BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 17 2014, pages 2399–2405
doi:10.1093/bioinformatics/btu199

Genome analysis

Advance Access publication May 7, 2014

PhenoNet: identification of key networks associated with disease phenotype

Rotem Ben-Hamo, Moriah Gidoni and Sol Efroni*

The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 5290002, Israel

Associate Editor: Inanc Birol

ABSTRACT

Motivation: At the core of transcriptome analyses of cancer is a challenge to detect molecular differences affiliated with disease phenotypes. This approach has led to remarkable progress in identifying molecular signatures and in stratifying patients into clinical groups. Yet, despite this progress, many of the identified signatures are not robust enough to be clinically used and not consistent enough to provide a follow-up on molecular mechanisms.

Results: To address these issues, we introduce PhenoNet, a novel algorithm for the identification of pathways and networks associated with different phenotypes. PhenoNet uses two types of input data: gene expression data (RMA, RPKM, FPKM, etc.) and phenotypic information, and integrates these data with curated pathways and protein–protein interaction information. Comprehensive iterations across all possible pathways and subnetworks result in the identification of key pathways or subnetworks that distinguish between the two phenotypes.

Availability and implementation: Matlab code is available upon request.

Contact: sol.efroni@biu.ac.il

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on December 17, 2013; revised on April 9, 2014; accepted on April 10, 2014

1 INTRODUCTION

Beginning in the mid-1990s, through unique conventions on public sharing and through the concentrated efforts of different groups around the world, an ever-growing abundance of transcriptome profiling sets of data have provided the systems medicine community with the foundations of what is now referred to as molecular signatures. Many of these signatures have concentrated on identifying sets of genes that classify patients into relevant clinical affiliations, followed by the use of enrichment tools to further affiliate sets of genes with biological processes. Additional layers of mechanism understanding have been introduced with the use of network metrics to quantify the orchestrated behavior of genes within the transcriptome (Ben-Hamo and Efroni, 2013; Emmert-Streib *et al.*, 2012). These network approaches and the resulting insight suggest that biological, cellular and disease outcome strongly depend on the complex of interactions between genes, proteins and other molecules through different pathways (Ziogas *et al.*, 2011).

The identification of biomarkers that, by definition, correlate with phenotypes, is an important task in biomedical research. By exploring the most significant molecular changes between groups, we may shed light on the structure, centrality, dynamics and operating principles of the disease network and reveal new features and structures. The introduction of network-based metrics carries with it the possibility of identifying biomarkers that have been hiding in plain sight (Khatri *et al.*, 2012).

The integration of experimental data and systems-level approaches, combined with the transition from 'classic' molecular-based analysis to network-based analysis, may be the basis for a new era of novel clinical implications in personalized medicine and for a deeper understanding of disease behavior.

Protein–protein interaction (PPI) networks are well-established tools in the field of systems biology. The Pathway Interaction Database (PID) (Schaefer *et al.*, 2009), Reactome (Joshi-Tope *et al.*, 2005) and Biocarta are examples of publicly available pathway-interaction datasets. By merging the identification of differentially expressed genes with curated network information, key novel and robust subnetworks may be detected as new biomarkers (Ben-Hamo and Efroni, 2012a, b; Efroni *et al.*, 2011; Greenblum *et al.*, 2011). Several pathway-level tools are available for incorporating pathway topology to interpret high-throughput datasets. For example, PPI spider tool (Antonov *et al.*, 2009) uses the Monte Carlo simulation procedure to compute the statistical significance of an inferred model based on the topology of the global PPI network. PARADIGM (Vaske *et al.*, 2010) is a method for inferring patient-specific genetic activities incorporating curated pathway interactions among genes. This method predicts the degree to which a pathway's activities (e.g. internal gene states, interactions or high-level 'outputs') are altered in a patient using probabilistic inference. PathOlogist (Greenblum *et al.*, 2011) is one of the few methods that provide the metrics for quantifying the behavior of interactions in a pathway. Using PathOlogist, every pathway can be quantified into a numeric value per sample.

A growing number of pathway-based works have been recently published and focuses on the identification of novel prognostic biomarkers. Fröhlich (Fröhlich, 2011) derived a consensus signature from seemingly different prognostic gene signatures in breast cancer by incorporating information on PPIs. They reported that the stability of the signature was significantly higher. Wang and Chen (2011) integrated gene expression data with PPI information for the development of a network-based biomarker for lung carcinogenesis. Pierobon *et al.* (2009) used reverse-phase protein microarray analysis of laser capture microdissected CRC tumor specimens to profile broad cell-signaling

*To whom correspondence should be addressed.

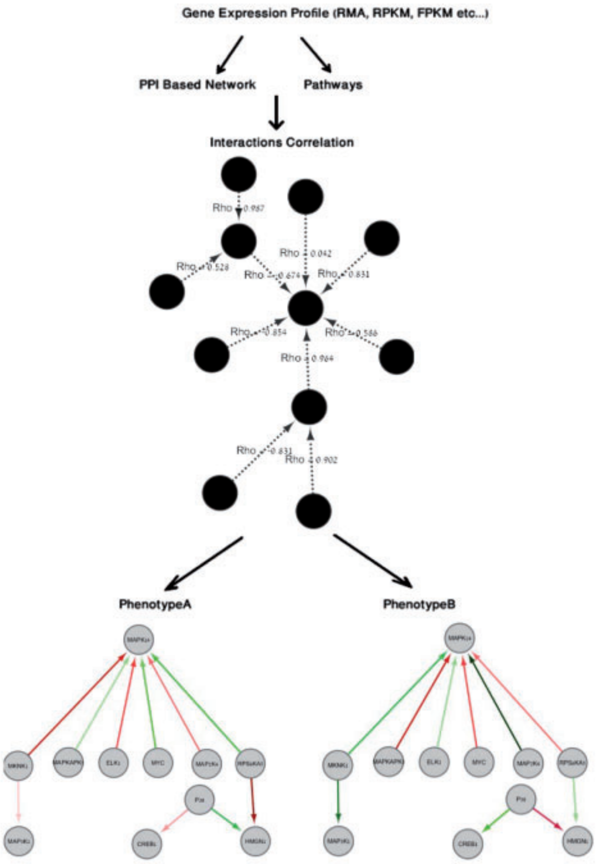

Fig. 1. PhenoNet algorithm pipeline

pathways in patients who exhibited liver metastasis versus patients who remained recurrence free after follow-up. Lee *et al*. (2008) integrated pathway information with disease classification procedure, to classify disease by looking at the activity signaling pathways or protein complexes. Su *et al*. (2009) proposed a classification method based on probabilistic inference of pathway activities. Teschendorff *et al*. (2010) developed a network method for estimating pathway activation in tumors from model signatures. In a recent work by Li *et al*. (2012), the authors provided a server for identifying network-based biomarkers that correlate with patient survival data. Kim and Gelenbe (2012) modified the Gillespie algorithm by adding the Hill function to describe protein activation. They applied their algorithm to simulate expression data that show switch-like and oscillatory behavior in the metabolite production system of microorganisms.

Here, we have developed PhenoNet (Fig. 1), which is a network-based phenotype classification tool that integrates gene expression data (RMA, RPKM, FPKM, etc.), clinical data and PPI knowledge for the identification of key subnetworks whose behavior and derived index stratifies the clinical states. PhenoNet works under the assumption that specific regulation mechanism can control specific phenotype, by doing so PhenoNet is able to detect phenotype-dependent gene-gene associations. By scanning all known PPIs, PhenoNet is able to detect a specific single subnetwork with strong regulation in phenotype A and with a lesser impact regulation in phenotype B. Such associations are identified through measuring correlation and lack of correlation among PPI components. Because of the fact that these specific regulations are present in only one phenotype, we can assume that the absent or present of them is meaningful for disease course.

To the best of our knowledge, PhenoNet is the first tool that is tuned to identify these changes according to their phenotype affiliations. Although available tools are structured to identify mutation or methylation modifications across phenotypes, they are not structured to detect subnetworks that undergo changes in their gene-gene associations. PhenoNet, while completely unconcerned with mutation or methylation status, is structured to identify modifications in associations between gene pairs across a subnetwork. The identification of these changes may be a key for catalyzing new treatment and understanding specific phenotype mechanisms.

At the core of the method is the calculation of correlations between all known PPIs, PhenoNet is able to identify the most varied subnetwork that stratifies tested phenotypes. PhenoNet returns a single subnetwork for every two phenotypes presented. This network is the most varied between the two phenotypes.

2 METHODS

The algorithm developed by us combines PPI information extracted from BIOGRID (Stark *et al*., 2011) and HPRD (Peri *et al*., 2003) with the PID (Schaefer *et al*., 2009), which integrates pathway knowledge from NCI, Kegg (Ogata *et al*., 1999) and Biocarta. This modeling strategy was based on the fact that gene contribution to the network is based on the role played by the gene and the way the gene influences its network neighborhood.

This approach makes use of a baseline model that is constructed using correlations from curated pathways from the PID (Section 2.1). We, then, compared this baseline with further extensions of the PPI network.

2.1 Baseline pathway model

Initially, we obtained all interactions and molecules in the pathway. We then used the gene expression data to calculate the Pearson correlation for each interaction in the pathway. A pathway score was then obtained as the absolute average over correlation coefficients in the pathway. This process was iterated throughout all pathways and across the set of patient samples. It was done simultaneously for both sets of samples to identify the most significant subnetwork that stratifies the two phenotype groups. Each pathway thus culminated in two scores: one score for its value in one phenotype and another score for its value in the other phenotype. Then, using Student's *t*-test, the *P*-value to describe the differences between the pathway scores in the two phenotypes, was obtained. When all pathway scores were calculated, the 'best' pathway was selected. This pathway set the threshold for the PPI network analysis that followed.

2.2 PPI network model

The PPI network compiled by us was composed of 10 489 genes and 62 643 interactions from the databases detailed above. We seeded the algorithm by identifying the single interaction whose correlation was most different between the two phenotypes. On identifying this single interaction, we searched the full PPI network using a greedy search, and built a subnetwork of interacting genes. This subnetwork combines from the associations with the most significant difference between the two phenotypes. Unlike the pathway-based approach, this identified set of genes was not constricted to the pre-carved pathway consensus.

This way, we discovered previously untagged interactions that are the core mechanism of the disease. Finally, we compared the obtained results with the results obtained through the pathway-based approach. The most significant result is tagged as the most effective biomarker.

2.3 Algorithm–pseudo code

```
1. for each pathway
  1.1 get molecules list
  1.2 for every interaction
    1.2.1 if(moleculeA == 'input' && moleculeB == 'output')
      1.2.1.1 PhenotypeA = correlation (moleculeA, moleculeB);
      1.2.1.2. PhenotypeB = correlation (moleculeA, moleculeB);
    1.2.2 End
  1.3 End
  1.4 P-value = t-test(PhenotypeA, PhenotypeB);
  1.5 PhenotypeA = average(abs(PhenotypeA));
  1.6 PhenotypeB = average(abs(PhenotypeB));
2. End
3. Threshold = best pathway;
4. for every interaction in PPI network
  4.1  PPI_Net (i,j) = Difference(gene-gene correlation in phenotype A,
                       gene-gene correlation in phenotype B);
  4.2 Max = highest difference
  4.3 while (Threshold < Max)
    4.3.1 Greedy walk on the network
  4.4 End
5. End
6. Return Best
```

The actual code is written in Matlab [23] and is available upon request.

2.4 Sample data for case study

Sample data were obtained from The Cancer Genome Atlas (TCGA) database, available at http://cancergenome.nih.gov/. The dataset combined from RNA-sequencing data and clinical characteristics of 244 breast cancer patients. RNA-sequencing data were aligned to the human reference genome build hg19 using TopHat (Trapnell *et al.*, 2009) and analyzed using Cufflinks (Trapnell *et al.*, 2010) to produce FPKM values. The TCGA dataset was analyzed by PhenoNet, thus identifying the network that significantly stratifies triple-negative (TN) patients from non-triple-negative (non-TN) patients.

2.5 Sample data for validation analysis

As TCGA is the only dataset that contains a large-enough collection of RNA-sequencing with its paired clinical data of breast cancer patients, we used, as validation to the robustness of the method, two additional microarray breast cancer datasets. Using PhenoNet, we stratified these data according to their luminal A and luminal B subtypes (the datasets containing TN data were too small for this type of analysis).

The two additional datasets obtained from GEO are: GSE20713 (Dedeurwaerder *et al.*, 2011) and GSE21653 (Sabatier *et al.*, 2011a, b). The datasets are composed of gene expression and clinical information from breast cancer patients. Luminal A and luminal B subtypes were extracted and used in this analysis owing to the shortage of TN information.

2.6 False discovery rate analysis

The analysis of large datasets has become mainstream. It is often the case that thousands of features in a genome-wide dataset are tested against the null hypothesis, and only a fraction of those features are expected to be significant. An overwhelming number of measurements require corrections for multiple hypotheses. We used Benjamini and Hochberg's false discovery rate (FDR) (Benjamini and Hochberg, 1995). We performed FDR analysis to determine the FDR for the network identified by PhenoNet and for individual edges in the network. The index we propose here quantifies the strength of the subnetwork found based on the differences between the two phenotypes. We term this index the edge correlation distance (ECD). ECD is measured as detailed in Equation (1):

$$\rho_{iA} = pearson\_correlation\_for\_phenotypeA$$

$$\rho_{iB} = pearson\_correlation\_for\_phenotypeB$$

$$\Delta\rho_i = \rho_{iA} - \rho_{iB}$$

$$(o - \Delta\rho_i)^2 = Squared\_distance\_from\_zero$$

$$ECD = \sqrt{\frac{\sum_{i=1}^{n} (o - \Delta\rho_i)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} \Delta\rho_i^{\,2}}{n}}$$

$$n = number\_of\_edges$$

Δρ_i indicates the delta over Pearson correlation coefficients between the two edges. This index forms a new distribution based on the differences between the two phenotypes, and calculates the distribution distance from zero that, in this case, represents a lack of difference in correlation. N represents the number of edges in the subnetwork. Overall, ECD metrics calculates, for every two edges, the squared distance of those edges from zero. ECD can range from zero to two (zero indicates no difference and two represents the biggest change possible, the difference between correlation of 1 in one group and a correlation of -1 in the other). By randomly constructing a large number of networks (we chose 10 000 in this simulation), we are able to calculate the ECD index for every random network, and to estimate the probability of obtaining the observed ECD in the networks found by PhenoNet, as calculated for the two datasets, by chance. The ECD for the first dataset (GSE20713) was found to be 0.96, with a *P*-value of 0, and the ECD for the second dataset (GSE21653) was 0.89, with a *P*-value of 3.21e-08.

The described FDR algorithm iteratively randomized phenotypic affiliation within patients 10 000 times, thus mimicking the results of a random, non-disease state. The results revealed that all edges were significant, as can be seen in Figure 2, meaning that the network found by PhenoNet is statistically highly significant and can only be accounted for through a disease-introduced mechanism.

3 RESULTS

3.1 Case study: PhenoNet algorithm identifies core mechanisms in TN breast cancer samples

TN breast cancer is of great interest in oncology research. TN cancers do not respond to hormonal therapies or to treatments targeted against human epidermal growth factor receptor 2 receptors. The only systemic therapy currently available is chemotherapy, and prognosis remains poor (Stockmans *et al.*, 2008). TNs are tumors without the observable expression of estrogen receptor (ER), progesterone receptor and HER2neu.

Here, using RNA-seq data from 244 breast cancer patients from the TCGA, we performed phenotype classification using PhenoNet algorithm. The two phenotypes chosen were a TN

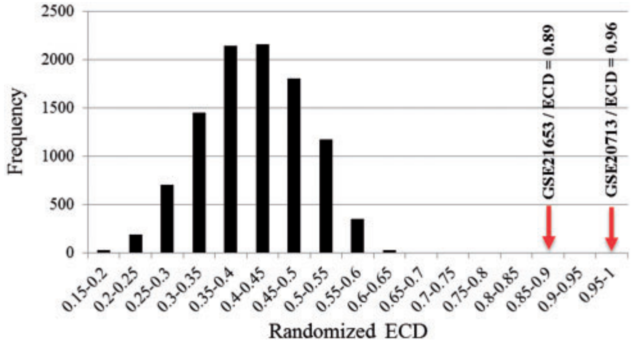

Fig. 2. False discovery analysis. The graph represents the ECD values calculated from 10 000 randomized iterations. By randomizing the clinical data, we were able to mimic a non-disease-related state. The results presented here show the randomized distribution, as opposed to the real disease state, in the two breast cancer datasets tested, which implies that the sub-networks found by PhenoNet are disease specific

group that included 52 patients, and a non-TN group that included 192 patients.

RNA-seq FPKM levels (Section 2), along with phenotypic groups, were provided as input. PhenoNet then iterated across all 464 pathways in the database, to identify the most variable pathway between the TN and the non-TN groups. Once pathway metrics for the entire collection of pathways were established, we chose the single most significant pathway according to its calculated *P*-value and its phenotype delta (see Section 2).

As previously discussed, the PPI network embedded within PhenoNet is composed of 10 489 genes and 62 643 interactions. As the lower threshold is set using the pathway approach described above, PhenoNet, using a greedy search methodology, iterates across the complete interaction matrix to identify interacting genes that create an optimal subnetwork (with the prerequisite of outperforming the previously selected single pathway). Once all iterations are exhausted, PhenoNet returns the best solution for the phenotype stratification. Figure 3 shows PhenoNet results for the network modules that stratify TN and non-TN groups in the analyzed breast cancer dataset. PhenoNet identified the network shown in Figure 3 as a highly modified version of the same mechanism in the two phenotype groups, with a highly significant *P*-value of 1.2e-009. Figure 3 shows us that the TN group is highly correlated; the colors of edges represent the correlation coefficients [low (red) to high (green) over a gradient], which indicate hypothesized regulation strength between the interacting genes. In contrast, the non-TN group produces low, practically zero, correlation values, indicating a possible loss of regulation within the participating genes.

The results presented here suggest this subnetwork as a candidate molecular mechanism involved in the control of TN breast cancer. When such control fails (and the correlation between the genes disappears), we see the network that characterizes the non-TN cases, which are more treatable and associated with better prognosis. Network-aware intervention may, therefore, seek to adjust the TN network to its non-TN behavior and, consequently, to control clinical presentation.

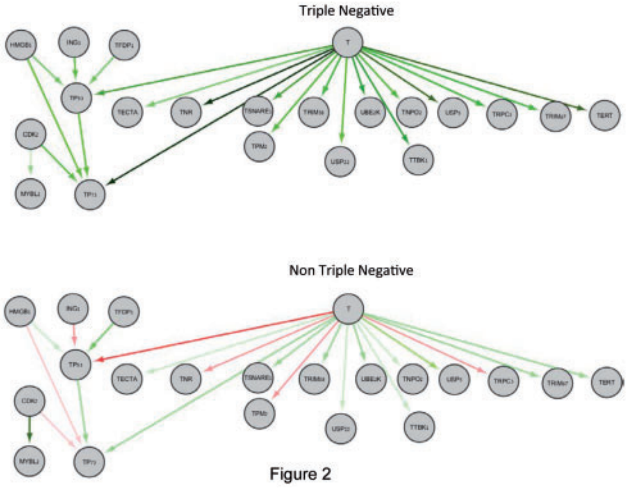

Fig. 3. Triple-negative versus non-triple-negative network. Every node represents a single protein and every edge represents a known physical interaction between the two nodes. A red-colored edge indicates a negative correlation and a green-colored edge represents a positive correlation. The strength of the edge's color represents the strength of the calculated correlation

3.2 Validation

Over the past few years, robustness proved to be highly important feature in molecular signatures (Ben-Hamo and Efroni, 2012a,b; Ben-Hamo and Efroni, 2013; Ein-Dor *et al.*, 2006; Venet *et al.*, 2011). Identified signatures are strengthened through multiple validations in additional, independent datasets. Ideally, to validate the algorithm, we would require re-application of the same approach of analysis on additional RNA-seq datasets produced from tumor samples taken from patients identified as having TN or non-TN breast cancer. Unfortunately, TCGA is the only available dataset with information of a significant number of patients. To overcome this limitation, we validated PhenoNet using two independent breast cancer microarray datasets with enough clinical information about patients' luminal A and luminal B status. As PhenoNet is impartial to the specific method used for transcriptome quantification (microarray, RNA-seq), the analysis of these two unrelated breast cancer datasets, GSE20713 (Dedeurwaerder *et al.*, 2011) and GSE21653 (Sabatier *et al.*, 2011b), that combine gene expression and clinical data, will identify subnetworks that significantly stratify the subtypes. Figure 4 shows PhenoNet results in the two datasets tested based on the luminal A and B classification.

Although the identified networks are not identical, they share key control genes and four overlapping edges. These overlapping edges provide robustness and highlight a possible role for this subnetwork in its involvement with the luminal A/B phenotypes. Figure 5 highlights differences in these phenotypes.

The networks identified using PhenoNet call special attention to YWHAG as a key regulator of the division between subtypes. YWHAG (also known as 14-3-3 γ) is part of the 14-3-3 protein family that participates in phosphorylation-dependent PPIs that control progression through the cell cycle, initiation and maintenance of DNA-damage checkpoints, activation of MAP kinases, prevention of apoptosis and coordination of integrin

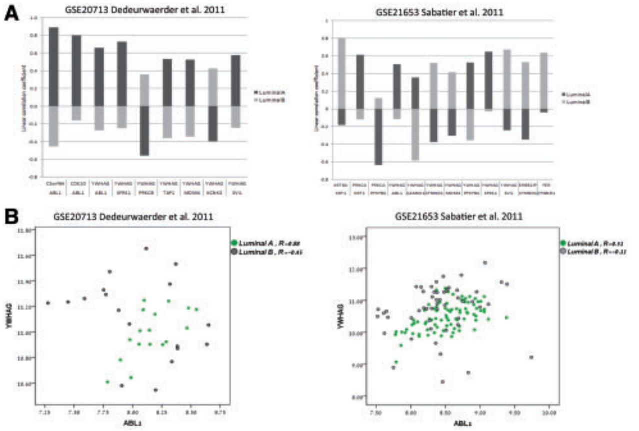

Fig. 4. Networks representing Breast cancer luminal A versus luminal B in two independent datasets. Each node represents a single protein and an edge represents a known physical interaction between the two nodes. Red edges indicate a negative correlation and green edges indicate a positive correlation. The edge color gradient represents the strength of the measured correlation

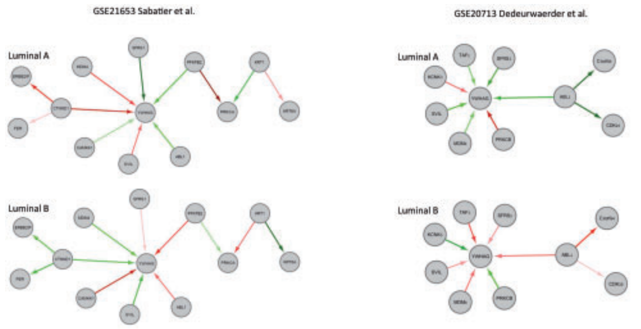

Fig. 5. Network behavior in breast cancer luminal A versus breast cancer luminal B. (**A**) This figure shows how values of the linear correlation coefficients in the two networks formed by PhenoNet are opposite. (**B**) A scatterplot showing the correlation between ABL1 and YWHAG in luminal A patients and the lack of correlation in luminal B patients in both datasets

signaling and cytoskeletal dynamics (Wilker and Yaffe, 2004). This family plays an important role in cancer. Song *et al.* (2012) recently found that the expression levels of the YWHAG protein in breast cancer were significantly higher than in non-cancerous mammary-gland tissues. Furthermore, they found a correlation between the expression levels of this protein and the clinicopathological features and prognosis of breast cancer patients.

Although this single gene, YWHAG, has been brought forward through this and previous research, it is important to emphasize that none of the genes in these subnetworks, by themselves, shows any stratification strength in separating the two subtypes, as can be seen in Supplementary Table S1; i.e. the gene components of the network, when taken separately and out of the network context of the other genes in the pathway, failed to provide biomedical meaning (Supplementary Table S1).

4 DISCUSSION

Cancer is not a static disease; disease progression correlates with molecular modifications to evade natural defenses and to adapt to new environmental and micro-environmental circumstances. Such drastic changes that stands at the core of the cancer phenotype are not isolated incidents but are systemic modifications; indeed, the initiation and progression of the disease are driven by a large set of factors such as: interactions among genes, proteins and the environment, rather than by single alterations in genetic variants (Correia and Bissell, 2012; Ziogas *et al.*, 2011). Systemic approaches enable the integration of biological and computational knowledge, and help address fundamental questions regarding the complexity and connectivity of this system-wide rewiring. Different molecular phenotypes are the manifestations of different molecular characteristics within the same, single disease. Such complex profiles of molecular phenotypes are determined using molecular and genetic information from tumor cells. Oncogenesis may be better considered not only as the rearrangement of chromosomes but also as the rearrangement of regulatory networks, their interactions and their interconnections. Cancer is not a disease of a few genes but of a cluster of genes, whose altered global interactions evolve into the transformed and malignant state. These considerations, as well as their multilayered association with network features, call for a network behavior tool with the ability to identify core network alterations that are at the root of phenotypic changes.

Here, we describe PhenoNet, a novel algorithm for the identification of subnetworks of genes or known pathways whose molecular co-behavior stratifies phenotypes. Correlations between the expression levels of interacting genes indicate regulatory control between network nodes. We search for subnetworks whose synchronization is visible in one phenotype and contrasted by another. Thus, when we identify a highly correlated subnetwork in one group of patients and then identify the loss of this correlation in another, the assumption is that this regulation is fragmented during the course of progression into or out of the phenotype. PhenoNet is able to identify such networks. Gelenbe (2007) wrote that ‘a regulatory network acts as the control system of a biochemical nanofactory with the rate of production of certain compounds being determined by the probability that certain sets of agents are activated’. The approach presented here uses the known regulatory PPI network to extract a subnetwork that acts as a control system distinguishing between the two phenotypes.

To determine PhenoNet robustness we compared it with a set of highly informative and well-structured tools: SurvNet (Li *et al.*, 2012), Biolayout Express (Theocharidis *et al.*, 2009) and C3NET (Altay and Emmert-Streib, 2011). For this analysis, we used the two breast cancer datasets that were tested (see above) on PhenoNet. PhenoNet results demonstrated a 50% overlap between the networks discovered for the two breast cancer datasets. This robustness was not repeated in the three tested tools. As informative and useful as these tools are, and without specific criticism of the important results that can be obtained using them, they did not repeat the results obtained from the two datasets (Supplementary Figs S1 and S2; Supplementary Tables S2–S5). This highlights a unique and important feature of PhenoNet, i.e. robustness, which is a necessity in transcriptomics-based biomarker discovery.

We first applied PhenoNet to stratify network rewiring in breast cancer clinical subgroups: TN (ER-, PgR-, Her2-) and the non-TN. Using breast cancer RNA-seq data from the TCGA, the network identified by PhenoNet as the most significant in distinguishing between those two groups, was dominated by T-gene (also known as Brachyury), as can be seen in Figure 3. The T-box family of genes is a group of highly conserved transcription factors (TF) that play an important role in embryonic development (Fan *et al.*, 2004). It has been found that the T-box TF Brachyury induces in tumor cells epithelial-mesenchymal markers, down regulation of epithelial markers and an increase in cell migration and invasion. Inhibition of Brachyury resulted in loss of cell migration and invasion, and reduced the ability of human tumor cells to form lung metastasis (Fernando *et al.*, 2010; Roselli *et al.*, 2012). These findings may support some of the results shown in this article. Although the subnetwork identified in the TN group is highly correlated, this same subnetwork is uncorrelated in the non-TN group. This network may be the cause of the high level of aggressiveness of the TN tumors. As long as this network is controlled by T-gene migration, invasion and metastasis are promoted. Once this control mechanism is lost (and with it the subnetwork correlation), the breast cancer phenotype becomes less aggressive and is manifested as non-TN.

The use of two additional breast cancer datasets to validate our approach using samples from tumors with receptor status [ER+, PgR+, and Her2- (classified as luminal A), and ER+, PgR+, and Her2+ (classified as luminal B)], demonstrated how PhenoNet directs the identification of a key player in the stratification of the subnetworks in both datasets, namely YWHAG. YWHAG belongs to the 14-3-3 family of proteins that regulate many cellular processes that are important in cancer biology, such as apoptosis and cell cycle check points (Hermeking, 2003). A recent study showed an association between single nucleotide polymorphisms in YWHAG and breast cancer subtype ER+ and Her2- (P-value$<$0.05) (Olson *et al.*, 2011). In addition, YWHAG was also found to be correlated with the expression pattern of ER protein in breast cancer in a way that correlated with patient outcome (Jenssen *et al.*, 2002). Taken together, this may suggest an important role played by YWHAG in controlling the networks and in stratifying the two subtypes.

The article presented here provides means to identify core processes that are tightly linked with different phenotypic manifestations of disease. We believe that network target identification is the key for constructing drug-target networks and for improving the understanding of disease etiology.

ACKNOWLEDGEMENTS

The results published here are fully or partially based on data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions constituting the TCGA research network can be found at the project website (http://cancergenome.nih.gov/).

Funding: Part of this work has been funded by the Israeli Cancer Association (grant no. 20132014).

Conflict of Interest: none declared.

REFERENCES

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

Antonov,A.V. *et al.* (2009) PPI spider: a tool for the interpretation of proteomics data in the context of protein-protein interaction networks. *Proteomics*, **9**, 2740–2749.

Ben-Hamo,R. and Efroni,S. (2012a) Biomarker robustness reveals the PDGF network as driving disease outcome in ovarian cancer patients in multiple studies. *BMC Syst. Biol.*, **6**, 3.

Ben-Hamo,R. and Efroni,S. (2012b) Correction: gene expression and network-based analysis reveals a novel role for hsa-miR-9 and drug control over the p38 network in glioblastoma multiforme progression. *Genome Med.*, **4**, 87.

Ben-Hamo,R. and Efroni,S. (2013) Network as biomarker: quantifying transcriptional co-expression to stratify cancer clinical phenotypes. *Syst. Biomed.*, **1**, 35–41.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.*, **57**, 289–300.

Correia,A.L. and Bissell,M.J. (2012) The tumor microenvironment is a dominant force in multidrug resistance. *Drug Resist. Updat.*, **15**, 39–49.

Dedeurwaerder,S. *et al.* (2011) DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol. Med.*, **3**, 726–741.

Efroni,S. *et al.* (2011) Detecting cancer gene networks characterized by recurrent genomic alterations in a population. *PLoS One*, **6**, e14437.

Ein-Dor,L. *et al.* (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl Acad. Sci. USA*, **103**, 5923–5928.

Emmert-Streib,F. *et al.* (2012) Harnessing the complexity of gene expression data from cancer: from single gene to structural pathway methods. *Biol. Direct*, **7**, 44.

Fan,W.W. *et al.* (2004) TBX3 and its isoform TBX3+2a are functionally distinctive in inhibition of senescence and are overexpressed in a subset of breast cancer cell lines. *Cancer Res.*, **64**, 5132–5139.

Fernando,R.I. *et al.* (2010) The T-box transcription factor Brachyury promotes epithelial-mesenchymal transition in human tumor cells. *J. Clin. Invest.*, **120**, 533–544.

Fröhlich,H. (2011) Network based consensus gene signatures for biomarker discovery in breast cancer. *PLoS One*, **6**, e25364.

Gelenbe,E. (2007) Steady-state solution of probabilistic gene regulatory networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **76** (Pt. 1), 031903.

Greenblum,S.I. *et al.* (2011) The PathOlogist: an automated tool for pathway-centric analysis. *BMC Bioinformatics*, **12**, 133.

Hermeking,H. (2003) The 14-3-3 cancer connection. *Nat. Rev. Cancer*, **3**, 931–943.

Jenssen,T.K. *et al.* (2002) Associations between gene expressions in breast cancer and patient survival. *Hum Genet*, **111**, 411–420.

Joshi-Tope,G. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.

Khatri,P. *et al.* (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.*, **8**, e1002375.

Kim,H. and Gelenbe,E. (2012) Stochastic gene expression modeling with Hill function for switch-like gene responses. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **9**, 973–979.

Lee,E. *et al.* (2008) Inferring pathway activity toward precise disease classification. *PLoS Comput. Biol.*, **4**, e1000217.

Li,J. *et al.* (2012) SurvNet: a web server for identifying network-based biomarkers that most correlate with patient survival data. *Nucleic Acids Res.*, **40**, W123–W126.

Ogata,H. *et al.* (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **27**, 29–34.

Olson,J.E. *et al.* (2011) Centrosome-related genes, genetic variation, and risk of breast cancer. *Breast Cancer Res. Treat.*, **125**, 221–228.

Peri,S. *et al.* (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.*, **13**, 2363–2371.

Pierobon,M. *et al.* (2009) Multiplexed cell signaling analysis of metastatic and nonmetastatic colorectal cancer reveals COX2-EGFR signaling activation as a potential prognostic pathway biomarker. *Clin. Colorectal Cancer*, **8**, 110–117.

Roselli,M. *et al.* (2012) Brachyury, a driver of the epithelial-mesenchymal transition, is overexpressed in human lung tumors: an opportunity for novel interventions against lung cancer. *Clin. Cancer Res.*, **18**, 3868–3879.

Sabatier,R. *et al.* (2011a) Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One*, **6**, e27656.

Sabatier,R. *et al.* (2011b) A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res. Treat.*, **126**, 407–420.

Schaefer,C.F. *et al.* (2009) PID: the pathway interaction database. *Nucleic Acids Res.*, **37**, D674–D679.

Song,Y. *et al.* (2012) Expression of 14-3-3γ in patients with breast cancer: correlation with clinicopathological features and prognosis. *Cancer Epidemiol.*, **36**, 533–536.

Stark,C. *et al.* (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.

Stockmans,G. *et al.* (2008) Triple-negative breast cancer. *Curr. Opin. Oncol.*, **20**, 614–620.

Su,J. *et al.* (2009) Accurate and reliable cancer classification based on probabilistic inference of pathway activity. *PLoS One*, **4**, e8161.

Teschendorff,A.E. *et al.* (2010) Improved prognostic classification of breast cancer defined by antagonistic activation patterns of immune response pathway modules. *BMC Cancer*, **10**, 604.

Theocharidis,A. *et al.* (2009) Network visualization and analysis of gene expression data using BioLayout Express (3D). *Nat. Protoc.*, **4**, 1535–1550.

Trapnell,C. *et al.* (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.

Trapnell,C. *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.

Vaske,C.J. *et al.* (2010) Inference of patient-specific pathway activities from multidimensional cancer genomics data using PARADIGM. *Bioinformatics*, **26**, i237–i245.

Venet,D. *et al.* (2011) Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput. Biol.*, **7**, e1002240.

Wang,Y.C. and Chen,B.S. (2011) A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med. Genomics*, **4**, 2.

Wilker,E. and Yaffe,M.B. (2004) 14-3-3 proteins - a focus on cancer and human disease. *J. Mol. Cell. Cardiol.*, **37**, 633–642.

Ziogas,D.E. *et al.* (2011) From traditional molecular biology to network oncology. *Future Oncol.*, **7**, 155–159.