A better understanding of the effects of gene/protein interaction may lead to the identification of cancer genes and correlated pathways, which, in turn, may offer better targets for drug development in cancer treatment.[1] Genomewide mRNA expression data provide a rich resource for studying the molecular mechanisms of cancer.[2] The reconstruction or "reverse engineering" of gene regulatory networks, which aim to find the underlying network of gene–gene interactions from the measurement of gene expression, is considered one of most important goals in systems biology.[3–5] Gene expression products are most often proteins. Large scale statistical analysis shows that the single gene signal transduction from gene expression to protein amount is not synchronous, but random.[6] The protein-protein interaction (PPI) network provides a fundamental basis for understanding the role of proteins within the cell by examining their physical and/or functional associations. Pathway Commons[7] is a comprehensive public pathway database (http://www.pathwaycommons.org/), integrating Human IntAct,[8] BioGrid,[9] human protein reference database,[10] Kyoto Encyclopedia of Genes and Genomes,[11] and 10 more famous PPI datasets. A large amount of gene network research is based on either gene or protein data.[12] However, according to literature,[13] the protein pairs encoded by coexpressed genes interact with each other more frequently than random proteins. This suggests that if gene-gene and PPIs are taken into account simultaneously in one computational method, it will increase the accuracy of identification of interactions of both types as well as allow the recognition of overlapping patterns between gene and protein networks. In this article, the gene expression profile of a cohort of patients with cancer is used to generate coexpression networks, which are then integrated with PPI networks for observing gene variation systematically.

Gene/protein interaction networks can guide us in understanding the gene module molecular mechanisms in a system biology level.[14] However, an exhaustive dataset that holds all of the interactions between genes or proteins in major

[1]Center for Computational Biology and Bioinformatics, School of Medicine, Indiana University, Indianapolis, Indiana, USA; [2]Department of Medical and Molecular Genetics, School of Medicine, Indiana University, Indianapolis, Indiana, USA; [3]Department of Pediatrics, Hematology/Oncology, School of Medicine, Indiana University, Indianapolis, Indiana, USA. *Correspondence: L Cheng (lijcheng@iupui.edu)
Received 5 December 2016; accepted 6 January 2017; published online on 7 March 2017. doi:10.1002/psp4.12167

biological pathways in cancer is difficult to find. Searching for condition-specific subnetworks can help us identify the most significant subsets of the holistic cancer genome. In general, subnetwork search algorithms can be broadly classified into two types[15–18]: (1) seed-initiated algorithms; and (2) scoring search algorithms.[19] Seed-based algorithms involve a set of predefined significant genes that aid in the detection of the subnetworks, such as SubNet,[20] which uses a version of Google's PageRank algorithm for seed detection. Scoring search algorithms work by treating all genes as having equal significance initially, deploy a scoring function to rank the genes, and finally use a search algorithm to extract the subnetworks. The typical scoring strategies include edge-based scoring, node-based scoring, and combined edge and node-based scoring. Guo *et al*.[21] introduce an edge-based scoring and search approach for extraction of a PPI subnetwork responsive to conditions related to some investigated gene expression profiles. However, the node (gene) function is ignored in the algorithm. Node-based scoring has also been used to identify disease-specific genes. Dezso *et al*.[22] score each node by the number of paths traversing that node in the disease-specific network in relation to the number of paths going via the same node in the global network. However, edge weights are not taken into account. Amgalan & Lee[23] proposed the weighted maximum clique method to identify a condition-specific subnetwork by both edge-based scoring and node-based scoring. However, drug targets are not discussed. So far, there is still a lack of a tool for extracting subnetworks from integrated gene-gene and PPI networks for the most significant dysregulated network in cancer, especially with regard to drug treatment in a specific condition. We propose a novel integrated optimization method for identifying the dysregulated subnetwork (IODNE) for a cancer cohort. Edge and node scoring jointly measure condition-specific changes to both gene-gene coexpressions and PPIs. Unfortunately, the derivation of a computational model that extracts the optimum subnetwork is a nondeterministic polynomial time-hard problem[24] and biomolecular networks are massive in scale. This huge dimension makes it tedious to extract globally optimum subnetworks. The Kruskal tree algorithm is a fast searching strategy to find the most connecting significant genes by the shortest spanning subtree of a gene connection network.[25]

Breast cancer is one of the most common cancer types and is the second leading cause of cancer deaths in women.[26] Nearly 40% of patients with breast cancer lack specific gene biomarker identification and have to receive chemotherapy over-treatment and suffer from its strong side effects in the clinic.[27,28] The estrogen receptor (ER) or progesterone receptor (PR) positive tumors (i.e., Luminal-A) usually receive endocrine (hormone) therapy as the standard treatment.[29] The *HER2* amplified tumor responds well to *HER2* targeted trastuzumab, which is currently the standard. However, triple-negative breast cancer (TNBC; *ER*−/*PR*−, *HER2*−) is the most aggressive tumor type and has a much shorter overall survival than the other tumor types. Currently, chemotherapy is still the main therapy for TNBC.[30] The Cancer Genome Atlas (TCGA)[31] provides comprehensive cancer genome profiles for more than 14,000 cancers, including gene expression profiles. This rich source of data provides us an opportunity to detect the molecular variation in subtype-specific breast cancer.

In this article, an integrated optimization method (IODNE) is proposed for the identification of the maximum dysregulated subnetwork for drug treatment in a subtype of patients. We use a strategy of subnetwork retrieval, which depends on the gene-gene control network merged with PPIs and rank subnetworks by scoring both edges and nodes. Comparison of transcriptomes between a subtype of patients with breast cancer and a corresponding normal group is used to construct the gene-gene control network. A modified Kruskal minimum spanning tree search strategy determines the maximum dysregulated subnetwork for drug treatment in a cohort of patients. The novel algorithm is validated by its ability to select previously known drug target genes in Luminal-A breast cancer from TCGA. IODNE is applied to TNBC for drug-target subnetwork identification.

METHODS

Materials

We attempted to find the major differences in network patterns between two groups from their specific gene regulatory networks and a prior knowledge of a protein-protein network. The gene expression profiles pertaining to Luminal-A and TNBC were collected from TCGA,[27] including tumor samples and their adjacent normal tissue. The expression data was derived from the Agilent-G450-2A Array-based platform, which consisted of 17,815 genes. There are nine pairs of TNBC samples with their respective adjacent-normal samples. The Luminal-A expression set consisted of 43 pairs of tumors and corresponding adjacent-normal samples. **Supplementary File S1** provides clinical information and subtypes of breast cancer. Pathway Commons 2 (version 7)[10] (http://www.pathwaycommons.org) is a well-known collection of biological pathways and PPI network knowledge. Pathway Commons consists of 31,698 pathways, 1,912,848 edges (PPI), and 14,863 nodes (genes), which were used for this analysis. Drug-targeted genes were annotated by DrugBank database version 4.0[32] (http://www.drugbank.ca/drugs). We collected 1,623 US Food and Drug Administration (FDA) approved drugs and their 1,770 targets (**Supplementary File S8**). According to the National Cancer Institute and the National Comprehensive Cancer Network annotation,[33] 134 FDA approved cancer drugs and 322 target genes are "signed," which are annotated as either enhancing or repressing the target (**Supplementary File S9**).

The preprocessing pipeline of IODNE is well structured to organize the bulk gene expression data into the desired input format suited for the main run, as shown in **Table 1**. The rows show each gene's expression and the columns show different samples. The first row shows the information of the group samples, such as tumor or adjacent normal group; and the second row shows the subtypes of tumors. IODNE is capable of recognizing the various subtypes of cancers and the type of tissue each sample pertains to from a single file. The PPI data can be simply a two-column comma-separated file with each row containing the names of interacting genes.

Table 1 Samples input formats for two groups' comparison

Groups	Adjacent normal	Adjacent normal	Tumor	Tumor
Subtypes	Basal-like	Luminal A	Basal-like	Luminal A
Samples genes	TCGA-A7-A0CE-11A-21R-A089-07	TCGA-A7-A0CH-11A-32R-A089-07	TCGA-A7-A0CE-01A-11R-A00Z-07	TCGA-A7-A0CH-01A-21R-A00Z-07
FKBPL	−0.92533	−0.639	−0.23283	0.324167
COL10A1	0.71875	2.121	4.655	6.3255
KIF26B	1.4585	0.28925	1.27575	2.76125

Integrated optimization algorithm

IODNE achieves detection of the subnetworks based on two groups' transcriptome comparison and existing PPIs, such as the tumors and the adjacent tumor normal groups. **Figure 1** shows the IODNE algorithm workflow. It consists of four modules: (1) gene expression-based scoring; (2) protein interaction-based scoring; (3) two networks node and edge scores integration; and (4) identification of the maximal scoring subnetwork algorithm. The genes are first scored based on the gene-expression data. For this, the concept of Pearson correlation is applied. Second, the protein interaction network scoring is based on the influence of the genes and gene pairs on the PPI network at large. This strategy is a partial adaptation of the Vandin *et al.*[34] HotNet algorithm. Third, edge scores and the node scores from the gene control network and the PPIs network score sets are scaled by an associated scale value for each network and combined by summation. The scoring functions that evaluate the edge scores of gene pairs and the node scores of individual genes are given below. Fourth, the scored network is subjected to the search strategy for the maximal edge-scoring subnetwork, with a size below a user-defined threshold number of genes. The search strategy uses a customized version of the modified minimum spanning tree Kruskal algorithm.[35,36] Each of the steps are described below, after descriptions of the data required for the analysis and the preprocessing.

Step 1. Gene-expression based scoring strategy on nodes and edges in construction gene regulatory network

Node score. To compare two groups' transcriptomes, such as tumor and normal groups, a two group *t*-test with equal variance for unequal sample sizes is performed to find differentially expressed genes (DEGs) as follows:

$$t = \frac{\bar{X}_n - \bar{X}_t}{S_{X_t X_n} \cdot \sqrt{\frac{1}{n_t} + \frac{1}{n_n}}}$$

where $S_{X_t X_n} = \sqrt{\frac{(n_t-1){S_{X_t}}^2 - (n_n-1){S_{X_n}}^2}{n_t+n_n-2}}$; $\bar{X}_t$ = mean of the tumor samples; $\bar{X}_n$ = mean of the normal samples; n_t = number of tumor samples; n_n = number of normal samples; S_{X_t} = standard deviation of the tumor samples; and S_{X_n} = standard deviation of the normal samples. Given the degrees of freedom: $df = n_t + n_n - 2$, checking the P value p associated with a t-value t for two-tailed t-test in the t-distribution. The sign of t means the gene is down (t >0) or up (t <0) regulation in tumors comparing with normal group. Here, we set gene (node) score as abstract t-value t in gene-gene interaction network: $NS_1 = |t|$.

Edge score. The scoring pipeline of IODNE starts with the scoring of the gene-expression profiles. For every pair of gene *x* and gene *y*, we calculate Pearson correlation coefficients between *x* and *y* as follows:

$$r = \sum \frac{(x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

where x, y = individual values of the gene in the cohort (tumor or normal group); $\bar{x}$, $\bar{y}$ = average of the gene across the samples in tumor or normal group. Thus, Pearson correlation coefficient of the gene pair (*x*, *y*) is calculated in the tumor group as $r_t(x, y)$ and the normal group as $r_n(x, y)$, respectively. The differential correlation of gene pair (*x*, *y*) shows the control distance difference between the tumor group and the normal group denote as:

$$\Delta r = |r_n - r_t|.$$

This represents the edge score of each gene pair (*x*, *y*) in the gene-gene interaction network: $ES_1 = \Delta r$.

Step 2. Protein-protein interaction based scoring strategy in nodes and edges

A Laplacian matrix is used to discover the relationship between network topological structure and its gene connection features in the PPI network, as in the literature.[35] Given a simple PPI graph, *G* with *n* genes (nodes), the symmetric normalized Laplacian matrix is defined as:

$$L = D - A$$

where $A = (a_{ij})_{n\times n}$ is the binary adjacency matrix of PPI graph G, a_{ij} is equal to 1 or a value ranging from 0 to 1. If gene *i* and *j* are linked in the network $a_{ij}=1$, otherwise $a_{ij}=0$; $D = (d_{ij})_{n\times n}$ is the degree matrix of graph *G*, where $d_{ii} = \sum_{j=1}^{n} a_{ij}$, and $d_{ij}=0$, for $i \neq j$.

A Hotness matrix is defined to reflect the influences of one gene against other genes[34] as follows:

$$H = e^{-Lc}$$

where $H = (h_{ij})_{n\times n}$ represents the connection influence between gene *i* and gene *j*, c is a constant, and default value is $c = 0.1$.

Node score. Node score σ_x is to evaluate the gene *x* influence to the whole PPIs network. It is calculated as the sum of all the influences of the genes with its connections, which is the sum of row margins to *Matrix* $A*H$. For instance, if a gene *x* has two connected genes *y* and *z*, the node score of gene *x* would be:

$$\sigma_x = h(x, y) + h(x, z).$$

Edge score. The edge score for each of the gene pairs (*x*, *y*) is calculated from the *max* node score of the individual genes of the pair multiplied with the binary adjacency score of the gene pair as follows:

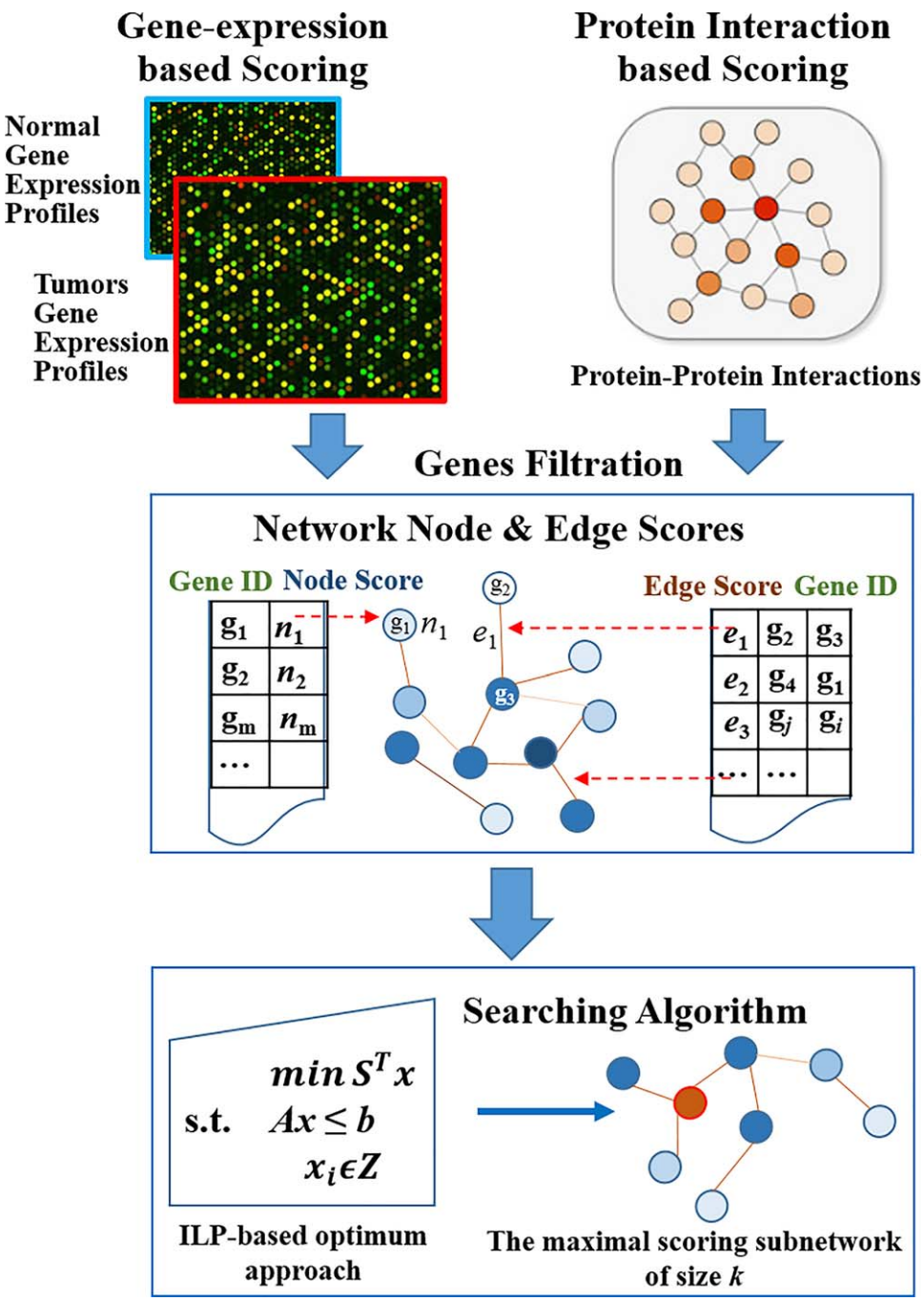

Figure 1 The overview of integrated optimization method for identifying the dysregulated subnetwork showing the scoring and search workflow.

$$w(x, y) = \max\{\sigma_x, \sigma_y\} * a_{xy}.$$

Denoted the node score and edge scores as $NS_2 = \sigma$, $ES_2 = w$.

Step 3. Node score and edge score integration of gene regulatory network and protein-protein integration network

In order to maintain the equal influences from the two networks in integration, scaling node score and edge score need be done by the ratio of maximum scores from either scoring to each of the genes in the two networks as follows:

$$\text{Node weights}: \ NS_{net} = NS_1 + \propto * NS_2$$

$$\text{Edge weights}: \ ES_{net} = ES_1 + \beta * ES_2$$

where $\propto = \{max\ NS_1/max\ NS_2\}$, $\beta = \{max\ ES_1/max\ (ES_2)\}$.

In the merged network, node weights (gene) carry two messages, one is the gene expression difference of tumors verse normal group, and another one is the gene influence

ability (connection degree) in PPIs. Meanwhile, edge weight for a pair of genes carries the similar two messages: the pair-gene connection difference of tumors vs. its normal group in the gene regulatory network, and the gene connection strength in PPIs.

Step 4. Search strategy

The IODNE algorithm uses both informative node weights and edge weights derived from prior knowledge. It sorts all the edge weights in a descending manner, and then extracts subnetworks with *k*-connected genes in sequence based on the modified Kruskal searching algorithm by drug-target gene weight ranking. Given an undirected graph $G = (V, E, w(v), w(e))$, where V is a set of nodes with weight $w(v)$, E is a set of edges with weight $w(e)$. IODNE algorithm is an integer linear programming-based optimum algorithm described as follows. The minimum spanning tree by the node and edge weight is used to search the subgraph to a limited number size k:

$$min\, C / \left(\sum_{i=1}^{k_1} w(v_i) + \sum_{j=1}^{k2} w(e_j) \right)$$

$$s.t.\ k_1 \leq k$$

C is constant.

IODNE is used to extract a subset of the edges in a given undirected graph with three properties: (1) it includes k vertices in the subgraph; (2) the total weight of all the edges is as largest as possible; and (3) there is no cycle in the subgraph. At the same time, two enforcing functions that counteract and refine the subnetworks are added: (1) orphan node pairs that are not connected to the parent subnetwork are removed; (2) for the nodes with an abnormally high number of leaves (greater than a declared threshold – defaulted to 30), the five edges with the largest weights are kept and the remaining are pruned out of the subnetwork. A more elaborate explanation is as follows:

Algorithm: Modified Kruskal minimum spanning tree algorithm.
Input: (1) A weighted, undirected graph $G = (V, E, w(v), w(e))$ and (2) the maximum allowed nodes per graph k.
Output: A maximally edge-weighted tree T.
Procedure: Sort the nodes (drug target genes) in V in decreasing order by node-based score (weight).
Select the largest score node $v_{largest}$ in V.
$T \leftarrow$ Select the largest edge $(\boldsymbol{u}, \boldsymbol{v})$ for node $\boldsymbol{v}_{largest}$ in its edge list.
Sort the edges in E in decreasing order by edge-based score (weight) around nodes u, v.
For each edge (u_1, v_1), in sorted order:
Determine the closest neighbor of the selected edge.
$x \leftarrow$ Find (u_1).
$y \leftarrow$ Find(v_1).
If $x \neq y$ then:
$T \leftarrow T \cup \{(u_1, v_1)\}$
Union (x, y)
Until the total node size in the network is $\boldsymbol{k}$.

Additionally, it is sometimes necessary to filter out nonsignificant genes before we use our algorithm on large datasets. A typical experiment is to select the DEGs by comparing the transcriptome profiles between an experimental group and its control group. The dysregulated genes are then selected using one of multiple possible criteria. For example, the dysregulated genes can be defined as those genes that surpass a fold change threshold and P value threshold after a two-group t-test. These genes will then be investigated for connected subgraphs using the IODNE algorithm.

Step 5. Software developing and code

IODNE code is written in Java and compiled with Java Development Kit version 1.7. Its installation and run notes are available in the **Supplementary Code File**. The original program is accessible to: https://drive.google.com/open?id=0B9-cfjp2iWgONjU4a3hwQ2RKcW8.

RESULTS

Here, we show that IODNE selects subnetworks, which contain *ER1*/*PR* and *HER2*, the well-known targets in clinical Luminal-A breast cancer subtypes, validating our results. In addition, IODNE is applied to TNBC, providing potentially novel gene targets for drug selection.

Subnetwork identification for Luminal-A subtype of breast cancer

DEGs to 43 pairs of TCGA tumor and corresponding adjacent-normal samples are calculated by unpaired two-group t-test in the tumor vs. the normal group first (**Supplementary File S2**). IODNE begins with the application of t-value or P value thresholds filters to the gene list that is input into the algorithm by user desired. In the analysis, nondysregulated genes were filtered by P value $P < 0.15$ to reduce the noise in the output as well as substantially increase the runtime performance of the algorithm. A relatively high P value threshold is selected here in order to keep many of the transcription factors and enzymes in IODNE, such as well-known genes *ESR1* and *PGR*. These genes exert important control over the gene network, despite that they do not show statistically significant expression variation between tumor vs. normal samples.

Our aim here was to use the DEG characteristics to identify possible targets for drug development in integrated gene and protein network. Genes that are expressed more strongly in tumors than the adjacent normal tissue (t-value $t < 0$ upregulation gene in tumors, $P < 0.05$) can potentially serve as drug targets. **Supplementary Files S4 and 5** provide the gene list of potential cancer drugs and FDA approved drugs, respectively. However, we need discover which targets are the optimum ones and the hub genes (the nodes with high degree in the network), which are yielded with the correspondingly high edge weights in the integrated gene and protein networks.

The highest edge weight subnetwork in Luminal-A breast cancer was obtained using the IODNE algorithm with size $k = 36$ and the FDA approved drug target list used for ranking nodes' weights. The analysis retrieved the subnetwork containing 36 genes connected with 58 edges. Cytoscape

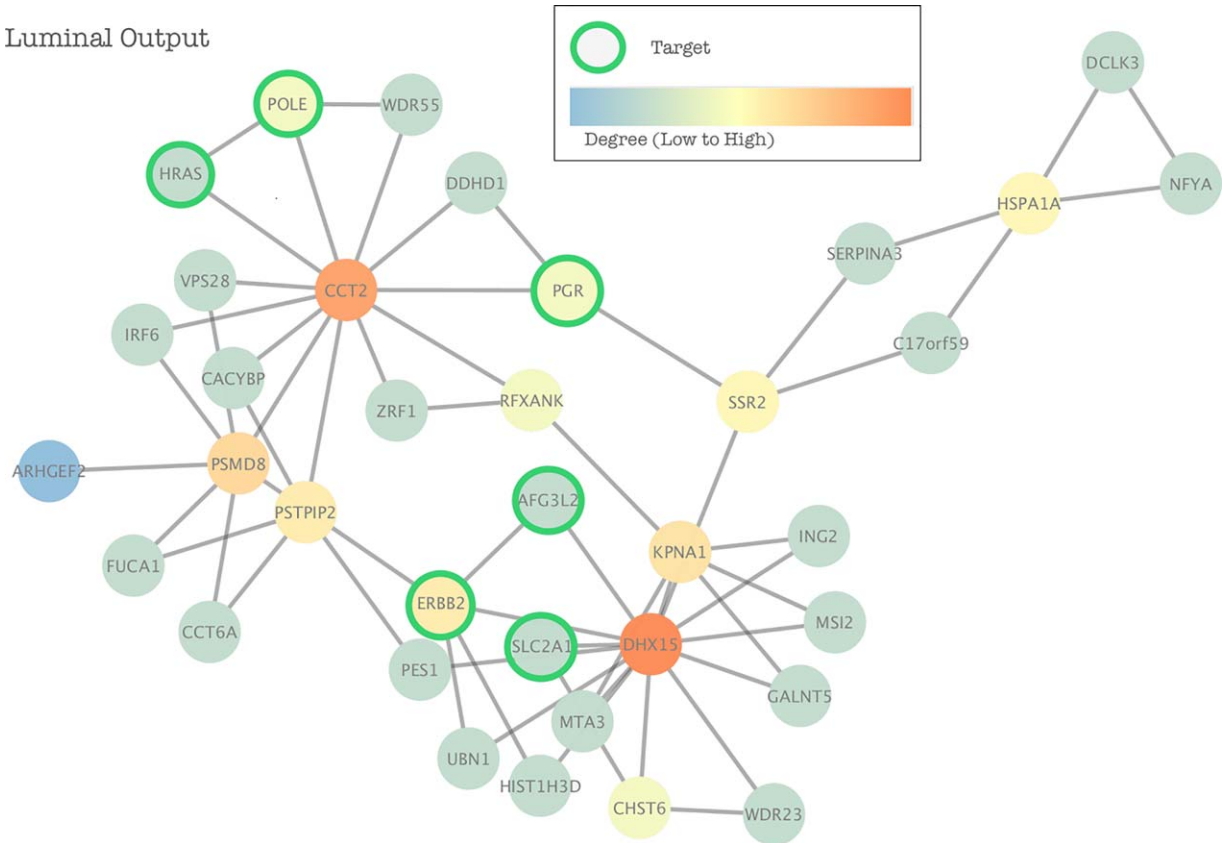

Figure 2 Subnetwork of subtype Luminal-A breast cancer.

3.4[37] was used to visualize these selected genes' interaction networks and denote the drug targeted genes from DrugBank annotations (**Figure 2**). Here, the nodes are colored according to degree to make obvious the hubs of the subnetwork. The subnetwork contains two very prominent biomarkers – PGR and ERBB,[38] as well as some new genes never reported as drug targets before, including *HRAS, POLE, AFG3L2,* and *SLC2A1*.

Subnetwork identification for the triple-negative breast cancer subtype of the breast cancer application

We analyzed the gene expression profiles of nine pairs of cancer and adjacent normal samples of patients with TNBC from TCGA. A listing of DEGs by unpaired two-group *t*-test for the tumors vs. the normal samples is provided in **Supplementary File S3**. IODNE was used to select the dysregulated optimum subnetwork for the TNBC subtype of breast cancer with the FDA approved drug target list used for ranking nodes' weights in the integrated gene and protein networks. The genes were filtered by *P* values (< 0.05) and the *t*-value (<0) for significant DEGs between the case and adjacent-normal groups. These significant DEGs were input into the IODNE algorithm, with a *k* value of 59. **Supplementary Files S6 and S7** list the target genes of potential cancer drugs and FDA approved drugs, respectively, for TNBC. **Figure 3** shows the subnetwork for the TNBC subtype of breast cancer consisting of 59 genes and 112 edges by Cytoscape 3.4 visualization. The network contains 12 drug targets by DrugBank annotation[39] that are highlighted in green in the subnetwork. The most connected hub is glyceraldehyde 3-phosphate dehydrogenase, which has been shown to play a crucial role in cancer regulation,[40] although it does not show significant differential expression between TNBC tumor samples vs. their adjacent-normal samples. The other targets include *HSP90B1*, *ME2*, *TUBB3*, *APRT*, *AXIN1*, *CAD*, *RDH13*, *POLE*, *DAPK1*, *HDAC6*, and *SLC2A1*.

DISCUSSION

Subnetwork analysis can explore complex patterns in the entire network of molecular pathways for the purpose of cancer genes and drug target identification. IODNE is a robust and powerful algorithm for the identification of subnetworks for cancer genes and drug targets. The application of the IODNE for subnetwork selection of drug target in breast cancer subtypes Luminal-A show well-known targets *PGR* and *ERBB2* in clinical are found. In addition, our results show IODNE can find genes in the subnetwork, which play a significant role in breast cancer. The *HRAS* gene is directly related to the tumor aggressiveness in

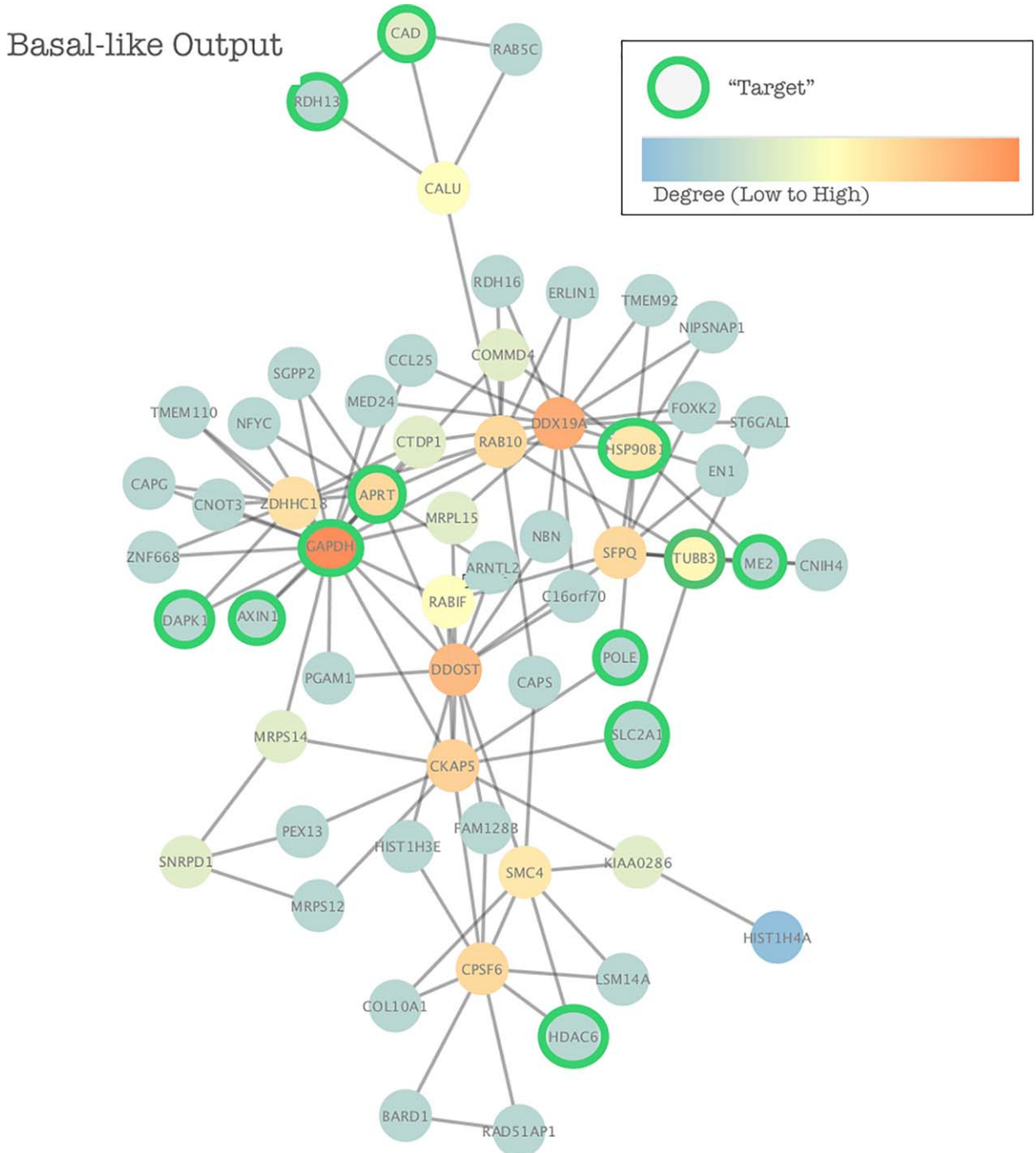

Figure 3 Subnetwork for triple-negative breast cancer subtype of breast cancer.

breast cancer[41] and the *POLE* gene has been linked to the increased risk of colorectal cancer.[42] The most connected hub genes in the subnetwork are *DHX15* and *CCT2*. *CCT2* is found to be necessary for growth/survival of breast cancer cells *in vitro*.[43] All of these results would seem to validate the utility of the IODNE algorithm. TNBC treatment is centered on chemotherapy. The discoveries in the molecular profiling of TNBC press the need to explore new targets in TNBC at the intersection of precision medicine and molecular profiling. IODNE is applied to TNBC data for the drug-target subnetwork identification. The genes *ERBB2*, *HRAS*, *PGR*, *CAD*, *POLE*, and *SLC2A1* are identified as significant in TNBC, which strong research evidence[28] identifies as drug-target candidates.

IODNE has many advantages, some of which include: (i) the scoring strategy takes into account both gene expression and PPI profiling. This gives a multidimensional assessment in finding the most significant subnetwork.

(ii) The node scores and edge scores are rationally combined through scaling parameters. This scaling is crucial in balancing both the gene expression and PPI metrics without a bias. (iii) The search algorithm is tailored to consider every edge in the network while deriving the most significant network with the largest connection. The final subnetworks are free from orphan edges increasing the comprehensiveness of the output. Unfortunately, the IODNE approach has limitations too. We did not define a re-sampling subnetwork mechanism to remove potential false-positive issues under these huge sets of gene interactions. In addition, IODNE cannot extend network-based drug target selection to precision medicine on an individual patient basis. Two potential avenues for extending the current algorithm include: (1) incorporating mutational and copy number variation profiling data along with gene expression profile and drug targets under a systems biology framework may lead to significant improvements in precision oncology.[44] Driver gene mutations tend to have a selective growth advantage in tumor cells and play a disproportionate role in cancer biology. The possibility of targeting driver mutations in a gene control network, which can be further studied in the emerging field of precision cancer medicine.[44–46] (2) Extending the IODNE algorithm to an individualized systems medicine approach to optimize precision cancer therapies to be more safe and effective for individual patients is another important direction for precision cancer medicine.[47] That would enable IODNE to extend its function to a single-patient level as well.

Acknowledgments. This work was supported by National Institutes of Health Funding 1U54HD090215-01 and the Walter Cancer Foundation, Walter Bioinformatics-Molecular Genomics/Genetics Join Indiana University-Purdue University Initiative Funding 0154.0.

Conflict of Interest. The authors declare no conflicts of interest.

Author Contributions. L.C., K.G.H., and S.Z. wrote the manuscript. J.R. and M.R. designed the research. S.M.I. and V.V. analyzed the data.

1. Barabási, A.L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
2. Liu, Y., Koyutürk, M., Barnholtz-Sloan, J.S. & Chance, M.R. Gene interaction enrichment and network analysis to identify dysregulated pathways and their interactions in complex diseases. *BMC Syst. Biol.* **6**, 65 (2012).
3. Rhodes, D.R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
4. Liu, Y. *et al.* Systems biology analyses of gene expression and genome wide association study data in obstructive sleep apnea. *Pac. Symp. Biocomput.* 14–25 (2011).
5. Zhang, X. *et al.* Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* **28**, 98–104 (2012).
6. Yazdanparast, A., Li, L., Radovich, M. & Cheng, L. Signal translational efficiency between mRNA expression and antibody-based protein expression for breast cancer and its subtypes from cell lines to tissue. *Int. J. Comput. Biol. Drug Design* (in press).
7. Cerami, E.G. *et al.* Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**(Database issue), D685–D690 (2011).
8. Aranda, B. *et al.* The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* **38**(Database issue), D525–D531 (2010).
9. Stark, C. *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* **39**(Database issue), D698–D704 (2011).
10. Prasad, T.S., Kandasamy, K. & Pandey, A. Human protein reference database and human proteinpedia as discovery tools for systems biology. *Methods Mol. Biol.* **577**, 67–79 (2009).
11. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
12. Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B. & Peyvandi, A.A. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol. Hepatol. Bed Bench* **7**, 17–31 (2014).
13. Grigoriev, A. A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. *Nucleic Acids Res.* **29**, 3513–3519 (2001).
14. Mutation Consequences and Pathway Analysis working group of the International Cancer Genome Consortium. Pathway and network analysis of cancer genomes. *Nat. Methods* **12**, 615–621 (2015).
15. Carey, L.A. *et al.* The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* **13**, 2329–2334 (2007).
16. Turner, N.C. & Reis-Filho, J.S. Tackling the diversity of triple-negative breast cancer. *Clin. Cancer Res.* **19**, 6380–6388 (2013).
17. Chen, L., Xuan, J., Riggins, R.B., Wang, Y. & Clarke, R. Identifying protein interaction subnetworks by a bagging Markov random field-based method. *Nucleic Acids Res.* **41**, e42 (2013).
18. Chen, L., Xuan, J., Riggins, R.B., Wang, Y., Hoffman, E.P. & Clarke, R. Multilevel support vector regression analysis to identify condition-specific regulatory networks. *Bioinformatics* **26**, 1416–1422 (2010).
19. Jiang, B. & Gribskov, M. Assessment of subnetwork detection methods for breast cancer. *Cancer Inform.* **13**(suppl. 6), 15–23 (2014).
20. Lemetre, C., Zhang, Q. & Zhang, Z.D. SubNet: a Java application for subnetwork extraction. *Bioinformatics* **29**, 2509–2511 (2013).
21. Guo, Z. *et al.* Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network. *Bioinformatics* **23**, 2121–2128 (2007).
22. Dezso, Z. *et al.* Identifying disease-specific genes based on their topological significance in protein networks. *BMC Syst. Biol.* **3**, 36 (2009).
23. Amgalan, B. & Lee, H. WMAXC: a weighted maximum clique method for identifying condition-specific sub-network. *PLoS One* **9**, e104993 (2014).
24. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18** (suppl. 1), S233–S240 (2002).
25. Thomas, H.C., Leiserson, C.E. & Rivest, R.L. *Clifford Stein Introduction to Algorithms* (Third edn.) (MIT Press, Cambridge, MA, 2009).
26. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**, 74–108 (2005).
27. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
28. Fanale, D., Amodeo, V., Corsini, L.R., Rizzo, S., Bazan, V. & Russo, A. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene* **31**, 2121–2128 (2012).
29. Onitilo, A.A., Engel, J.M., Greenlee, R.T. & Mukesh, B.N. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin. Med. Res.* **7**, 4–13 (2009).
30. Shah, S.P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
31. Ciriello, G. *et al.* Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
32. Wishart, D.S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**(Database issue), D901–D906 (2008).
33. Cheng, L., Schneider, B.P. & Li, L. A bioinformatics approach for precision medicine off-label drug drug selection among triple negative breast cancer patients. *J. Am. Med. Inform. Assoc.* **23**, 741–749 (2016).
34. Vandin, F., Upfal, E. & Raphael, B.J. Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* **18**, 507–522 (2011).
35. Huang, L., Liao, L. & Wu, C.H. Inference of protein-protein interaction networks from multiple heterogeneous data. *EURASIP J. Bioinform. Syst. Biol.* **2016**, 8 (2016).
36. Kruskal, J.B. Jr. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* **7**, 48–50 (1956).
37. Cline, M.S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382 (2007).
38. Zurrida, S., Montagna, E., Naninato, P., Colleoni, M. & Goldhirsch, A. Receptor status (ER, PgR and HER2) discordance between primary tumor and locoregional recurrence in breast cancer. *Ann. Oncol.* **22**, 479–480 (2011).
39. Wishart, D.S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**(Database issue), D668–D672 (2006).
40. Zhang, J.Y. *et al.* Critical protein GAPDH and its regulatory mechanisms in cancer cells. *Cancer Biol. Med.* **12**, 10–22 (2015).
41. Yong, H.Y. *et al.* Identification of H-Ras-specific motif for the activation of invasive signaling program in human breast epithelial cells. *Neoplasia* **13**, 98–107 (2011).

42. Briggs, S. & Tomlinson, I. Germline and somatic polymerase ε and δ mutations define a new class of hypermutated colorectal and endometrial cancers. *J. Pathol.* **230**, 148–153 (2013).
43. Guest, S.T., Kratche, Z.R., Bollig-Fischer, A., Haddad, R. & Ethier, S.P. Two members of the TRiC chaperonin complex, CCT2 and TCP1 are essential for survival of breast cancer cells and are linked to driving oncogenes. *Exp. Cell Res.* **332**, 223–235 (2015).
44. Cheng, F. *et al.* Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* **8**, e1002503 (2012).
45. Cheng, F., Zhao, J., Fooksa, M. & Zhao, Z. A network-based drug repositioning infrastructure for precision cancer medicine through targeting significantly mutated genes in the human cancer genomes. *J. Am. Med. Inform. Assoc.* **23**, 681–691 (2016).
46. Cheng, F.,Zhao, J. & Zhao, Z. Advances in computational approaches for prioritizing driver mutations and significantly mutated genes in cancer genomes. *Brief. Bioinform.* **17**, 642–656 (2016).
47. Cheng, F., Hong, H., Yang, S. & Wei, Y. Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief. Bioinform.*; e-pub ahead of print (2016).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://psp-journal.com)