

SCIENTIFIC REPORTS



OPEN

Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer

Esra Gov & Kazim Yalcin Arga

Ovarian cancer is one of the most significant disease among gynecological disorders that women suffered from over the centuries. However, disease-specific and effective biomarkers were still not available, since studies have focused on individual genes associated with ovarian cancer, ignoring the interactions and associations among the gene products. Here, ovarian cancer differential co-expression networks were reconstructed via meta-analysis of gene expression data and co-expressed gene modules were identified in epithelial cells from ovarian tumor and healthy ovarian surface epithelial samples to propose ovarian cancer associated genes and their interactions. We propose a novel, highly interconnected, differentially co-expressed, and co-regulated gene module in ovarian cancer consisting of 84 prognostic genes. Furthermore, the specificity of the module to ovarian cancer was shown through analyses of datasets in nine other cancers. These observations underscore the importance of transcriptome based systems biomarkers research in deciphering the elusive pathophysiology of ovarian cancer, and here, we present reciprocal interplay between candidate ovarian cancer genes and their transcriptional regulatory dynamics. The corresponding gene module might provide new insights on ovarian cancer prognosis and treatment strategies that continue to place a significant burden on global health.

Ovarian cancer remains the leading cause of death from gynecologic malignancy with an estimated death ratio of 5% in whole cancer types¹. Five -year survival rate of women with ovarian cancer was reported as 35% in US, and the necessity in development of effective treatment strategies was emphasized since treatment of ovarian cancer with current therapies is harder than any other type of female reproductive tract cancers².

The most common type of primary malignant ovarian tumors is the epithelial carcinoma that begins in the ovary tissues³. This carcinoma is the most dangerous of all types of ovarian cancers. Unfortunately, it is not diagnosed until the disease is advanced in stage⁴.

Over the last decade, enormous researches have been made to understand mechanisms of ovarian cancer pathogenesis and to identify diagnostic and prognostic targets. However, disease-specific and effective biomarkers were still not available, since studies have focused on individual genes associated with ovarian cancer, ignoring the interactions and associations among the gene products. On the other hand, integration of biological data at any level (gene, transcript, protein, metabolite, etc.) with biomolecular networks (i.e., gene co-expression⁵, protein-protein interaction⁶, transcriptional regulatory⁷, and metabolic networks⁸) provides valuable insights on elucidation of the disease mechanisms^{9,10} and identification of molecular signatures of human diseases¹¹⁻¹⁴.

Co-expression networks are reconstructed from gene expression data using pairwise correlation metrics⁵. Altered co-expression patterns of genes between two states (for instance, healthy vs. tumor) are called differential co-expression¹⁵, which represent significant potential to identify gene clusters affected by state transition. Construction of differential co-expression networks and their topological analysis provide us valuable information on the alterations in biological systems in response to environmental and biological perturbations, such as disease formation and gene mutation^{16,17}. In several studies, differential co-expression networks were studied to identify disease associated genes and gene modules in human diseases including chronic lymphocytic leukemia¹⁸, obesity¹⁹, tumor-associated macrophages²⁰, breast cancer²¹, and ovarian cancer^{22,23}.

Zhou and coworkers²² built an integrated co-alteration network via utilization of copy number, methylation and mRNA expression data, and identified 155 ovarian cancer associated genes. In another study, through weighted correlation network analysis, 3095 differentially expressed genes were identified from genome

Department of Bioengineering, Marmara University, 34722, Goztepe, Istanbul, Turkey. Correspondence and requests for materials should be addressed to K.Y.A. (email: kazim.arga@marmara.edu.tr)

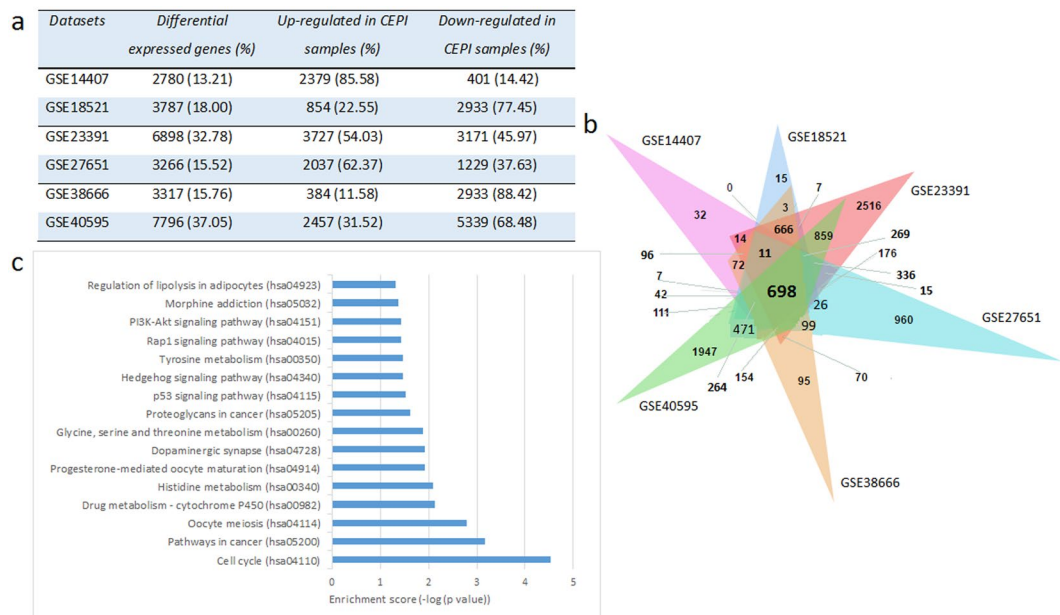


Figure 1. Differentially expressed genes (DEGs) in ovarian cancer datasets. **(a)** The table represents the numbers (and percentages in parenthesis) of differentially expressed genes between laser micro-dissected epithelial cells from ovarian tumor (CEPI) and ovarian surface epithelia (OSE) samples. The direction of regulation (i.e., up or down-regulation) of the genes was also specified. **(b)** Venn diagram of DEGs across the six ovarian cancer associated datasets. **(c)** Statistically significant KEGG pathways obtained through the pathway enrichment analysis of the 698 mutual DEGs of all datasets (adjusted p-value < 0.05).

expression profiles of ovarian cancer patients in early and advanced cancer stages, and 6 prognosis related genes were selected out as novel candidates for clinical biomarkers²³.

In the present study, considering the importance of the choice of tissue in gene expression studies²⁴, we performed meta-analysis of transcriptome datasets including samples from laser micro-dissected epithelial cells from ovarian tumor and healthy ovarian tissues, differential co-expression networks were reconstructed at two different (healthy and diseased) states, and the modules (clusters of highly connected network components) of co-expression networks were comparatively analyzed. We identified a novel prognostic gene module, which was differentially co-expressed in ovarian cancer when compared to healthy tissues. Topological and functional enrichment analyses were performed to understand the molecular mechanisms. Prognostic transcriptional regulatory elements (i.e., transcription factors and miRNAs) of the module were also investigated.

Results

Differential gene expression in ovarian cancer. In the present study, we analyzed transcriptome datasets from six independent studies associated with ovarian serous carcinoma by comparing gene expression levels between laser-micro-dissected epithelial cells from ovarian tumor (CEPI) and healthy ovarian surface epithelial (OSE) samples. To identify differentially expressed genes (DEGs) in each dataset, statistical analyses were performed, which reported statistically significant (adjusted p-value < 0.01 and fold-change > 2.0 or < 0.5) alterations in 13–37% of the genome (22% in average) between healthy and diseased states (Fig. 1a). Although the numbers of DEGs and their expression patterns (i.e., down or up-regulation) among datasets were incompatible, there were still 698 mutual DEGs (Fig. 1b). Functional enrichment analysis of proteins encoded by these genes indicated several biological pathways mainly associated with cell cycle, cancer signaling, drug and amino acid metabolisms (Fig. 1c).

Co-expression profiles in ovarian cancer. Possible correlations between expression profiles of 698 mutual DEGs were identified employing Pearson correlation coefficients (PCCs). For this purpose, the expression values of mutual DEGs in diseased (CEPI) and healthy (OSE) states were employed separately. The calculated PCCs were normally distributed with a mean and standard deviation of 0.05 and 0.39, respectively (data not shown). PCC cut-offs of > 0.815 (positive correlation) and < -0.815 (negative correlation), which correspond to p-value < 0.05, were employed to determine statistical significance of the pairwise correlations. Two co-expression networks (CNC and CNO) were constructed within the significantly co-expressed DEGs. CNC, which reveals the co-expression network in CEPI samples, consisted of 5444 links between 237 genes (Fig. 2a); whereas CNO, representing the co-expressions in OSE samples, consisted of 4947 links among 455 genes (Fig. 2b). When degree distributions of networks were investigated, both networks possessed scale-free topology ($R^2 > 0.80$), indicating the presence of hubs. However, CNC and CNO did not share mutual hubs. Though the number of genes exhibiting co-expression pattern was almost two-fold higher in healthy state, the density (0.194) and the clustering coefficient (0.72) were higher in CNC when compared to those in CNO (0.048 and 0.48, respectively).

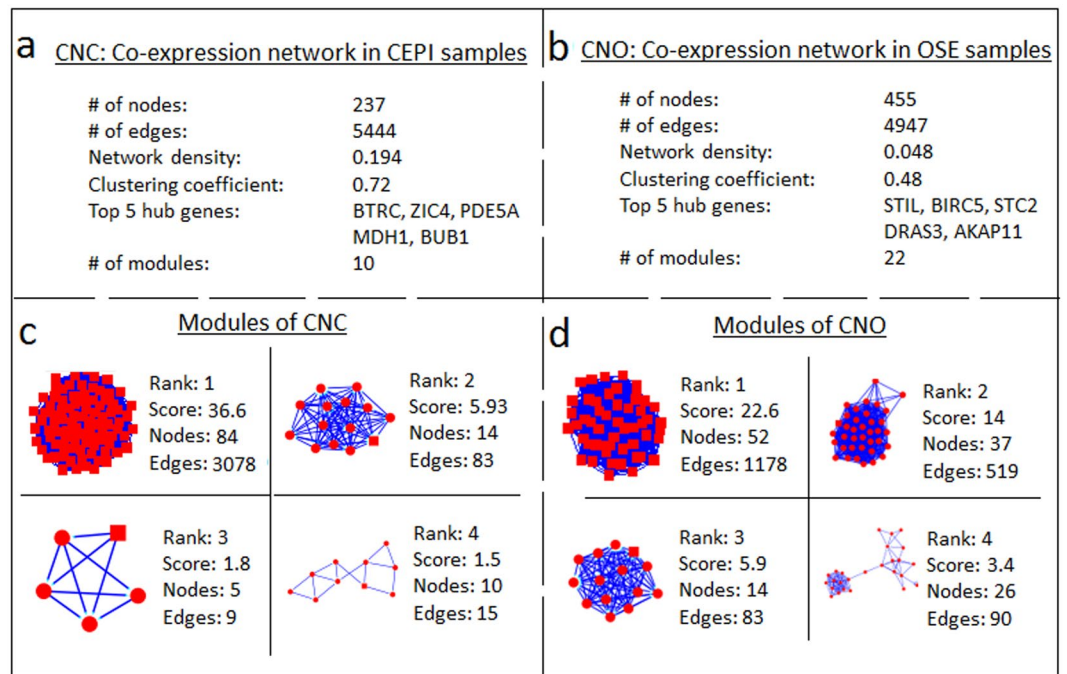


Figure 2. Characteristics of co-expression networks and belonging modules in laser micro-dissected epithelial cells from ovarian tumor (CEPI) and ovarian surface epithelia (OSE) samples. **(a)** Topological properties of the co-expression network in CEPI samples (CNC). **(b)** Topological properties of the co-expression network in OSE samples (CNO). **(c)** Top four co-expressed modules of CNC representing the diseased state. **(d)** Top four co-expressed modules of CNO representing the healthy state. DEGs were represented as nodes, and the statistically significant co-expression associations between DEGs were represented as edges.

Co-expressed gene modules in diseased and healthy states. To identify network modules, the topological structure of the co-expression networks (CNC and CNO) were analyzed. We identified 10 and 22 modules in CNC and CNO, respectively (Fig. 2c,d). Since we are looking for a prognostic gene module that may represent signatures in ovarian cancer pathogenesis, we focused on the modules in CNC. In addition, we expect the candidate prognostic modules to contain considerable amount of genes and represent high intensity of the connectedness within the module in order to maintain high precision in their predictive capability. Considering the topological parameters (the number of nodes, the module density and the average connectivity) of the modules in CNC (Supplementary Table S1), we selected the module with 84 genes, which also represented high average connectivity (with 36.6) and density (with 0.88), and considered in further analyses.

On the other hand, top scoring module of CNO (with 52 genes), which might give clues on the dysregulated pathways in healthy tissues, was also analyzed. However, functional enrichment analyses of this module presented expected results; the majority of the elements of the module was significantly enriched with the cell cycle pathway and cellular organization related processes (data not shown).

The module was differentially co-expressed in ovarian cancer. The selected module of CNC, so-called CMC, consisted of 3078 links among 84 genes (Fig. 3a). We also searched the CMC module within the CNO network to analyze the co-expression pattern of the module genes in the healthy state. In this case, only 100 links were observed among 48 of the 84 nodes (Fig. 3b). Furthermore, the comparative analysis of the CMC in both states indicated the significant alterations in co-expression patterns of module genes between two states. The network density in CMC was 0.883 in diseased state, whereas significantly lower (0.029) in the healthy state. These results pointed out the differential co-expression of the module in ovarian cancer. It may be suggested that the co-expression pattern among 84 genes (Supplementary Table S2) was activated in response to the change of state, and this set of genes may be considered as a “systems-biomarker” for therapeutics and prognosis in ovarian cancer.

We performed pathway enrichment analysis on the module genes using information from several data sources including KEGG and Reactome databases. However, statistically significant enrichment results couldn't be obtained; interestingly, the majority of these genes (90.5%) was taking roles in distinct pathways and/or not annotated in any pathway. Afterwards, to characterize the pathway or process annotation of each gene, we manually searched GeneCards Human Gene Database. A significant portion (35 genes, 42%) of the genes was not categorized in any biological pathway (Fig. 4a), whereas 22 genes (26%) were associated with GPCR signaling (Fig. 4b). When GPCR signaling pathways were investigated, 5 of 22 genes (BTRC, DYNC2H1, IGHD, NEDD4L and PIK3C3) were involved in immune systems response. In addition, 3 genes, (ADCYAP1, CDC14A, and LEPR) were taking roles in interleukin receptor SHC signaling pathway, and 6 genes (AMDHD1, CYP19A1, GADL1, GATM, HM13 and HSD17B2) were encoding proteins with metabolic activities in several biological processes (Fig. 4c). On the other hand, products of the remaining genes were distributed in several pathways or biological processes.

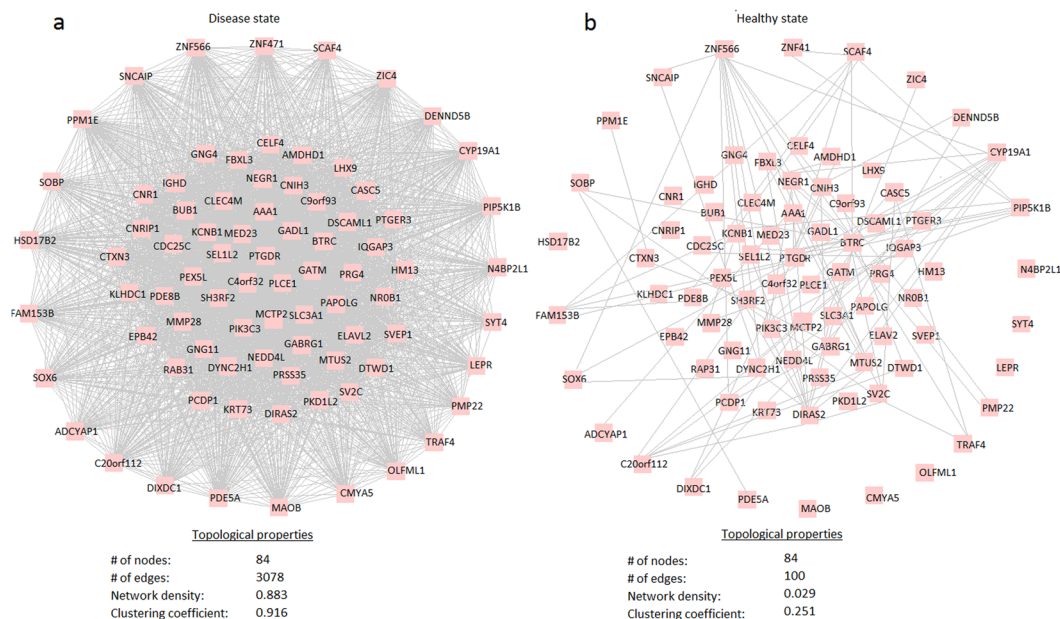


Figure 3. Differential co-expression module (CMC) of ovarian cancer consisting of 84 genes. (a) The topological features of the CMC representing the dense co-expression pattern in diseased samples. (b) The topological features of the CMC in healthy samples.

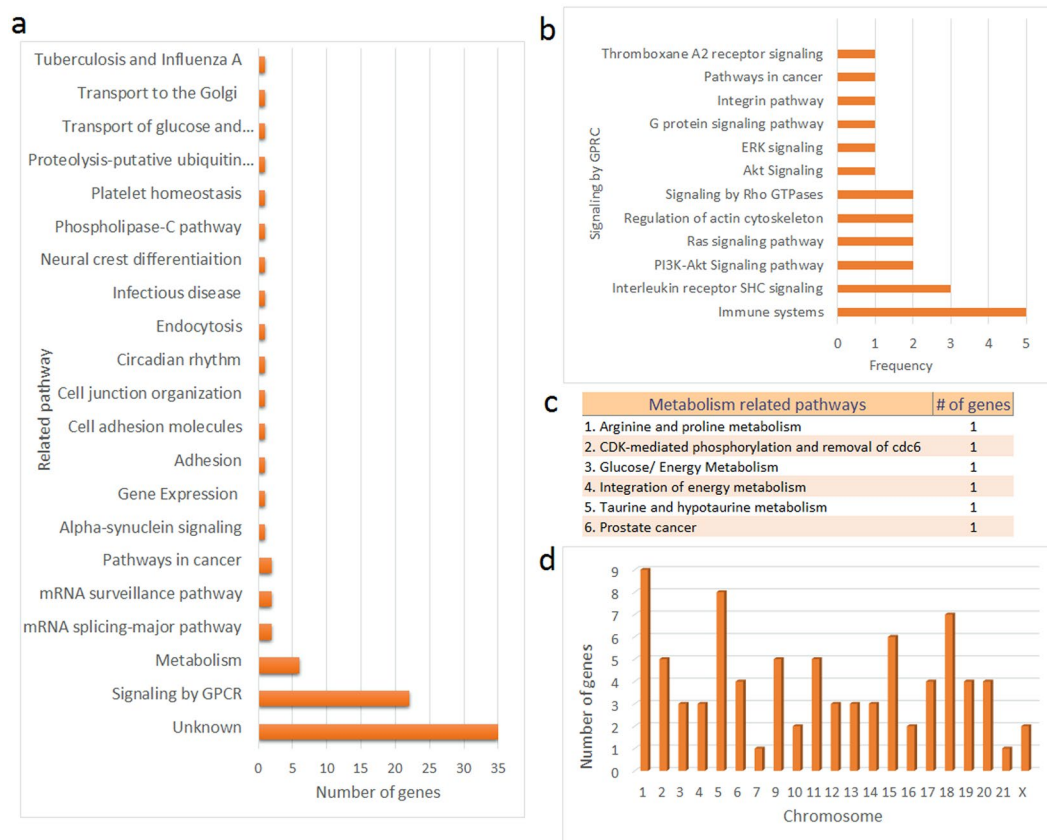


Figure 4. The biological processes, pathways and chromosomal locations associated with the prognostic genes. (a) The distribution of the prognostic genes into biological processes or molecular pathways. Process or pathway annotations of genes were obtained from GeneCards database. (b) The distribution of the prognostic genes into the “signaling by GPCR” subcategory. (c) The distribution of the prognostic genes into metabolism related pathways. (d) Distribution of the prognostic genes into chromosomal locations.

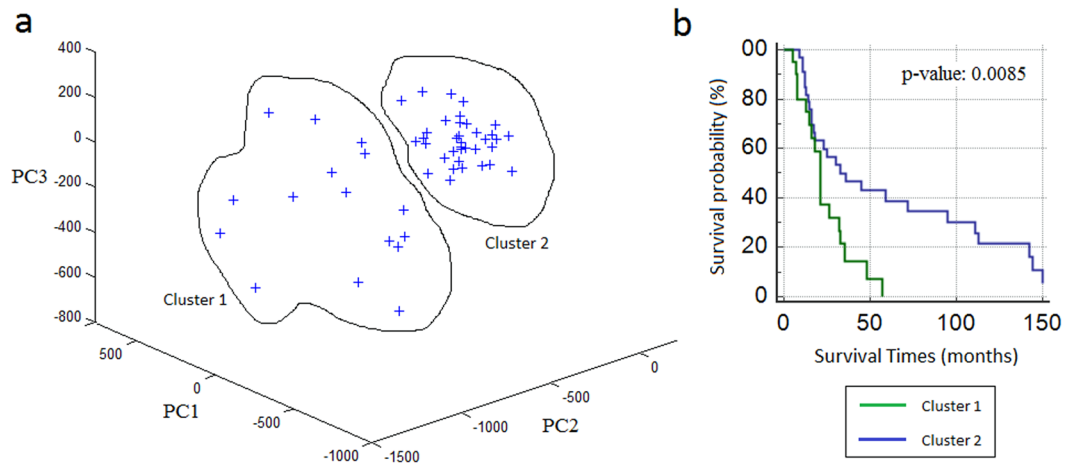


Figure 5. Prognostic performance of the module. (a) Clustering of samples using principle components (PC1, PC2, PC3) of the expression matrix of module genes. (b) Kaplan-Meier analysis of ovarian cancer dataset using the clusters identified through differentially co-expressed module. The p-values are computed using log-rank test.

The module genes were not concentrated in any pathway or biological process, surprisingly. Therefore, we analyzed the chromosomal location of the genes, hypothesizing that the dense co-expression pattern of this gene set may be as a result of the co-localization of the genes on the same chromosomes. However, the genes were distributed to almost all human chromosomes (except chromosome 8 and 22) at different ratios (Fig. 4d).

Prognostic performance of the gene module. Since the module was differentially co-expressed in ovarian cancer, and module genes were distributed in several pathways or biological processes and chromosomal locations, we hypothesized that this module may represent some common traits of ovarian cancer and can be potential marker for ovarian cancer prognosis. Therefore, we tested its prognostic capability in ovarian cancer using principle component analysis (PCA) and Kaplan-Meier survival curves. The first three principle components of the expression matrix of module genes, which describes 74.8% of the total variance, separated the samples into two clusters (Fig. 5a). Then, survival curve was plotted based on both clusters (Fig. 5b). The gene module was predictive of patient survival in ovarian cancer (p-value = 0.0085, log-rank test). Through Cox-proportional hazard analysis, and the hazard ratio was estimated as 2.36 with a 95% confidence interval of 1.21 to 4.62 (p-value = 0.0117).

Transcriptional regulators of the module genes. It must be revealed how accrued differential co-expression depends on conditions and which mechanisms cause these phenomena. Especially, transcriptional regulation of gene expression comes into prominence to evaluate the condition specific gene expression alterations. Transcription factors (TF) and microRNAs (miRNA) are the major regulators in transcriptional control. Therefore, relationships between TF-miRNA-module genes were investigated. It was found that six TFs (GATA2, YBX1, AR, ETS1, FOXP3, and PRDM14) were regulating the majority (89%) of module genes (Fig. 6a). In addition, 70 of 84 genes were regulated by more than one TF. Furthermore, 35 genes were co-regulated by GATA2, YBX1 and AR (Fig. 6b). When miRNA-target gene interactions were also evaluated; miR-335, miR-4284, miR-190a, miR-16, and miR-26b came into prominence with regulating at least seven target genes (Fig. 6c). Unlike TFs, miRNAs represented a diverse regulatory pattern, where module genes were regulated by distinct miRNAs (Fig. 6d).

Differential expression of the module genes in different tumor types. Genes of the differentially co-expressed CMC module were investigated across nine tumor tissues to determine specificity of the module to ovarian cancer, and to examine the expression pattern of the module in different tumor tissues. For this purpose, gene expression datasets associated with nine tumors (ovarian, breast, cervical, prostate, lung, colorectal, pancreatic, thyroid, and leukemia) were employed and statistically significant DEGs were identified. Module genes were screened in identified DEGs for each cancer type, and the expression pattern of the module genes in different tumor tissues were analyzed. The module genes were not detected in any of the tumor tissues as a whole (Fig. 7). However, 50% of the co-expressed module genes were determined as differentially expressed in serious ovarian tumor. Differential expressions of the genes were observed in leukemia, colorectal, thyroid, cervical and lung cancers at lower coverages (3–28%). Furthermore, none of the module genes were differentially expressed in breast, prostate and pancreatic tumor tissues. Among the module genes, BUB1, GATM, PLCE1, NEGR1, PTGER3 and SH3RF2 were differentially expressed in three or more cancer tissues (Table 1). These results indicated that the expression pattern of the module in different tumor tissues was distinct, and the module was specific to ovarian cancer.

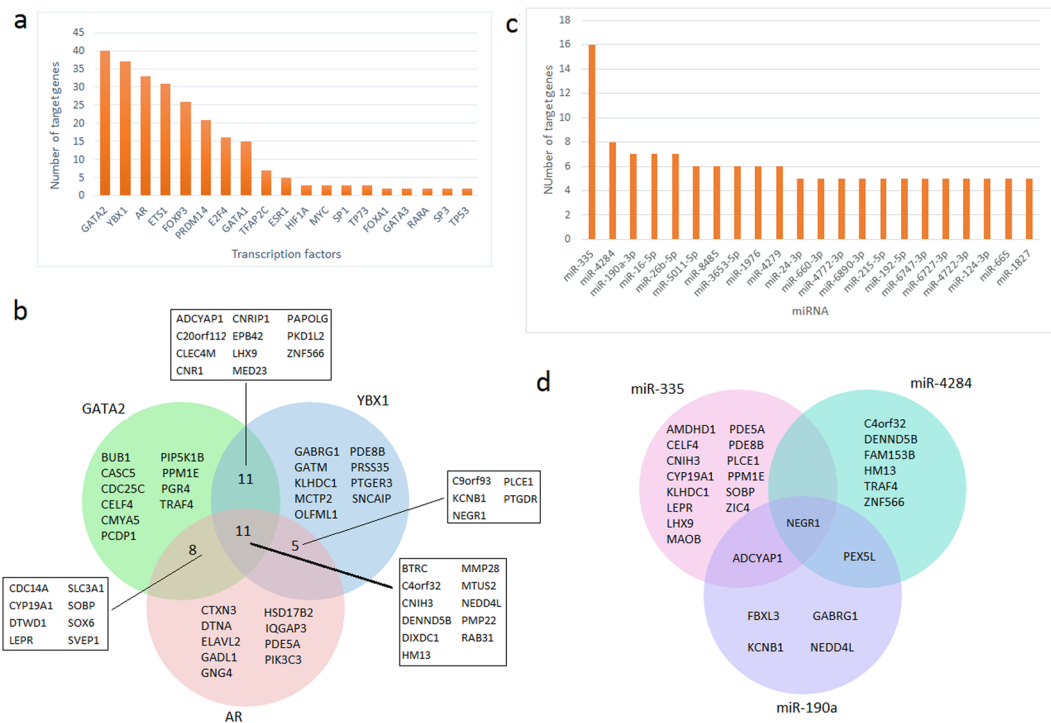


Figure 6. Transcriptional regulators of the prognostic genes. **(a)** Bar graph representing the distribution of transcription factors (TFs) which regulate the prognostic genes. **(b)** Venn diagram of the prognostic genes regulated by the most significant (top 3) TFs. **(c)** Bar graph representing the distribution of microRNAs (miRNAs) regulating the prognostic genes. **(d)** Venn diagram of the prognostic genes regulated by the most significant (top 3) miRNAs.

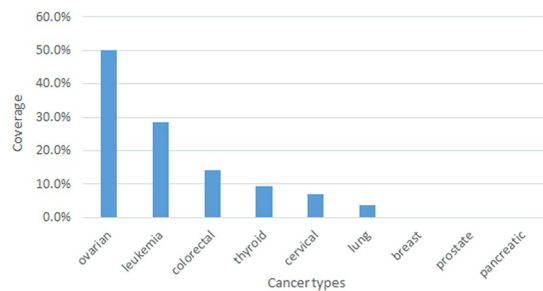


Figure 7. Differential expression of the prognostic genes in different tumor tissues. Bar graph represents the coverage of the module in each tumor sample, i.e. the ratio of the number of differentially expressed prognostic genes to the number of all genes in the prognostic module.

Discussion

Identification of alterations in co-expression patterns of genes among disease and healthy samples provides information about disease specific gene modules. Previous studies that aim to predict and prioritize candidate prognostic gene modules have focused on modules of co-expression and protein-protein interaction networks constructed around differentially expressed genes in ovarian cancer^{22,23}. However, the altered co-expression patterns of module genes during transition from healthy to diseased state were neglected in these studies. Here, we performed a differential co-expression analysis to identify disease genes and their co-expression couplings. As a result, the novel co-expressed gene module presented here might be considered as “systems biomarkers” that pioneer rational design of effective strategies in ovarian cancer diagnosis and prognosis within the perspective of precision medicine²⁵.

In the present study, in order to increase dimensionality and breadth of information that can be extracted, we didn't limit the analyses with a single transcriptome dataset, but performed a meta-analysis of various transcriptome datasets. In addition, a single type of microarray was chosen as the investigation platform (i.e. Affymetrix) and each dataset was analyzed independently to reduce the confounding factors in the data analysis, such as batch effects and variabilities originated from array platforms. It was reported that the genetic basis of carcinogenesis involves a process of acquiring multiple genetic mutations in epithelial cells resulting in an activated stroma²⁶. Considering this fact as well as the complexity of tumor tissues comprised of epithelial, stromal and immune

Mutual genes	Cancer types	Regulation pattern	FC	Mutual genes	Cancer types	Regulation pattern	FC
BUB1	colorectal	up	12.11	GATM	cervical	down	3.15
	cervical	up	2.68		leukemia	up	25.34
	leukemia	down	2.43		ovarian	down	2.48
	ovarian	up	2.38		thyroid	down	3.01
PLCE1	colorectal	down	5.47	PTGER3	cervical	down	2.26
	ovarian	down	4.55		leukemia	up	3.44
	leukemia	up	2.95		ovarian	down	4.99
NEGR1	ovarian	down	17.44	SH3RF2	colorectal	down	2.54
	leukemia	up	3.18		lung	down	2.10
	thyroid	down	2.06		ovarian	down	2.89

Table 1. Mutual differentially expressed module genes in different tumor types.

cells, we analyzed data from laser micro-dissected epithelial samples from carcinoma cell population (CEPI) and healthy ovarian surface epithelial (OSE) tissues, only. Consequently, analysis of gene expression datasets from six independent studies associated with ovarian serous carcinoma by comparing gene expression levels between CEPI and OSE samples resulted with 698 mutual DETs, which were overrepresented with cell cycle, cancer signaling, drug and amino acid metabolisms.

The co-expression networks constructed around mutual DEGs in CEPI samples (CNC) and in OSE samples (CNO) represented observable differences in topological properties such as network size, density and clustering coefficient, and dispersed around different hub genes. Though the number of genes exhibiting co-expression pattern was almost two-fold higher in healthy state (i.e., CNO), network connectivity was higher in CNC. Consequently, a distinct tendency in gene expression correlations within diseased and healthy states was observed: the communication among genes decreases in diseased state; however, higher connectivity was developed between the genes.

Topological analysis of the CNC and CNO networks resulted with several modules. Among those, we identified a group of genes (84 genes) among mutual DEGs with strongly correlated co-expression pattern in CEPI samples, but not in OSE samples. The cooperation among these genes might be activated in response to the alteration of state, i.e., tumorigenesis. These genes may have a cooperative role in tumorigenesis process or this cooperative behavior may be a response of cells to the tumorigenesis process. In addition, the module could separate samples (patients) with long and short survival in ovarian cancer, and the death rate between groups was approximately twice, and at least 1.2 times. Therefore, this gene module represents a significant potential of “systems biomarkers” for therapeutics and prognosis in ovarian cancer.

Interestingly, the majority of these differential co-expressed genes was uncharacterized or encoding proteins taking roles in distinct signaling and metabolic pathways. Signaling proteins were mostly associated with signaling by GPCR. Dysregulation of G-protein and GPCR signaling leads to the initiation and progression of malignant tumor and their metastatic spread. More specifically mitogen-activated protein kinases (MAPK) signaling pathways such as Interleukin receptor SHC, PI3K-Akt, and Ras-Raf-Mek-Erk signaling, in which defect or alteration cause uncontrolled growth in terms of tumorigenesis^{27–29}. Thus, it was speculated that module genes of these pathways might be considered as tumorigenesis related genes in ovarian cancer. The metabolism associated genes were encoding enzymes of energy, amino acid and protein metabolisms as expected, and represent potential for metabolic markers in ovarian cancer therapy.

The dense co-expression pattern of this gene set (CMC) may be as a result of the localization of the genes on the same chromosome or their transcription may be controlled through common transcriptional regulatory elements. When the chromosomal locations of the genes were analyzed, we observed that the module genes were distributed to almost all human chromosomes at different ratios. On the other hand, the transcription of the majority (89%) of module genes were regulated by six TFs, especially GATA2, YBX1 and AR. Unlike TFs, miRNAs represented a diverse regulatory pattern, where module genes were regulated by distinct miRNAs. This observation supports the hypothesis that TFs play general roles in transcriptional regulatory mechanisms; however, miRNAs take more specific duties⁷.

Screening the module genes across nine tumor tissues (ovarian, breast, cervical, prostate, lung, colorectal, pancreatic, thyroid, and leukemia) indicated distinct expression patterns of the CMC in different tumor tissues and strongly suggested that CMC is specific to ovarian cancer and the module genes play an important role in epithelial cell originated tumorigenesis.

The grand challenge in ovarian cancer research is that the origin of epithelial ovarian cancer is not clearly defined. Recent studies suggested that ovarian cancer may originate from precursor lesions located in the fallopian tubal epithelium³⁰. This issue should be also considered in further researches. However, harboring a stem cell niche, OSE differs from the other differentiated epithelial cells, and malignant transformation may be explained by the presence of stem cell niches in those areas³¹. From this point of view, the results of the present study may have important implications for understanding ovarian cancer pathogenesis. Future research investments in experimental validations of corresponding disease biomarkers might provide new insights on ovarian cancer prognosis and treatment strategies that continue to place a significant burden on global health.

Conclusions

Our study has demonstrated the presence of a novel prognostic gene module, which was differentially co-expressed and predictive of patient survival in ovarian cancer. Moreover, genes of the co-expressed module were co-regulated by a certain number of TFs and various miRNAs in an integrated manner, but did not represent a common regulatory pattern across different cancer types, and therefore the co-expressed and co-regulated module was specific to ovarian cancer. The proposed gene module will enable valuable insights through understanding the tumor formation and progression in ovarian epithelial cells. The set of genes in the module could be considered as systems biomarkers which may be used for screening or therapeutic purposes in ovarian carcinoma; more efforts are required to experimental and clinical validation of the findings.

Methods

Gene expression datasets. Taking into account the heterogeneous nature of ovarian cancer, we focused on micro-dissected epithelial samples from high grade serous ovarian carcinoma^{32–37}. As a result of an extensive screening on ovarian cancer associated transcriptome datasets, six datasets (GSE14407³², GSE18521³³, GSE23391³⁴, GSE27651³⁵, GSE38666³⁶, and GSE40595³⁷) that included laser micro-dissected epithelial samples from carcinoma cell population (CEPI) and healthy ovarian surface epithelial (OSE) tissues were considered. The raw data consisting of 191 samples (140 CEPI and 51 OSE samples) were obtained from the NCBI Gene Expression Omnibus (GEO)³⁸, which is a public functional genomics data repository supporting MIAME-compliant data submissions.

Identification of differentially expressed genes. To characterize differentially expressed genes (DEGs), each dataset was normalized by means of the Robust Multi-Array Average (RMA) expression measure³⁹ as implemented in the “affy” package⁴⁰ of R/Bioconductor platform (version Rx64 3.3). DEGs were identified from the normalized log-expression values using the multiple testing option of LIMMA (linear models for microarray data)⁴¹. Benjamini-Hochberg’s method was used to control the false discovery rate. An adjusted p-value threshold of 0.01 with a fold-change cutoff of 2 was used to determine the statistical significance of differential expression.

Construction of co-expression networks in diseased and healthy states. Mutual DEGs of all datasets were determined. Expression profiles of CEPI and OSE samples were separated, and two new data subsets (consisting of 140 and 51 samples, respectively) were constructed using the expression profiles of mutual DEGs within diseased (CEPI) and healthy (OSE) states. To eliminate batch effects and ensure a similar empirical distribution of each array, both data subsets were normalized via RMA expression measure. In case of the repetitive DEGs, the mean expression values were computed and employed in further analyses. We computed the Pearson correlation coefficient (PCC) of the expression profiles between every pair of DEGs in each dataset. The resultant PCCs were approximately normally distributed in each case. A PCC cut-off, corresponding to p-value < 0.05, was determined and employed to identify statistical significance of the pairwise correlations. We established two co-expression networks, CNC and CNO, representing diseased and healthy states (i.e., CNC: co-expression network in CEPI samples, CNO: co-expression network in OSE samples), between the significantly co-expressed DEGs.

Determination of network modules and their differential co-expression. To identify network modules of these networks, co-expression networks were analyzed using MCODE plugin of Cytoscape (v.2.8.3)⁴². Ranking of modules was based on MCODE scores (i.e., average connectivity). In selection of the modules for further analyses, modules with at least 10 nodes (genes), average connectivity ≥ 10 and network density ≥ 0.80 were considered. Since they might give important clues on disease pathogenesis and prognosis, we focused on modules of CNC, and co-expression pattern of these modules were also investigated in healthy state (i.e., within the CNO). Modules with significantly altered topological metrics (i.e., network density and clustering coefficient) were considered as differentially co-expressed modules between diseased and healthy states.

Topological and functional enrichment analyses. Local and global topological features of networks and their modules were represented by several metrics, including degree, betweenness connectivity, network density, and clustering coefficient, and were determined via NetworkAnalyzer⁴³ and Cytohubba⁴⁴ plugins of Cytoscape (v.2.8.3). The dual-metric approach^{12, 14} incorporating degree as a local metric and betweenness centrality as a global metric was employed at identification of hub molecules, and top five molecules in terms of any of the metrics were presented as hubs.

Pathway enrichment analyses of gene sets (within networks or modules) were performed through DAVID bioinformatics tool (v.6.8)⁴⁵ using KEGG⁴⁶, and Reactome⁴⁷ as the data sources. P-values were determined through Fisher Exact test and adjusted via Benjamini-Hochberg’s method. A threshold of adjusted p-value < 0.05 was used to determine the statistical significance of enrichment results. In addition, to characterize the pathway or process annotation of each gene, we manually searched GeneCards Human Gene Database⁴⁸.

Prognostic power analysis. We extracted gene expression profiles of module genes from the ovarian cancer dataset (GSE18521³³), which supports clinical information for survival analysis. Then, PCA was performed based on gene expression profiles of 84 module genes, and samples were separated into two clusters using k-means algorithm taking into consideration the first three principle components. The survival time statistics were calculated by log-rank test and visualized in Kaplan-Meier survival curve. Cox (proportional hazards) regression was also employed to estimate hazard ratio.

Identification of transcriptional regulatory networks. To investigate the common transcriptional regulators (i.e., TFs and/or miRNAs) of genes within the differentially co-expressed module (CMC), experimentally validated transcriptional regulatory interactome data⁷ were updated via addition of TF-target gene interactions from the latest version of HTRI database⁴⁹ and miRNA-target gene interactions from mirTarBase (ver.6.0)⁵⁰. For each component of CMC, the number of TFs and miRNAs regulating the mentioned gene was counted, and the mutual TFs and miRNAs regulating the significant portions of the module components were determined.

Screening the differential expression of the module in different tumor types. Genes of the differentially co-expressed module in CEPI samples were investigated across nine tumor tissues (i) to determine specificity of the module to ovarian cancer, and (ii) to analyze the expression pattern of the module genes in different tumor tissues. For these purposes, gene expression datasets associated with ovarian cancer (GSE19352⁵¹, GSE26712⁵²), breast cancer (GSE15852)⁵³, cervical cancer (GSE9750)⁵⁴, prostate cancer (GSE6919)⁵⁵, lung cancer (GSE19804)⁵⁶, colorectal cancer (GSE32323)⁵⁷, pancreatic cancer (GSE28735)⁵⁸, thyroid cancer (GSE3467)⁵⁹ and leukemia (GSE26725)⁶⁰ were employed, and statistically significant DETs were identified (as described before) comparing tumor vs. healthy tissues. Module genes were screened within the identified DEGs for each cancer type, and the expression pattern of the module genes in different tumor tissues (comprising of stroma, immune and epithelial cells) were analyzed.

References

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA: Cancer J. Clin.* **64**, 9–29 (2014).
2. American Cancer Society. Cancer Facts & Figures 2017. *Atlanta: American Cancer Society* (2017).
3. Hippisley-Cox, J. & Coupland, C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *BMJ*. **344**, doi:10.1136/bmj.d8009 (2012).
4. Buys, S. S. *et al.* Effect of screening on ovarian cancer mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *Jama* **305**, 2295–2303 (2011).
5. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–55 (2003).
6. Karagoz, K., Sevimoglu, T. & Arga, K. Y. Integration of multiple biological features yields high confidence human protein interactome. *J. Theor. Biol.* **403**, 85–96 (2016).
7. Gov, E. & Arga, K. Y. Interactive cooperation and hierarchical operation of microRNA and transcription factor crosstalk in human transcriptional regulatory network. *IET Syst Biol.* doi:10.1049/iet-syb.2016.0001 (2016).
8. Mardinoglu, A., Gatto, F. & Nielsen, J. Genome-scale modeling of human metabolism – a systems biology approach. *Biotechnol. J.* **8**, 985–996 (2013).
9. Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome Networks and Human Disease. *Cell* **144**, 986–998 (2011).
10. Sevimoglu, T. & Arga, K. Y. The role of protein interaction networks in systems biomedicine. *Comput. Struct. Biotechnol. J.* **11**, 22–27 (2014).
11. Calimlioglu, B. *et al.* Tissue-Specific Molecular Biomarker Signatures of Type 2 Diabetes: An Integrative Analysis of Transcriptomics and Protein-Protein Interaction Data. *OMICS* **19**, 563–73 (2015).
12. Karagoz, K., Sinha, R. & Arga, K. Y. Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. *OMICS* **19**, 115–130 (2015).
13. Sevimoglu, T. & Arga, K. Y. Computational Systems Biology of Psoriasis: Are We Ready for the Age of Omics and Systems Biomarkers? *OMICS* **19**, 669–687 (2015).
14. Kori, M., Gov, E. & Arga, K. Y. Molecular signatures of ovarian diseases: Insights from network medicine perspective. *Syst Biol Reprod Med.* **62**, 266–82 (2016).
15. de la Fuente, A. From ‘differential expression’ to ‘differential networking’ – identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26**, 326–33 (2010).
16. Ideker, T. & Krogan, N. J. Differential network biology. *Mol Syst Biol* **8**, 565 (2012).
17. Hsu, C. L., Juan, H. F. & Huang, H. C. Functional Analysis and Characterization of Differential Coexpression Networks. *Sci Reports* **5**, 13295, doi:10.1038/srep13295 (2015).
18. Zhang, J. *et al.* Using gene co-expression network analysis to predict biomarkers for chronic lymphocytic leukemia. *BMC Bioinformatics* **11**, S5, doi:10.1186/1471-2105-11-S9-S5 (2010).
19. Walley, A. J. *et al.* Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue. *Int J Obes (Lond)*. **36**, 137–47 (2012).
20. Doig, T. N. *et al.* Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. *BMC Genomics* **14**, 469, doi:10.1186/1471-2164-14-469 (2013).
21. Wolf, D. M., Lenburg, M. E., Yau, C., Boudreau, A. & van ‘t Veer, L. J. Gene co expression modules as clinically relevant hallmarks of breast cancer diversity. *PLoS One* **9**, e88309, doi:10.1371/journal.pone.0088309 (2014).
22. Zhou, Y. *et al.* ICan: an integrated co-alteration network to identify ovarian cancer-related genes. *PLoS One* **10**, e0116095, doi:10.1371/journal.pone.0116095 (2015).
23. Cai, S. Y. *et al.* Gene expression profiling of ovarian carcinomas and prognostic analysis of outcome. *J Ovarian Res.* **8**, 50, doi:10.1186/s13048-015-0176-9 (2015).
24. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **2**, 397–406 (2014).
25. Dandara, C. *et al.* Precision Medicine 2.0: The Rise of Global Innovation, Superconnectors, and Design Thinking. *OMICS* **20**, 493–495 (2016).
26. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* **19**, 1423–37 (2013).
27. Altomare, D. A. & Testa, J. R. Perturbations of the AKT signaling pathway in human Cancer. *Oncogene* **24**, 7455–7464 (2005).
28. Ravichandran, K. S. Signaling via Shc family adapter proteins. *Oncogene* **20**, 6322–30 (2001).
29. McCubrey, J. A. *et al.* Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochim Biophys Acta* **1773**, 1263–84 (2007).
30. Erickson, B. K., Conner, M. G. & Landen, C. N. Jr. The role of the fallopian tube in the origin of ovarian cancer. *AJOG*, doi:10.1016/j.ajog.2013.04.019 (2013).
31. Flesken-Nikitin, A. *et al.* Ovarian surface epithelium at the junction area contains a cancer-prone stem cell niche. *Nature* **495**, 241–245 (2013).
32. Bowen, N. J. *et al.* Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med Genomics* **2**, 71, doi:10.1186/1755-8794-2-71 (2009).

33. Mok, S. C. *et al.* A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2. *Cancer Cell* **16**, 521–32 (2009).
34. Shahab, S. W. *et al.* Evidence for the complexity of microRNA-mediated regulation in ovarian cancer: a systems approach. *PLoS One* **6**, e22508, doi:10.1371/journal.pone.0022508 (2011).
35. King, E. R. *et al.* The anterior gradient homolog 3 (AGR3) gene is associated with differentiation and survival in ovarian cancer. *Am J Surg Pathol.* **35**, 904–12 (2011).
36. Lili, L. N., Matyunina, L. V., Walker, L. D., Benigno, B. B. & McDonald, J. F. Molecular profiling predicts the existence of two functionally distinct classes of ovarian cancer stroma. *Biomed Res Int.* **2013**, 846387, doi:10.1155/2013/846387 (2013).
37. Yeung, T. L. *et al.* TGF- β modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res.* **73**, 5016–28 (2013).
38. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* **41**, 991–995 (2013).
39. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
40. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy- analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
41. Smyth, G. K. Limma: linear models for microarray in bioinformatics and computational biology solutions using R and bioconductor 397–420 (Springer, 2005).
42. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
43. Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008).
44. Chin, C. H. *et al.* CytoHubba: identifying hub objects and sub networks from complex interactome. *BMC Syst Biol.* **8**, S11, doi:10.1186/1752-0509-8-S4-S11 (2014).
45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
46. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, 199–205 (2014).
47. Fabregat, A. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **44**, D481–D487 (2016).
48. Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database*, baq020, doi:10.1093/database/baq020 (2010).
49. Bovolenta, L. A., Acencio, M. L. & Lemke, N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* **13**, 405 (2012).
50. Chou, C. H. *et al.* miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* **44**, D239–47 (2016).
51. Iorio, E. *et al.* Activation of phosphatidylcholine cycle enzymes in human epithelial ovarian cancer cells. *Cancer Res.* **70**, 2126–35 (2010).
52. Vathipadiekal, V. *et al.* Creation of a Human Secretome: A Novel Composite Library of Human Secreted Proteins: Validation Using Ovarian Cancer Gene Expression Data and a Virtual Secretome Array. *Clin Cancer Res.* **21**, 4960–9 (2015).
53. Pau Ni, I. B. *et al.* Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context. *Pathol Res Pract.* **206**, 223–8 (2010).
54. Scotto, L. *et al.* Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. *Genes Chromosomes Cancer* **47**, 755–65 (2008).
55. Chandran, U. R. *et al.* Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer* **7**, 64, doi:10.1186/1471-2407-7-64 (2007).
56. Lu, T. P. *et al.* Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev.* **19**, 2590–7 (2010).
57. Khamas, A. *et al.* Screening for epigenetically masked genes in colorectal cancer Using 5-Aza-2'-deoxycytidine, microarray and gene expression profile. *Cancer Genomics Proteomics* **9**, 67–75 (2012).
58. Zhang, G. *et al.* Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin Cancer Res* **19**, 4983–93 (2013).
59. He, H. *et al.* The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci USA* **102**, 19075–80 (2005).
60. Vargova, K. *et al.* MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. *Blood* **117**, 3816–25 (2011).

Acknowledgements

Financial support by Marmara University, Scientific Research Projects Committee (BAPKO) through project FEN-C-DRP-110915-0445.

Author Contributions

E.G. analyzed the data and evaluated the results. E.G. and K.Y.A. designed the algorithms and the analysis framework. K.Y.A. conceived and directed the study. E.G. and K.Y.A. wrote the paper. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-05298-w

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017